The Journal of Prediction Markets 2012 Vol 6 No 3 pp 27-41

SIMPLIFYING MARKET ACCESS: A NEW CONFIDENCE-BASED INTERFACE

Florian Teschner
teschner@kit.edu
Institute of Information Systems and Management, Karlsruhe Institute of Technology

David Rothschild
David@ResearchDMR.com
Yahoo! Research

ABSTRACT

Markets are a strong instrument for aggregating dispersed information, yet there are flaws. Markets are too complex for some users, they fail to capture massive amounts of their users' relevant information, and they suffer from some individual-level biases. Based on recent research in polling environments, we design a new market interface that captures both a participant's point estimate and confidence. The new interface lowers the barrier to entry, asks market's implicit question more directly, and helps reduce known biases. We further utilize a novel market rule that supplements the interface with its simplicity. Thus, we find that market participants using our new interface: provide meaningful information and are more likely to submit profitable orders than using a standard market interface.

1 INTRODUCTION

We test a new web-based (wizard) interface in this article. The wizard captures the users' confidence range at self-selected intervals. Participants first set a range for the answer then provide a confidence level for that range. We utilize the mid-point of the confidence range as a point estimate. We then assume normality to create a full probability distribution for each user. This information then automatically creates orders in the underlying market.

People have massive amounts of dispersed information about upcoming events, including: economic, financial, and political events. The most efficient and prominent current method of gathering individual-level information is a market where the users buy and sell contracts on the event. The contracts are set to outcome levels, and the quantity and price that a user is willing to wager is a proxy for their estimation of and confidence in that outcome.

Markets' success is a combination of many elements; for this paper we focus on three: the users, the questions, and the aggregation. First, markets attract a self-selected group of informed people that, by playing the market, provide their information. Second, markets implicitly ask the users meaningful questions and properly incentivize their answers. Third, markets aggregate this dispersed information in a unique method that is a proxy for the users' level of information or confidence in that information.

There are still major deficiencies in these three key elements of markets. First, market complexities might hinder participants with new information from participating or participating efficiently. Second, markets only collect a small sliver the of the users' information. For every trade users execute or order they place in the order book, users calculate multitudes of estimations the market fails to record. Third, due to several biases, both rational and behavioral, quantity and price is not an perfect proxy for confidence.

The new web-based interface ameliorates those three concerns, while still taking advantage of the questions, aggregation techniques, and incentives that make a market so effective. First, the interface simplifies the investment into its core information and eliminates the second step required in a market of translating expectation into efficient investment. Second, the interface collects the users' expectation on a regular basis, whether or not he ultimately invests at that moment. Third, the lowered saliency of investment part of markets moots the behavioral biases.

We test this new interface in a linear prediction market for economic outcomes. This simplified market rule enhances all of the advantages of the wizard interface.

There are three main results. First, we show that inexperienced users are more likely to use such a trading support mechanism. Secondly, we find that the method works in extracting well-calibrated individual estimates. Third, we find that participants using our trading wizard make a higher profit.

We structure the article as follows. In second section we review related work in the domains of prediction markets, information extraction and market interfaces. In the third section we explain the experimental design. In the fourth section we outline the estimation strategy and results. We conclude with a discussion in section five.

2 BACKGROUND

Prediction markets have a long track of successful application in a wide area ranging from political to sport events, sometimes outperforming established forecast methods (Berg et al., 2000; Rothschild, 2009; Luckner and Weinhardt, 2008). The accurate and real-time aggregated expectations about events facilitate and support decision making (Arrow et al., 2008; Hahn and Tetlock, 2006). The granular nature of the data allows for increasingly precise event studies of these events (Snowberg et al. 2011).

The most basic trading mechanism for prediction markets is a continuous double auction for one stock which represents the outcome of an event. The stock will pay $1 if an event has the predicted outcome and $0 if not. Market participants form expectations about the outcome 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.

Attracting users with dispersed information into the market is a key element of making a market successful; yet, responding to the challenge posed 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. (2010a) 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." Hence the goal is to lower the entrance barriers (e.g. market complexities) for non-sophisticated users to participate in markets.

There are two methods of achieving simplification: first we discuss adapting the market rules and second we discuss changes to the user interface.

