A macroeconomic forecasting market

Florian Teschner & Christof Weinhardt

Journal of Business Economics
Zeitschrift für Betriebswirtschaft

ISSN 0044-2372
Volume 85
Number 3

J Bus Econ (2015) 85:293-317
DOI 10.1007/s11573-014-0741-5

Springer

J Bus Econ (2015) 85:293–317
DOI 10.1007/s11573-014-0741-5

ORIGINAL PAPER

A macroeconomic forecasting market

Florian Teschner · Christof Weinhardt

Published online: 25 July 2014

Abstract Macro-economic forecasts are used extensively in industry and government even though the historical accuracy and reliability is disputed. Modern information systems facilitate participatory, crowd-sourced processes that harness the collective intelligence. One instantiation of such *wisdom of the crowds* are prediction markets which have proven to successfully forecast the outcome of elections, sport events and product sales. Consequently we specifically design a prediction market for macro-economic variables in Germany. The proposed market design differs significantly from previous ones. It solves some of the known problems such as low liquidity and partition-dependence framing effects. 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. Analyzing the market-generated forecasts, we find that forecast accuracy improves constantly over time and that generated forecasts performed well in comparison to the Bloomberg-survey forecasts. From an individual perspective, market participants interact in a repeated decision-making environment closely resembling decision-making in financial markets. We analyze the impact of cognition, risk-aversion and confidence on trading activity and success.

F. Teschner (✉) · C. Weinhardt
Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
e-mail: teschner@kit.edu

C. Weinhardt
e-mail: weinhardt@kit.edu

Keywords Prediction Markets · Macro-economic forecasting · User heterogeneity · Market-engineering · Crowd sourcing

JEL Classification D40 · D02 · E37

1 Introduction

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 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). 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 makers) 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 forecasts to form a common informative basis.

Advances in information systems are changing information aggregation in many contexts: political institutions increasingly open up for grassroots democracy and open discussion of societal innovation (Campbell et al. 2009; Brabham 2008), ad-hoc communities use social media to coordinate (Kittur and Kraut 2008) and companies gradually shift decisions towards a broad basis of employees and allow for user-driven innovation (Gillen and Plott 2012; Cowgill et al. 2009). An underlying theme of this trend is using the collective intelligence and wisdom of the crowd. As one way to facilitate the collective intelligence, 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 stock markets do. 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; Cowgill and Zitzewitz 2013).

To sum up, evidence so far suggests that prediction markets provide considerable advantages in terms of continuous forecasting, participation and information revelation. This paper aims to analyze the predictive power of a prediction market for macro-economic forecasts. In contrast to previous studies, we analyze the accuracy of a prediction market for long forecast horizons (up to 2 months) over a comparatively long period (2 years). Analyzing the participant heterogeneity in our market, we 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 is directly negatively affected if participants make systematically biased decisions. This is a relatively well known, but still not well understood or studied, hypothesis of the behavioral finance literature. Our setup is well-suited to studying the behavioral aspects of decision making because, in contrast to financial markets (1) the value of shares in our market is ultimately known and (2) we can measure the participants' ex-post trading performance.

2 Research questions

The research questions are grouped in three interlinked categories of market design, long-term participant motivation and individual trading behavior.

2.1 Market design

The main research question is how to design a participatory, crowd-based platform to facilitate information aggregation of macro-economic variables. Compared to previous market designs we propose a different contract structure which allows for longer forecast horizons. We evaluate our market design by comparing the forecast accuracy of the crowd-sourced platform with an expert panel.

2.2 Participant motivation

This subsequently leads to the question of how participants can be motivated to contribute and share their information for longer time horizons. Previous studies show that prediction markets work for short forecast horizons. We want to see if they work even beyond the cited novelty factor. We accomplish this by first implementing a specifically designed market environment. Secondly, we design a play-money incentive scheme which rewards participants according to their performance. Varying the incentive size we measure the effect of incentives on overall market activity and forecast performance.

2.3 Individual trading behavior

In order to further motivate participants intrinsically we need to understand their heterogeneity. Hence the next question is: which participants clustered by their personality traits are more motivated to contribute? More importantly, are certain participants more likely to perform above average? If we find any differences in the activity level and performance, do the additional contributions improve the crowd-sourced forecast?

3 Related work

The following section reviews related work in the domains of market engineering and prediction markets. It details the advantages of prediction markets over surveys and presents previous market designs for economic forecasting markets.

3.1 The wisdom of crowds and market engineering

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. 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 believe the weather to be good for the orange harvest, and vice versa. The market price for these futures is the aggregate belief of all individual forecasts. Roll (1984) shows that orange-concentrate prices forecast weather extremes more precisely than the US national weather service. Thereby markets provide incentives for information revelation and act as a mechanism for aggregating information.

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) and Weinhardt et al. (2006) propose the market engineering framework. 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, he 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 assess 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.

3.2 Prediction markets

One way to harness the wisdom of the crowd is prediction markets. 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. A famous example is the Iowa political stock market (PSM) which predicts the outcome of US 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 US presidential election outcome. The US 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. 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).

According to Arrow et al. (2008) prediction markets reflect a fundamental principle of market-based pricing: a mechanism to collect and aggregate dispersed information. Following the Hayek-hypothesis, the price mechanism is an efficient way to aggregate asymmetric distributed information (Hayek 1945). Efficient Market Hypothesis maintains that market prices fully reflect all available information (Fama 1970, 1991). In an informationally efficient market, price changes must be unpredictable if they are properly anticipated, that is if they fully incorporate the information and expectations of all market participants (Lo 2007).