The standard market rules for economic indicators are complex. In an attempt to set up a market to predict economic variables in 2002 Goldman Sachs and Deutsche Bank created the so called 'Economic Derivatives' market. It tries to predict macroeconomic outcomes such as ISM Manufacturing, change in Non-Farm Payrolls, Initial Jobless Claims and consumer price index (Gadanecz et al., 2007; Mbemap, 2004). The traded contracts are securities where payoffs are a function of macroeconomic data releases. The instruments trade as a series (between 10-20) of binary options. For example a single data release of the retail sales in April 2005 was 18 stocks. In order to maximize liquidity the market operators use 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 provides hedging opportunities against event risks and a short horizon market forecast of certain economic variables. By analyzing the forecast effeciency Gurkaynak and Wolfers (2006) find that market generated forecasts are very similar but more accurate than survey based forecasts. In an attempt to forecast inflation changes in Germany, Berlemann and Nelson (2005) set up a series of markets. The markets also feature continuous trading of binary contracts. In a similar field experiment Berlemann et al. (2005) again used this system in order to aggregate information about inflation expectations in Bulgaria. These series of binary outcome markets have some known problems such as the partition-dependence' and favorite-longshot bias (Sonnemann et al. 2008; Wolfers and Zitzewitz 2006).

Addressing these problems Teschner et al. (2011a) propose a linear market design in which one outcome is represented by a single stock. The theoretical improvements are threefold; first the number of traded stocks is reduced leading 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. The next section (experimental setting) will detail their approach. In a first evaluation they show that market forecasts perform well compared to the Bloomberg survey forecast (Teschner et al. 2011b).

The second method to achieve market simplification is to change the market interface; following the idea of hidden market design, Seuken et al. (2010b) design a market-based P2P backup application. Conducting a usability study they find that the hidden market interface activates the right mental model. By analyzing the effect of different trading interfaces on trading performance Teschner and Weinhardt (2011) empirically show that trading performance can be improved using hidden market interfaces. However, it remains unclear how to design these interfaces and how this affects trading performance.

As there are no guidelines on how to design simplified trading interfaces we rely on work on polling interfaces. Polls frequently ask laypeople and experts alike to estimate imperfectly known quantities, or outcomes. Answers to such questions often come with considerable uncertainty. Respondents express uncertainty in two ways, by providing: (a) a probability attached to the most likely estimate or (b) a range judgment (Teigen and Jorgensen, 2005). One common way to combine range and probability estimate is to ask questions for the 90% prediction interval such as “What is the 90% interval for the unemployment-rate in 2012? Please provide an answer so that the minimum limit and the maximum limit include the correct answer in 9 out of 10 times.” (e.g. Moder et al. 1995). This procedure produces subjective confidence intervals. However, it is a common finding that both laymen and experts set their prediction intervals too narrow (Soll and Klayman 2004; McKenzie et al. 2008). Researchers interpret this as overconfidence.

Our new interface, which needs to capture confidence, combines the most accepted literature on polling for confidence; from an interface design perspective, research shows that the level of overconfidence massively depends on the question format. Using frequencies vs. probabilities reduces overconfidence (Klayman et al. 2006). Asking the same estimate from one individual stimulating different trains of thought also improves accuracy (Vul and Pashler 2008). Finally, Teigen and Jorgensen 2005 also show that letting participants self-assign a confidence level to a range estimate produces well calibrated estimates.

3 EXPERIMENTAL DESIGN

In October 2009 a play money prediction market was launched specifically designed to forecast the following economic indicators in Germany: GDP, inflation, Ifo-index, investments, export, and unemployment figures. The goal of the market is gather individual-level information about the upcoming economic indicators and aggregate it into forecasts. The nature of this market allows us to create an aggregated forecast of the economic indicators earlier than the most efficient public sources and continuously.

The market called Economic Indicator Exchange (EIX) (www.eix.handelsblatt.com), 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. We thus expect the users to have no problems understanding the indicators and the concept of trading. The market is publicly available over the Internet and the newspaper actively invited their readers to join. The registration is free and requires, besides a valid email address, just minimal personal information.

The market design features a continuous double auction without a market maker. Participants can submit marketable limit orders with 0.01 increments through the web- interface. After registration participants begin with an endowment of 1,000 stocks of each contract and 100,000 play money units. Our simplification over the standard market method is to represent continuous outcomes with one stock and define a linear payout function. 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 market is able to account for changes in the range of -10% to infinity. To represent the whole outcome range from -100%, α could be set to one. Hence we propose to scale the minor changes to a certain level. Looking at historical data there were no events where 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. Economic Variables

Indicator	Unit	Data Release Cycle	Payout Function
Unemployment	Absolute level (in millions)	monthly	$100 + \frac{ABS(Number)}{100}$
GDP	percent	quarterly	$100 + \alpha\left(\frac{I_t - I_{t-1}}{I_{t-1}}\right)$
Exports	percent	monthly	$100 + \alpha\left(\frac{I_t - I_{t-1}}{I_{t-1}}\right)$
Inflation	percent (over year)	monthly	$100 + \alpha\left(\frac{I_t - I_{t-12}}{I_{t-12}}\right)$
Investments	percent	quarterly	$100 + \alpha\left(\frac{I_t - I_{t-1}}{I_{t-1}}\right)$
Ifo Index	absolute level	monthly	$100 + \alpha(I_t - I_{t-1})$

Table 1 summarizes the economic variables tradable on the market. Due to the payout function and the selection of the corresponding units it is safe for users to expect stock prices to fall roughly between 50 and 150.

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. To facilitate longer forecast horizons every indicator has three independent stocks each representing the next three data releases: $\mathbf{t_1}, \mathbf{t_2}, \mathbf{t_3},$. As a consequence, the initial forecast periods vary from 1 month for monthly released indicators up to 3 quarters for quarterly released variables. We halt trading for any stock on day prior to the release of the associated economic indicator. Finally, after the announcement of the economic indicator, all stocks liquate according the payout function noted in Table 1.

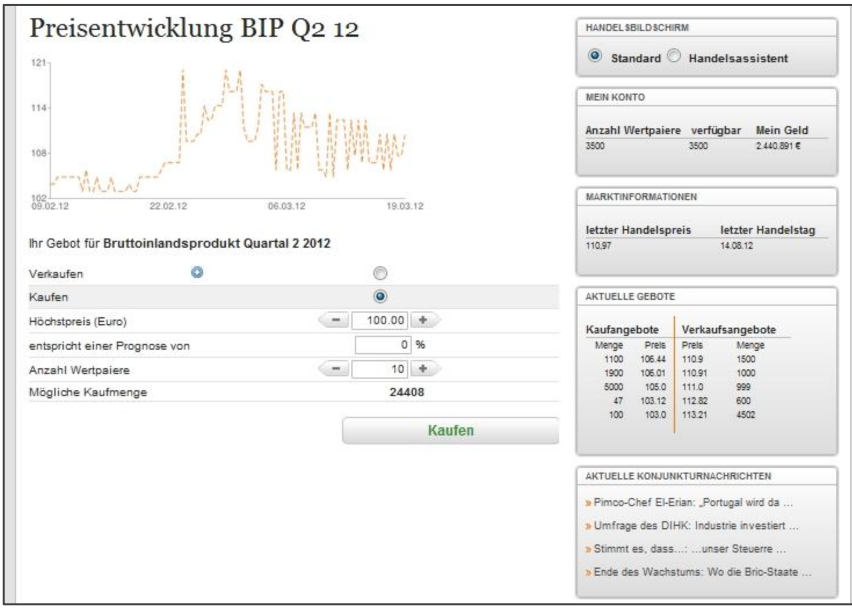
As soon as the trading in one stock stops a new stock of the same indicator (i.e., $\mathbf{t_4}$) begins trading. This means that participants received 1000 new stocks of the respective indicator. All in all participants are able to continuously trade 18 stocks at all times.

In order to incentivize users to provide as meaningful information as they have, we use two methods. First, we have two interface features that breed competition: traders can follow their performance on a leader board and they can form groups with others to spur competition with their social network. Previous research in the field of prediction markets has shown that play-money perform as well as real-money markets predicting future events (Wolfers and Zitzewitz 2004; Servan-Schreiber et al. 2004). Second, due to the legal restrictions on gambling the EIX prediction market has to rely on play money. But, to increase participants' motivation and to provide incentives to truly reveal information we hand out prizes worth 12,000 Euro. 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 3 yearly prizes (total value 4,000 Euro) are handed out according to the portfolio ranking at the end of the market. The monthly prizes are shuttled among participants who fulfilled two requirements for the respected month: (i) they increased their portfolio value and (ii) they actively participated by submitting at least five orders. Both incentives are clearly communicated through the 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.