On the other hand, if markets incorporate all available information participants have to assume that their private information is reflected by prices as well. Assuming that they are rational agents they should choose not to trade. This paradox is often called the no trade theorem (Milgrom and Stokey 1982). Wolfers and Zitzewitz (2006a) point out that it remains an open theoretical question "why there is trade in prediction markets". The markets' predictive power basically comes from three sources: an incentive to truthfully reveal expectations, incentives to research new information and an algorithm for aggregating opinions (Wolfers and Zitzewitz 2004).

Through the price formation process, markets aggregate information across traders, solving what would otherwise be complex aggregation problems. Additionally, price levels provide continuously updated dynamic forecasts.

Evidence suggests that such markets give unbiased, relatively accurate forecasts well in advance of outcomes and can outperform existing alternatives (Berg and Rietz 2006). It is therefore hard to find information which is not embodied in the market (Hanson 2006). In general, traders have five reasons to participate: the novelty factor, confidence in one's knowledge relative to other traders, confidence in one's ability to interpret news relative to other traders, confidence in one's talents as a trader and risk-seeking behavior (Forsythe et al. 1992).

In his seminal work Roth (2009) summarizes some guiding principles for designing markets. He highlights three key elements: Thickness, congestion, and safety. Having sufficient thickness means that the market attracts enough participants to make it thrive. Congestion refers to markets that are too thick too fast. It is a problem in markets in which transactions are heterogeneous and offers cannot be made to the whole market at once. Safety refers to an environment in which all participants feel secure enough to make decisions based on their best interests, rather than attempt to game a flawed system. It also implies that the market system is sufficiently simple to participate as opposed to transacting outside the market. These general principles of market design translate well into the prediction market domain.

Thickness is most often addressed by shuffling lottery prizes among top traders. The prediction topic (e.g. sports) itself provides the users with an engaging environment, which motivates sustainable long-term participation. Congestion is addressed by providing a standardized way to trade expectations about future events. Previous work (Teschner et al. 2011; Seuken et al. 2012) shows that prediction markets should provide concepts to hide and reduce the complexity, in order to enable the users to intuitively express their beliefs about an event. Finally, in order to fully exploit the "wisdom of the crowds", participation must be easy and access frictionless.

4 Advantages of prediction markets over surveys

Prediction markets offer a number of advantages over surveys. Table 1 summarizes the key distinguishing elements. Prediction markets are continuous and ongoing, allowing immediate revelation of new information (Rothschild 2009). As they are open around the clock, participants can trade whenever they receive new information and therefore react to news immediately (Snowberg et al. 2007). Although some surveys offer a small incentive in return for participation, the incentives earned by traders in a prediction market increase in proportion to the quality of the information provided. Unlike surveys, a market provides immediate feedback to participants, allowing them opportunities to reassess their own information and to respond. The feedback enables participants to learn 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 (Teschner et al. 2011).

The market interface is interactive and the setting gamified, in marked contrast to most surveys, providing further incentives for participation. Most surveys rely on random samples for validity and accuracy. In prediction markets, on the other hand, those with the best information are the best participants—the very individuals who are most likely to self-select into the market. Additionally, as successful participants accumulate their profits they gain forecasting weight over time compared to less successful participants. With surveys, this process of self-selection would introduce a sampling bias, but with markets, the incentive structure forces low performers out of the market. Turning to the disadvantage of markets over surveys, one has to mention the higher complexity burdening participants (Graefe 2010). First, they have to understand the trading mechanism and secondly have to understand how events are related to contracts. This process is more structured and better researched for surveys. The forecast performance of prediction markets is still in debate. On the positive side, they have proven repeatedly to be very potent information aggregation mechanisms (Berg et al. 2008; Ledyard et al. 2009; Bennouri et al. 2011). Other evidence however suggests that the relative performance advantage of markets may be small compared to surveys or polls (Goel et al. 2010; Erikson and Wlezien 2008; Rothschild 2009).

4.1 Manipulation in prediction markets

Manipulation can be defined as influencing market prices to knowingly wrong levels. Allen and Gale (1992) considered three types of manipulation: action-based (changing the underlying fundamentals), information-based (spreading false information), and trade-based (buying, selling of shares). Deck et al. (2013) provide a recent review on manipulating prediction markets. A common line of argument is that a manipulator is just another kind of noise trader (Hanson 2006; Hanson and Oprea 2007). If well-informed traders seized the opportunity to profit from manipulative trades, market accuracy might actually increase. An example is the Internet-based prediction market In trade being temporarily manipulated in the 2012 US presidential election (Thomason 2012). However, empirical studies show that if manipulation occurs, it is typically only short-lived (Wolfers and Zitzewitz 2004). The prevailing opinion is that long-term price manipulation does not occur or succeed and that manipulators are unable to distort price accuracy (Hanson et al. 2006; Berg et al. 2008; Oliven and Rietz 2004). Wolfers and Zitzewitz (2004) concluded that there “have been several known attempts at manipulation of these markets, but none of them had much of a discernible effect on prices, except during a short transition phase”. Deck et al. (2013) summarize that “research suggests prediction markets are robust to manipulation attacks”.

Table 1 Prediction markets and surveys compared

	Prediction markets	Surveys
Continuous forecasts	+	–
Reacting to news	+	–
Feedback/participant performance	+	o
Incentives	+	–
Complexity	–	+
Forecast performance	o	o

Positive (+), neutral (o), and negative (−) ratings for both methods

4.2 Prediction markets for macro-economic events

Markets for macro-economic variables have been used since the 1980s. 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.

4.3 A case for a new market design

As outlined in the last section, previous research focuses on binary contracts. 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, he has to pre-estimate a reasonable range of possible outcomes. Secondly, he 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 call "partition-dependence-framing-effect". 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 systematic errors. In particular, people underweight outcomes which 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 (2006a, b) 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 paper, tries to circumvent the presented problems by representing events as linearly paid out contracts.