Standard Market Interface: In the standard trading screen (Figure 1), participants have convenient access to the order book with 5 levels of visible order book depth, the price chart, the account information and market information such as the last trading day. As additional information the Handelsblatt provides access to an up-to-date economic news-stream. Moreover, trading screen shows a short description of the market comprising the respective payout function.

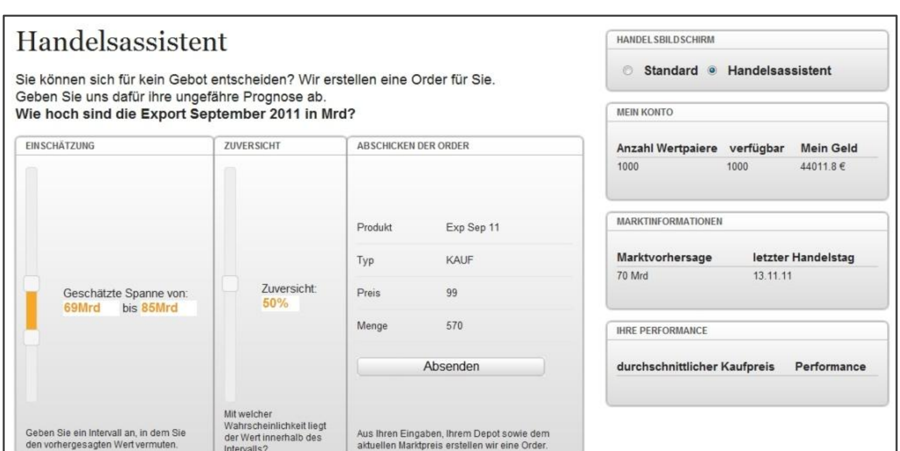
Additional to the standard trading interface, participants have the choice to switch to a trading wizard guiding their trading decisions.

Figure 1. Standard Market Interface



Experimental Wizard Interface: We build on the most recent methods of surveying expectations to create a graphical, interactive web-based interface that gathers individual forecasts: range-estimates and confidence. We employ a polling mechniasm based on Teigen and Jorgensen, (2005) in a market setting. Figure 2 displays the setup.

Figure 2. Experimental wizard interface.



The interface is 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 the user states 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. It is noteworthy that the wizard provides far less information than the standard interface.

As this experiment is part of a prediction market the user input generates an order. In the previously described process the participant provides three input parameters; a lower forecast bound (LB), an upper forecast bound (UB), and a confidence estimate (C) on the probability that the outcome is within the range. From this input, we calculate the point estimate (PE) as the mean of lower and upper bound. Implicitly the user defined the area between LB and UB with the confidence measure (C); assuming normal distribution of beliefs, we can use this area to calculate sigma. Mu equals the point estimate (PE). Hence, the user input specifies a full probability distribution of the outcome as $N(\mu, \sigma2)$.

Given that we can ex-ante determine a transaction price (TP), it follows that PE >TP (PE <TP) results in a buy (sell) order. Moreover, a rational trader would buy (sell) up to $PE - \varepsilon = TP$ ($PE + \varepsilon = TP$). However taking risk-aversion into account, the order size should be inverse to the expected probability of a loss. The higher the distance between TP and PE the lower is probability of a loss. Hence, the closer the transaction price is to the point estimate, the smaller should be the submitted order size. We express this relationship as a percentage of the possible investment; I = 2*probability(TP, μ, $\sigma2$). The highest probability of a loss is 0.5 and occurs in the case that TP equals PE. Therefore, we multiply the probability with the factor 2. The order size is then calculated as; order size = I * available cash amount. As participants might want to invest in more than one stock we operationalize this by limiting the maximum order size to 5,000.

4 ESTIMATION STRATEGY & RESULTS

The following data covers the time span from October 1, 2011 until March 5, 2012. In total 1,359 participants registered for the EIX market.