4.4 Personality traits and trading behavior

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). A promising approach to describe and explain financial decision making may be the explicit consideration of psychological factors such as risk aversion, cognitive reflection and (over)confidence.

Risk aversion has been cited by several authors as a reason for certain market behavior (Subrahmanyam 1991). Risk-aversion may cause participants to not make profitable but risky trades in a market. If all participants suffer from this aversion, valuable information may not be impounded into prices, thereby reduce the predictive power of a market. Unfortunately, useful insights can only rarely be obtained from empirical data on security prices since risk aversion measures must be obtained independently of trading data. By merging household investment decisions with data from external risk questionnaires, Wärneryd (1996) cannot find a relationship between risk-aversion and portfolio choice. This is in line with findings from an empirical asset market in which participants' portfolio choice was unrelated to a risk aversion proxy (Güth et al. 1997). In contrast to portfolio choice, individual market behavior seems to be influenced by risk aversion. Fellner and Maciejovsky (2007) find that the higher the degree of risk aversion the lower the observed market activity. They also find a gender difference in both risk aversion and market activity. They find that women are more risk averse than men and submit fewer orders. Similarly, Kirchler and Maciejovsky (2002) find that the higher the degree of risk aversion the lower the total number of contracts traded.

In an early experimental study, Ang and Schwarz (1984) separated participants in two markets according to their degree of risk aversion. They show that the market with lower risk aversion (speculators) exhibit greater volatility but it also tended to converge closer and faster to the expected equilibrium price than the risk-averse (conservative) market. Finally, the interaction between risk attitude and overconfidence with respect to trading activity deserves further attention. Theoretical finance models predict higher market activity as a consequence of overconfidence (Barber and Odean 2001). Overconfidence refers to the habit of overestimating one's ability to perform a task. People tend to be overconfident about their capabilities and level of knowledge. This could also negatively impact the information content of prices. According to Barber and Odean (2000) overconfidence causes excess trading that can be risky to financial well-being and detrimental to market quality.

5 An experiment on crowdsourcing macro-economic forecasts

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. After the first year (named version one, or round one) we slightly adapted the system for the second year (version two, or round two).

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

5.1 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 to equations in the table below.

An Export (V1) or GDP contract is worth: $100 \pm \alpha$ times the percentage change for an indicator in play money. We set α to 10 (e.g. a change of 2.1 % results in a price of 121). Therefore, the representable outcome values 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 event (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 express reasonable accuracy. Additionally, German statistical data releases rarely come with more than one decimal. Table 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 (t1, t2, t3). 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 functions defined in Tables 1 and 2.

As soon as the trading in one stock stops a new stock of the same indicator (e.g. t4) 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. 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) economic indicators. The exact definitions are given on the web page. The selection and precise definition follows what the media usually reports in a casual way.

5.2 Incentives

As mentioned, the market is a free to join play money market. In order to motivate participants intrinsically, we have provided two interface features; traders could follow their performance on a leader board and they could 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).

Table 2 Economic variables and their payoff functions

Indicator	Unit	Data release cycle	Number of payouts	Payout function
Exports (V1)	Relative changes$_{t-1}$	Monthly	12 (–)	$100 + \alpha \times \left(\frac{I_{t0}-I_{t-1}}{I_{t-1}}\right)$
Exports (V2)	Billion (absolute)	Monthly	2)	ABS (Number) + 30
GDP	Relative changes$_{t-1}$	Quarterly	4 (4)	$100 + \alpha \times \left(\frac{I_{t0}-I_{t-1}}{I_{t-1}}\right)$
Ifo index (V1)	Absolute changes$_{t-1}$	Monthly	3 (–)	$100 + \alpha \times (I_{t0} - I_{t-1})$
Ifo index (V2)	Points (absolute)	Monthly	– (12)	ABS (Points)
Inflation	Relative changes$_{t-12}$	Monthly	11 (12)	$100 + \alpha \times \left(\frac{I_{t0}-I_{t-12}}{I_{t-12}}\right)$
Investments	Relative changes$_{t-1}$	Quarterly	5 (4)	$100 + \alpha \times \left(\frac{I_{t0}-I_{t-1}}{I_{t-1}}\right)$
Unemployment	Million (absolute)	Monthly	12 (12)	$100 + \left(\frac{\text{ABS(Number)}}{100.000}\right)$

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. The payout functions determine how the corresponding stock is valued given a certain outcome

Due to the legal restrictions on gambling the EIX prediction market had to rely on play money. To increase participants' motivation and to provide incentives to contribute information, we have handed 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 issue. The incentives are divided in two parts: (1) monthly prizes and (2) yearly prizes. The eight yearly prizes (total value of 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: (1) they increase their portfolio value and (2) 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.

5.3 Interface

The default trading interface is displayed in Fig. 1. 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. He can change the limit price three ways; by (1) inserting the number directly (with a maximum of two decimals) or (2) by typing in her prediction, which is then translated into a limit price, or (3) 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 currently 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 the rest of their trading interface individually. By clicking on the small arrows, the 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 indicators' last year 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 order, the chosen interface is saved per user. When the user returns, the system opens the previously used interface elements on default.

As additional information, 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.

6 Results

In the following section we analyze the market's forecast performance. By splitting the market period in two phases, the first and the second round (year or version), we

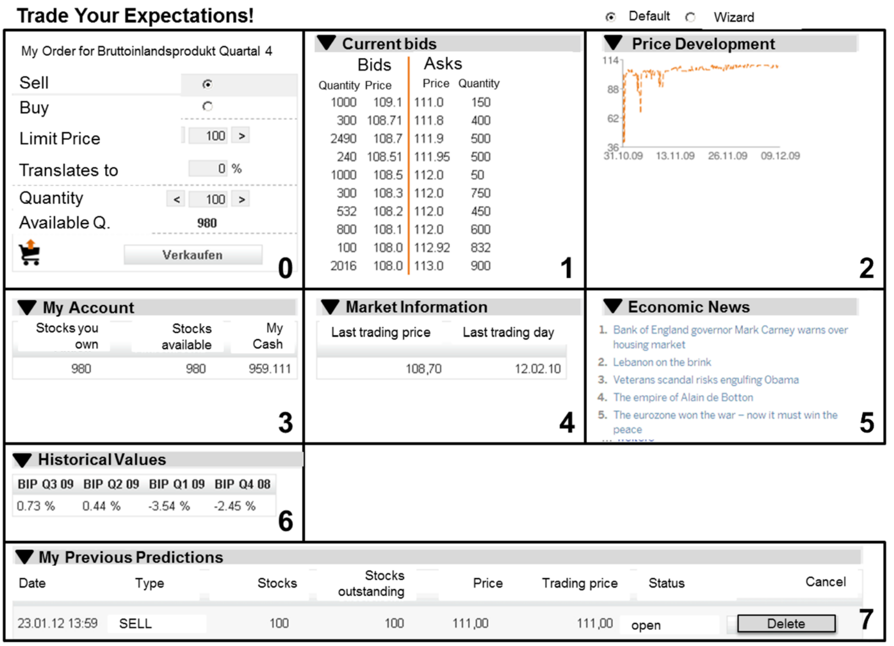

Fig. 1 Trading interface (English translation)

try to draw conclusions for market engineering. We start by presenting some descriptive market statistics and then evaluate the market design. 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. Analyzing individual participant behavior we (4) quantify the influence of a participant's behavioral traits on market activity and performance.

6.1 Participant activity

The following data includes the time span from 30th October 2009 till 31st of October 2011. In total, 1,342 (1,056 in the first round) participants registered at the EIX market, of those 809 (680) submitted at least one order. On average, every stock was traded 645 times (standard deviation 424). Figure 2 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 there for the rest of the experiment period.

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. In order to keep participants active and informed we sent out a weekly newsletter summarizing up-to-date economic news. In general, one can see that trading activity was lower in the second version (on average 127 vs. 86 orders per day). In order to better understand the difference in trading activity we ran two

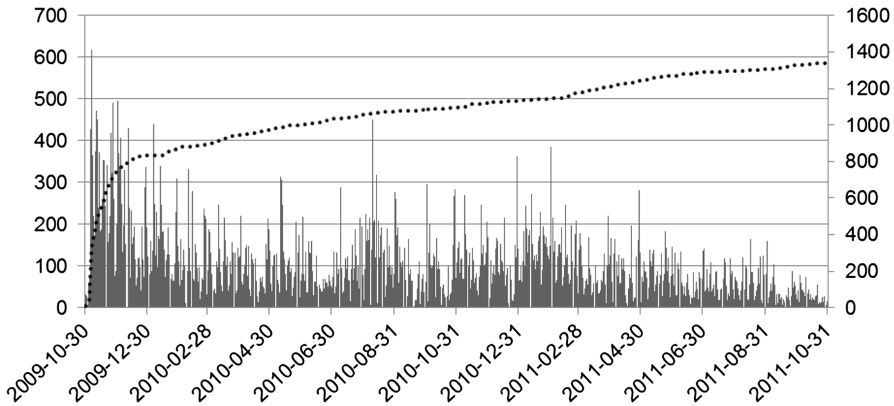

Fig. 2 Market activity over time (number of orders per day, *left scale*) and cumulative number of registered participants (*dotted line*, *right scale*)

OLS-regressions (Results depicted in Table 3). First we quantify the difference between the two versions (binary variable Version 2), second we add a continuous variable capturing the decreasing effect over time. In the second regression we add a novelty dummy for the first two months of market. We see that the second regression better captures the underlying effects (Adj. R2 25 vs. 20 %). Hence we conclude that there is indeed a substantial novelty effect of an additional 71 orders per day. Moreover taking the effect into account, the activity per day is only 18 orders lower on average.

As detailed in the previous section, participants could win non-cash prizes worth 36.000€ in the first version and 12.360€ in the second version. Hence, the potential reward for an order was substantially different between the two versions. Viewed from the market operator perspective; he invested 78 cents per order in version 1 and only 34 cents per order in version 2. If there are no differences in forecast performance this result indicates that the size of rewards is less important as theory suggests.

The market activity, quantified as the number of orders submitted, is displayed in four dimensions in Fig. 3. 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 in the morning and 1 am at night, with slight peaks during noon.

There is high activity during working hours, and higher activity during the week compared to the weekend. It seems that people trade while being at work. On average, users submitted orders with an order size of 759 units (median 300). With market prices between 90 and 140, the volume turnover is on average 93,320 (median 38,690). In Fig. 3 (lower right hand side), the number of submitted orders in each trade size category is displayed. Over 70 % of all orders are sized between 50 and 1,000 stock units.