We only study stocks which have been paid out, which means we can rate all orders depending on their performance. Altogether participants submitted 14,480 orders resulting in 5,215 executed transactions. In the respected time frame 19 stocks were paid out. On signup we randomly set the default trading screen to either the standard or wizard trading interface, but we do not restrict users from crossing back over.

We are able to draw significance across all 19 stocks, as the payout function provided remarkable consistency across all six indicators. Table 2 shows the standard deviations are well within the same magnitude and the trades are reasonably distributed across indicators.

Table 2. Comparable Payouts Across Indicators

Indicator	Mean Payoff Value	SD of Payoff Value	Trades
Unemployment	129.6	1.7	2,117
GDP	98.6	2.0	911
Exports	118.5	1.8	1,200
Inflation	121.6	1.4	1,466
Investments	110.8	0.7	1,021
IFO Index	108.5	1.1	1,261

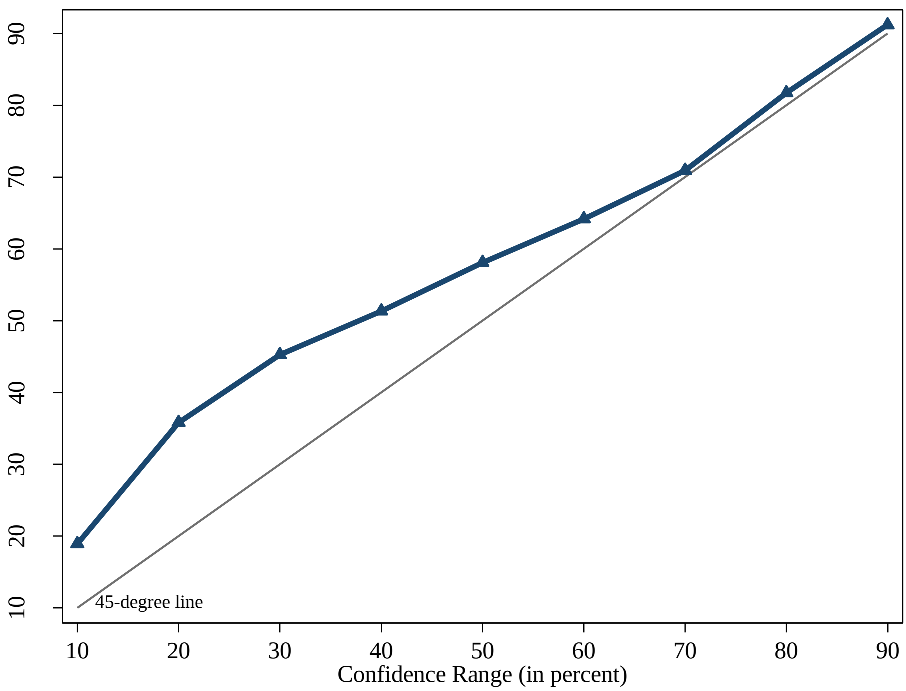
Users of the experimental wizard interface provide well calibrated confidence levels. In order to quantify the level over/under-confidence, we readjust the user provided ranges and confidence levels to a certain confidence level. This procedure follows Speirs-Bridge et al. (2010) who, for comparison reasons, re-adjusted all responses to the 80% level. We readjust all responses to the 10%-90% intervals to see how well calibrated they are at any given range (e.g., the 20% range is the value between 40% and 60% in the respondent's probability distribution). In Figure 3, the dark blue line with the triangles plots the quantity of outcomes that hit in a given range. If participants are well calibrated the outcome should follow the 45 degree line (e.g., 20% of outcomes should occur within the middle 20% range). Participants are slightly under-confident in the small ranges (e.g., about 35% of answers fall in the middle 20% range). Yet, as the ranges move up to 50% and larger they are extremely well calibrated (e.g., 81.8% of answers fall within the 80% range).

The confidence of the users, the inverse of the variance, is inversely proportional to absolute error of the point-estimate and outcome. The users have not only calibrated their confidence well, but they have provided accuracy with it.