Table 3 Analyzing trading activity and novelty effect

	Orders (OLS)	Orders (OLS)
Novelty	–	71.79***
Version 2	−36.36***	−18.19**
Days	−0.32***	−0.21***
Intercept	221.77***	164.98***
N	765	765
Adj. R2	20.03 %	25.06 %

Significance code: ***0.001, **0.01

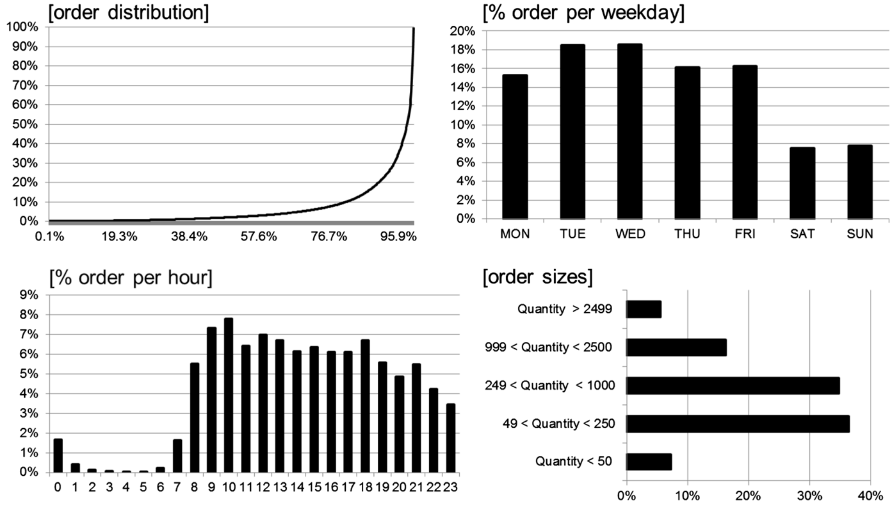

Fig. 3 Market activity from different perspectives

6.2 Measuring 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, b). The simplest form of forecast performance evaluation is to calculate the difference between the mean of beliefs and the outcome. In linear outcome markets, prices do not reflect the probability of an outcome but the market participants' 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 the $error_{t,i}$.

Studying the error terms we can analyse the forecast properties. To begin with, a forecast is unbiased if it has zero mean. Secondly forecast errors are supposed to fall as more information arrives over time. 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). A detailed discussion of the common measures can be found in (Fildes and Stekler 2002).

A problem with the standard error functions 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 $\text{Theil}'\text{s U} = \frac{\text{RMSE}_{\text{forecast}}}{\text{RMSE}_{\text{naiive}} - \text{RMSE}_{\text{forecast}}}$.

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.

6.3 Aggregate 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.

On an aggregated level, we compare the market generated forecasts eight days before the data release ($\text{forecast}_{8,i}$) to the fundamental value. Table 4 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 and the second version. As the mean error in the second version seems 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, $p < 0.1$ %).

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.

Table 4 Forecast performance comparison of version one and two

	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	–

Average price difference (error) between fundamental values and stock prices in t-8. Testing for a systematic bias (if the mean error equals 0)

Thus, we conclude that market forecasts 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.

An important question is if the market continuously aggregates information. In Fig. 4 the average and median absolute errors over time are depicted. Additionally, we plot the number of stocks (N) on which the average (median) is based upon. As the stocks have different time horizons (monthly and quarterly), the number increases. One can see a steady decreasing absolute error in the last 70 days. We run an OLS-regression to quantify the error reduction per day. In order to control for different indicator effects, we add dummies for each market indicator category (e.g. Export, GDP …). In the last 70 days, the average error is reduced by 0.12 price points per day (t-stat. 7.0; p value $\leq$0.1 %). We conclude that forecast uncertainty was reduced over time, information aggregation took place and hence the absolute error was reduced.

Finally, we compare the market forecast performance to an external benchmark forecast. Table 5 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 1. We compare the forecast errors (1) within the market and (2) 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 have provided macroeconomic forecasts for industrialized countries since October 1989. Every Friday consensus surveys 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.

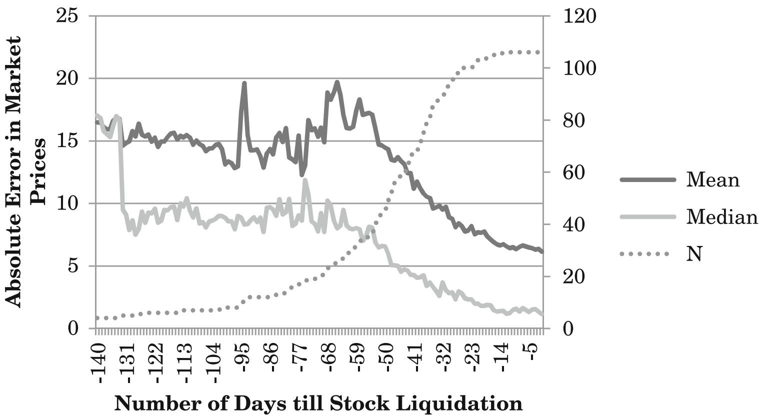

Fig. 4 Absolute forecast error over time

Table 5 Market forecast errors, compared within the market and benchmarked against the Bloomberg (BB) survey forecast

Indikator	Unit	MAE		MAPE		RMSE		Theil's U	
		EIX	Bloom-berg (BB)	EIX	BB	EIX	BB	EIX	BB
Export V1	%	3.56	2.88	123.2	69.4	4.79	3.73	0.41	0.52
Export V2	Bil	2.09	2.1	2.39	2.39	2.79	2.76	0.96	0.94
GDP	%	0.44	0.29	124.1	78.7	0.58	0.4	0.39	0.42
Inflation	%	0.18	0.11	21.91	12.3	0.27	0.16	0.74	0.39
Ifo V2	Points	1.34	0.89	1.21	0.81	1.71	1.06	0.81	0.57
Investments	%	1.21	–	121.4	–	1.44	–	0.5	–
Unemployment	Num	37,75	89,792	1.16	2.80	51,92	12,394	0.38	0.95

MAE mean absolute error, *MAPE* mean absolute percentage error, *RMSE* root mean squared error

As the indicators have different baselines and units, the absolute error (MAE, Table 5) 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 others. 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 is 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 (AR(1)-Model). 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 at least equally well. One should keep in mind that the market forecasts have always a longer forecast horizon than the Bloomberg forecast. That is they are continuously available and updated previous to the data release.

6.4 Individual participation and forecast performance

Finally, in order to obtain general feedback of the running platform, we run a survey in September 2010 at the end of version one. As part of the survey 104 participants completed the risk aversion questionnaire from Holt and Laury (2000) and Frederick's cognitive reflection test, as a proxy for the participants' cognitive abilities (Frederick 2005). Upon registration we asked participants to self-assess their market-knowledge and their knowledge of the German economy. We combined the scales to a confidence proxy. We split the participants in two groups the confident group (Confidence = 1) and the rest (Confidence = 0). The data for risk aversion (risk-averse) and the cognitive reflection (CRT-score) is coded the following way; for risk-averse traders risk-averse equals 1 and the CRT-Score equals 1 for cognitive reflective traders.

In the following section, we link behavioral aspects of the market participants and the quality of their decisions. We start by analyzing the trading activity on a user per user basis. In a second step, we document the influence of personality traits on trader behavior and profits. As Table 6 shows risk-averse participants trade less than other market participants. Additionally, we find that confident participants trade more often than other market participants.

Both findings confirm previous work by Fellner and Maciejovsky (2007); Kirchler and Maciejovsky (2002), who find that that risk-averse traders trade less on average and Barber and Odean (2001) showing that confidence is positively correlated with market activity. Turning to trading behavior, we find that confident traders use above average and cognitive reflective traders below average order sizes. Risk-aversion does not seem to play a role here. In order to analyze the influence of a trader's personal attributes on her performance, we conduct a logit-regression on whether a trade was profitable or not on a per order basis (Table 6, Profit model). We find a significance influence of the participant's risk-aversion and confidence level. In line with previous studies on over-confidence, we also find that confident traders perform below average. However, contrary to Oliven and Rietz (2004) and Barber et al. (2009) in our market, risk-averse traders perform worse than average. Extending previous studies we surprisingly find that the effect for higher cognitive reflection a proxy for IQ has a low to no significant influence on trading success. All in all, we conclude that participant heterogeneity plays a significant role in how traders act and perform in such prediction markets. This hints that preselecting participants (avoiding over-confident traders) might increase market quality.

Table 6 Analyzing trader heterogeneity: Activity, order size and profits

	Log (orders) (OLS)	Order size (OLS)	Profit (logit)
Confidence	0.97*	122.36*	−0.13***
Risk-averse	−0.99.	−90.16	0.46***
CRT-Score	0.32	−98.39*	0.05.
(Intercept)	3.35***	1,315.74***	−0.52***
N	104	27,335	27,335
Adj.R2	6.67 %	0.44 %	1.04 %

Significance code: ***0.001, *0.05, '.' 0.10

7 Conclusions

In this paper we provided an in-depth analysis of a new market for macroeconomic outcomes. The following section highlights the findings from three perspectives and gives a short outlook.

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 be 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 find a strong novelty effect increasing participation in the first months. Lowering the incentives in the second round even increases the activity per Euro "invested". Hence, it seems sustainable to run such markets. We strongly believe that by rigorously analyzing market properties there is potential for further improvement. This is indicated by early studies showing that the market interface (Teschner and Weinhardt 2012) and the information provided (Jian and Sami 2012) play a significant role in how traders interact in such markets. Moreover, the influence of the size as well as composition of the crowd might influence the predictive power.

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 market-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. 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.

Individual perspective: We furthermore analyze user actions in a repeated market environment where information processing plays a key role. Participants interact in a repeated decision-making environment closely resembling decision-making in financial markets. In our field experiment we analyze the impact of personality, cognition and risk on trading activity and individual profits. Summing up, it is rather simple to categorize (potential) traders ex ante using answers from short questionnaires. The implications of our results are straightforward; from a market operator's point of view, these results can be useful, since they can ex ante identify potential traders that do not have the 'right' predispositions.

This study offers a couple of future research directions. First off, the comparison between the market designs (digital vs. linear) could also be studied in lab experiments. Moreover, in our field experiment we used a continuous double auction as market mechanism. Hanson (2003) proposes the market scoring rule as a viable alternative. Using the market scoring rule would allow for simplified trading interfaces. Market interfaces generally seem to play a key role in the adoption and the success of complex markets (Seuken et al. 2012). Hence, future research should take the market interface into consideration.

References

Abramowicz MB (2004) Information markets, administrative decision making, and predictive cost-benefit analysis. Univ Chicago Law Rev 71:933–1020

Allen F, Gale D (1992) Stock-Price Manipulation. Rev Financ Stud 5(3):503–529

Ang JS, Schwarz T (1984) Risk aversion and information structure: an experimental study of price variability in the securities markets. J Financ 40(3):825–844

Armstrong JS (2008), 'Combining Forecasts', principles of forecasting: a handbook for researchers and practitioners, J. Scott Armstrong, ed., Norwell, MA: Kluwer Academic Publishers, 2001

Arrow KJ, Forsythe R, Gorham M, Hahn R, Hanson R, Ledyard JO, Levmore S, Litan R, Milgrom P, Nelson FD, Neumann GR, Ottaviani M, Schelling TC, Shiller RJ, Smith VL, Snowberg E, Sunstein CR, Tetlock PC, Tetlock PE, Varian HR, Wolfers J, Zitzewitz E (2008) Economics: the promise of prediction markets. Science 320(5878):877–878

Barber BM, Odean T (2000) Trading is hazardous to your wealth: the common stock investment performance of individual investors. J Financ 55:773–806

Barber BM, Odean T (2001) Boys will be boys: gender, overconfidence, and common stock investment. Quart J Econ 116(1):261–292

Barber BM, Lee Y-T, Liu Y-J, Odean T (2009) Just how much do individual investors lose by trading? Rev Financ Stud 22(2):609–632

Bennouri M, Gimpel H, Robert J (2011) Measuring the impact of information aggregation mechanisms: an experimental investigation. J Econ Behav Organ 78(3):302–318