This confirms our use of the confidence in the heuristic that determined the users' orders. The theory was that the higher the sigma of the standard distribution the user submits the lower the precision of the user's estimation. We test this assumption using a few versions of the following regression:

$$\mathbf{AE} = \boldsymbol{\alpha} + \mathbf{Sigma} + \mathbf{Decision\ Time} + \mathbf{Number\ of\ Orders} + \sum_{\mathbf{i=1}}^{\mathbf{5}} \mathbf{M_i}$$

where sigma is derived from the users' imputed probability distribution, decision time is time a user takes to submit that order and the number of order a user has made is a proxy for experience level.



Finally, we include dummy variables for each indicator to control for varying levels of underlying uncertainty. We showed in Table 2 that this should not be a meaningful factor. While individual dummies are occasionally statistically significant, we do not show them in the table. 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, so we show clustered standard errors by user. 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.

As expected we find that a higher sigma predicts a higher absolute error. Hence, the rule we use to create orders from the wizard correctly reduces the order size with increasing sigma. This result holds with clustered errors in column II. Further, column III shows that this result also holds within users, not just between users. This coefficient of around 0.06 is not just significant, but meaningful. The standard deviation of sigma is 14.2; a one standard deviation movement in sigma correlates to a change of the absolute error of 0.85. The mean standard error is 4.6.

The other two obvious variables that could have indicated the accuracy of the user's estimates, decision time and experience, do not provide any additional identification. Column IV shows that these variables are small and insignificant in a simple Fair-Shiller (1989 and 1990) regression.

Table 3. Predicting the Absolute Individual-level Forecast Error: $\boldsymbol{AE = \alpha + Sigma + Decision\,Time + Num\,Order + \sum_{i=1}^{5} M_i}$

	I	II	III	IV
Sigma	0.063** (0.025)	0.063* (0.035)	0.058* (0.034)	0.062** (0.025)
Decision Time				-0.003 (0.003)
Number of Orders				-0.005 (0.004)
Clustered Standard Error by user		x		
User Fixed Effects			x	

Notes: ** denotes statistically significant coefficients at the 5%, * at the 10% level. (Standard errors in parentheses).

Participants provide more meaningful information in the web-based interface version than the standard market version. One way to test this is to see if market participants using the trading wizard perform better than participants using the default trading interface. The following regression model quantifies the effect:

$$\mathbf{Profit = \alpha + Interface + Decision\ Time + Num\ Order + \sum_{i=1}^{5} M_i}$$

We measure the effect of the choosing the interface on the profit on an order by order basis. As before, we control for decision time, experience and indicator type.

Table 4. Profits Depending on the Interface (Interface =1 if the polling interface is used): $\boldsymbol{Profit = \alpha + Interface + Decision\,Time + Num\,Order + \sum_{i=1}^{5} M_i}$

	I	II	IV
Interface	31,688** (14,029)	31,688 (22,151)	32,050** (14,083)
Decision Time			101** (32)
Number of Orders			7.2 (5.3)
Clustered Standard Error by user		x	

Notes: ** denotes statistically significant coefficients at the 5% level. (Standard errors in parentheses).

We see that the participants using the interface have a higher profit per order, an average of 31,688 currency units. This holds in both the clustered error regression and the regression with decision time and number of orders. While number of orders still proves to be a statically insignificant variable, decision time correlates with more profitable decisions when both the wizard users and standard users are included. We do not have meaningful identification on users moving between interfaces, just four users switched, so we do not include a fixed-effect model.

The result that users of the wizard make more profitable trades is likely downwardly biased as the users of the interface are much less knowledgeable than their counterparts in the market. In an ideal setting we would have randomly placed users in the two different settings, which we did, and forced them to stay trading in those settings, which we could not do. Yet, we do know that the ultimate users of the new interface are less knowledgeable about macroeconomics and markets than the users of the standard market. On signup participants are asked to rate their economic knowledge and their experience with markets. Table 5 illustrates that the average participants using the interface rather than the market rate themselves lower on both scales.

Table 5. Percent of Users with "Good Knowledge" in a Given Category

	Standard	Wizard	Difference
Economy	68.5%	45.2%	23.3*
Market	65.4%	36.5%	28.9*

5 DISCUSSION

Markets are a strong instrument for aggregating dispersed information, yet there are flaws. Markets are too complex for some users, they fail to capture massive amounts of their users' relevant information, and they suffer from some individual-level biases.