Berg JE, Rietz TA (2006) The iowa electronic markets: stylized facts and open issues. In: Hahn RW, Tetlock PC (eds) Information markets: a new way of making decisions. AEI Press, Washington, pp 142–169

Berg J, Forsythe R, Nelson F, Rietz TA (2000) 'Results from a dozen years of election futures market research', working paper

Berg JE, Nelson FD, Rietz TA (2008) Prediction market accuracy in the long run. Int J Forecast 24(2):285–300

Berlemann M, Nelson F (2005) 'Forecasting inflation via experimental stock markets some results from pilot markets', ifo working papers (Ifo working papers No. 10)
Berlemann M, Dimitrova K, Nenovsky N (2005) 'Assessing market expectations on exchange rates and inflation: a pilot forecasting system for Bulgaria', working paper
Brabham DC (2008) 'Crowdsourcing as a model for problem solving convergence'. The international journal of research into new media technologies, Sage publications, 2008, 14, 75
Brenner M, Eldor R, Hauser S (1999) 'The price of options illiquidity', technical report, New York University, Leonard N. Stern School of Business
Camerer CF, Lowenstein G (2003), Behavioral Economics: Past, Present, Future' Advances in Behavioral Economics', *Princeton University Press*, pp. 3–51
Campbell DW, Sareen J, Stein MB, Kravetsky LB, Paulus MP, Hassard ST, Reiss JP (2009) Happy but not so approachable: the social judgments of individuals with generalized social phobia. Depress Anxiety 26(5):419–424
Clements M, Hendry D (2002) 'An overview of economic forecasting', A companion to economic forecasting, 1–18
Cowgill B, Zitzewitz E (2013) Corporate prediction markets: evidence from Google, Ford, and Koch industries. Working paper
Cowgill B, Wolfers J, Zitzewitz E (2009) Using prediction markets to track information flows: Evidence from google. In S. Das, M. Ostrovsky, D. Pennock, and B. Szymanksi (Eds.), *Auctions, Market Mechanisms and Their Applications, Volume 14 of Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pp. 3–3. Berlin: Springer
Croushore D, Stark T (2001) A real-time data set for macroeconomists. Journal of econometrics 105(1):111–130
Deck C, Lin S, Porter D (2013) Affecting policy by manipulating prediction markets: experimental evidence. J Econ Behav Organ 85:48–62
Erikson RS, Wlezien C (2008) Are political markets really superior to polls as election predictors*?*. Public Opinion Quart 72(2):190–215
Fama EF (1970) Efficient capital markets: a review of theory and empirical work. J Financ 25(2):383–417
Fama EF (1991) Efficient capital markets: ii. J Financ 46(5):1575–1617
Fellner G, Maciejovsky B (2007) Risk attitude and market behavior: evidence from experimental asset markets. J Econ Psychol 28(3):338–350
Fildes R, Stekler H (2002) The state of macroeconomic forecasting. J Macroecon 24(4):435–468
Forsythe R, Nelson F, Neumann GR, Wright J (1992) Anatomy of an experimental political stock market. Am Econ Rev 82(5):1142–1161
Forsythe R, Rietz TA, Ross TW (1999) Wishes, expectations and actions: a survey on price formation in election stock markets. J Econ Behav Organ 39:83–110
Frederick S (2005) Cognitive reflection and decision making. J Econ Perspect 19(4):25–42
Gadanecz B, Moessner R, Upper C (2007) 'Economic derivatives', BIS quarterly review, 68
Gillen BJ, Plott CR, Shum M (2012) Information aggregation mechanisms in the field: Sales forecasting inside intel. Working paper
Gjerstad S, Hall M (2005), 'Risk aversion, beliefs, and prediction market equilibrium', *Economic Science Laboratory, University of Arizona*
Goel S, Reeves DM, Watts DJ, Pennock DM (2010, June). Prediction without markets. In Proceedings of the 11th ACM conference on Electronic commerce (pp. 357-366). ACM
Graefe A (2010) Are prediction markets more accurate than simple surveys? Foresight. Int J Appl Forecast Issue 19:39–43
Gurkaynak R, Wolfers J (2006) 'Macroeconomic Derivatives: an initial analysis of market-based macro forecasts, uncertainty, and risk', technical report, National Bureau of Economic Research, Inc, No. 11929
Güth W, Krahnen JP, Rieck C (1997) Financial markets with asymmetric information: a pilot study focusing on insider advantages. J Econ Psychol 18(2–3):235–257
Hanson R (2003) Combinatorial information market design. Inf Sys Front 5(1):107–119
Hanson R (2006) 'Foul play in information markets'. In: Hahn RW, Tetlock PC (ed.) Information markets: a new way of making decisions. pp. 126–141. Washington D.C.: *AEI Press*. Incollection George Foul Play in Information Markets
Hanson R, Oprea R (2007) A Manipulator Can Aid Prediction Market Accuracy. Economica 76:304–314
Hanson R, Oprea R, Porter D (2006) Information Aggregation and Manipulation in an Experimental Market. J Econ Behav Organ 60(4):449–459

Hayek FA (1945) The use of knowledge in society. Am Econ Rev 35(4):519–530
Holt CA, Laury SK (2000) Risk aversion and incentive effects. Am Econ Rev 92:1644–1655
Jian L, Sami R (2012) Aggregation and manipulation in prediction markets: effects of trading mechanism and information distribution. Manage Sci 58(1):123–140
Kirchler E, Maciejovsky B (2002) Simultaneous over- and underconfidence: evidence from experimental asset markets. J Risk Uncertain 25(1):65–85
Kittur A, Kraut RE (2008) Harnessing the wisdom of crowds in wikipedia: quality through coordination. Proceedings of the 2008 ACM conference on Computer supported cooperative work (CSCW '08). ACM, New York, pp 37–46
Ledyard J, Hanson R, Ishikida T (2009) An experimental test of combinatorial information markets. J Econ Behav Organ 69(2):182–189
Leitch G, Tanner J (1991) Economic forecast evaluation: profits versus the conventional error measures. Am Econ Rev 81(3):580–590
Lo AW (2007) 'Efficient markets hypothesis'. The new Palgrave: a dictionary of economics 2
Luckner S, Schroeder J, Slamka C (2008), 'On the forecast accuracy of sports prediction markets'. In: Gimpel H, Jennings NR, Kersten G, Ockenfels A, Weinhardt C (eds) "Negotiation, auctions and market engineering" 1, 227–234
Manski CF (2006) Interpreting the predictions of prediction markets. Econ Lett 91:425–429
Mbemap M (2004) Economic derivatives and the art and science of investment. J Deriv Account 1:5–8
McNees SK (1992), 'How large are economic forecast errors?', New England Economic review(Jul), 25–42
Milgrom P, Stokey N (1982) 'Information, trade, and common knowledge'. Journal of Economic Theory 26(1)
Neumann D (2004) 'Market Engineering: A structured design process for electronic markets', PhD thesis, Karlsruhe
Oliven K, Rietz TA (2004) Suckers are born but markets are made: individual rationality, arbitrage, and market efficiency on an electronic futures market. Manage Sci 50(3):336–351
Oller L-E, Barot B (2000) The accuracy of European growth and inflation forecasts. Int J Forecast 16(3):293–315
Osterloh S (2008) Accuracy and properties of german business cycle forecasts. Appl Econ Quart 54(1):27–57 (formerly: Konjunkturpolitik)
Prakash, Loungani (2001) How accurate are private sector forecasts? Cross-country evidence from consensus forecasts of output growth. Int J Forecast 17(3):419–432
Roll R (1984) Orange juice and weather. Am Econ Rev 74(5):861–880
Roth EA (2009) What have we learned from market design? Innovation policy and the economy, Vol 9. University of Chicago Press, 79–112
Rothschild D (2009) Forecasting Elections Comparing Prediction Markets, Polls, and Their Biases. Public Opin Quart 73(5):895–916
Schuh S (2001) 'An evaluation of recent macroeconomic forecast errors', New England Economic Review, 35–56
Servan-Schreiber E, Wolfers J, Pennock DM, Galebach B (2004) Prediction markets: does money matter? Electron Markets 14(3):243–251
Seuken S, Parkes DC, Horvitz E, Jain K, Czerwinski M, Tan D (2012) 'Market user interface design'. In Proceedings of the 13th ACM Conference on Electronic Commerce (pp. 898–915). ACM
Shiller RJ (1993), Macro markets: creating institutions for managing society's largest economic risks/ Robert J. Shiller, Clarendon, Oxford
Snowberg E, Wolfers J (2005) 'Explaining the favorite-longshot bias: is it risk-love, or misperceptions?', University of Pennsylvania, mimeo
Snowberg E, Wolfers J, Zitzewitz E (2007) Partisan impacts on the economy: evidence from prediction markets and close elections. Q J Econ 122(2):807–829
Sonnemann U, Camerer CF, Fox CR, Langer T (2008) 'Partition-dependent framing effects in lab and field prediction markets', working paper
Spann M, Skiera B (2004) Einsatzmöglichkeiten virtueller Börsen in der Marktforschung. Zeitschrift für Betriebswirtschaft 74:25–48
Stathel S, Luckner S, Teschner F, Weinhardt C, Reeson A, Whitten S (2009), 'AKX - An exchange for predicting water dam levels in Australia', in Proceedings of the 4th International Symposium on Information Technologies in Environmental Engineering, pp. 78-90

Subrahmanyam A (1991) Risk aversion, market liquidity, and price efficiency. Rev Financ Stud 4(3):417–441

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. Doubleday, New York

Teschner F, Weinhardt C (2012) Evaluating hidden market design'. auctions, market mechanisms, and their applications. Springer, New York, pp 5–17

Teschner F, Mazarakis A, Riordan R, Weinhardt C (2011) "Participation, feedback & incentives in a competitive forecasting community," Proceedings of the International Conference on Information Systems (ICIS), Shanghai, China, Paper 16, 1–14

Thaler RH, Sunstein CR, Balz JP (2010) 'Choice architecture', working paper

Thomason, D. (2012). "Should presidential campaigns spend more money manipulating intrade?" The Atlantic, October 23, 2012

Tversky A, Kahneman D (1974) Judgment under Uncertainty: heuristics and Biases. Science 185(4157):1124–1131

Tziralis G, Tatsiopoulos I (2007) Prediction markets: an extended literature review. J Predict Markets 1(1):75–91

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

Wärneryd KE (1996) Risk attitudes and risky behavior. J Econ Psychol 17(6):749–770

Weinhardt C, Gimpel H (2006) Market engineering: an interdisciplinary research challenge. In: Jennings N, Kersten G, Ockenfels A, Weinhardt C (eds) Dagstuhl seminar proceedings 06461 negotiation and market engineering. Internationales Begegnungs- und Forschungszentrum (IBFI), Schloss Dagstuhl, Germany

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

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

Wolfers J, Zitzewitz E (2004) Prediction Markets. J Econ Perspect 18(2):107–126

Wolfers J, Zitzewitz E (2006a) Five open questions about prediction markets. In: Hahn Robert W, Tetlock Paul C (eds) Information markets: a new way of making decisions. AEI Press, Washington D.C., pp 13–36

Wolfers J, Zitzewitz E (2006b) 'Interpreting prediction market prices as probabilities', Technical report, Federal Reserve Bank of San Francisco