Based upon recent polling research, we build a new web-interface for markets that is a more efficient method of gathering individual-level information than the currently utilized methods. The web-interface capture's the users' confidence range at self-selected intervals. Adding such a wizard to our market system has three advantages. It lowers the barrier to entry, asks market's implicit question more directly, and might reduce known biases such as non-risk-neutrality. We find that market participants using our new interface: provide meaningful information and are more likely to submit profitable orders than using a standard market interface.

This research adds to three separate research streams. First, from a polling perspective, we add to previous work by showing that overconfidence in range estimates is an artifact of the question format and not a participant characteristic. By letting participants self-select their confidence in their range participants exhibit slight under-confidence, rather than overconfidence. Second, this experiment is a test of the hidden market paradigm, which states that reducing the default market interface can improve participation and efficiency. Our wizard interface is very much in line with this idea and hides most market features. Comparing profits between the standard and the wizard interface we find that participants using the wizard gain more profit. Finally, for prediction market domain this work suggests to take the interface into account when designing and implementing prediction markets. Moreover, prediction markets should provide alternative and simplified trading interfaces for inexperienced users.

Further research will expand on this wizard interface design and how the additional information can be utilized. It seems reasonable to aggregate the confidence forecasts to provide an additional benefit of indicating forecast uncertainty. Furthermore, as the wizard extracts a full probability distribution, one could imagine a similar system in a serie of binary outcome markets in which one user input could trigger trade in multiple contracts at once.

6 REFERENCES

Armstrong, J. S. (2008). “Combining forecasts.” in Principles of Forecasting: A Handbook for researchers and practitioners.

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.

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

Berlemann, M., K. Dimitrova, and N. Nenovsky (2005). “Assessing market expectations on exchange rates and inflation: A pilot forecasting system for bulgaria.” SSRN eLibrary.

Berlemann, M. and F. Nelson (2005). “Forecasting inflation via experimental stock markets some results from pilot markets.” Ifo Working Papers No. 10, Ifo Institute for Economic Research at the University of Munich.

Fair, Ray, and Robert Shiller, 1989. “The Informational Content of ex-Ante Forecasts.” *Review of Economics and Statistics* 71(2):325-31.

-----------, 1990. “Comparing Information in Forecasts From Econometric Models.” *American Economic Review* 80(3):375-89.

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

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 (2006). Introduction to information markets. In R.W. Hahn, & P. C. Tetlock (Eds.), Information markets: A new way of making decisions in the public and private sectors (pp. 1−12). Washington DC: AEI-Brookings Joint Center.

Klayman, J.; Soll, J. B.; González-Vallejo, C. and Barlas, S. (1999). 'Overconfidence: It Depends on How, What, and Whom You Ask'. *Organizational Behavior and Human Decision Processes* 79(3).216 - 247.

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).

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

McKenzie, C.; Liersch, M. & Yaniv, I. (2008). 'Overconfidence in interval estimates: What does expertise buy you?', *Organizational Behavior and Human Decision Processes* 107(2), 179--191.

Moder, J. J., Phillips, C. R. and Davis, E. W (1995). Project management with CPM, PERT and precedence diagramming, Wisconsin: Blitz Publishing Company.

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

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

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

Rothschild, D. (2009), 'Forecasting Elections', Public Opinion Quarterly 73(5), 895--916.

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

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.; Jain, K. and Parkes, D. C. (2010). Hidden Market Design, in 'Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI'10)'. pp. 1498-1503.

Seuken, S.; Tang, J. and Parkes, D. C. (2010). Accounting Mechanisms for Distributed Work Systems, in 'Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI'10)'. pp. 860-866.

Shiller, R. J. (1993). Macro markets: creating institutions for managing society's largest economic risks (Clarendon Lectures in Economics).Oxford University Press.

Soll, J. and Klayman, J. (2004). 'Overconfidence in interval estimates.'. Journal of Experimental Psychology: *Learning, Memory, and Cognition* 30(2), 299.

Sonnemann, U.; Camerer, C. F.; Fox, C. R.; and Langer, T. (2008) 'Partitiondependent framing effects in lab and field prediction markets', SSRN eLibrary.

Snowberg, E.; Wolfers, J.; Zitzewitz, E. (2011) 'How Prediction Markets Can Save Event Studies', NBER Working Paper No. 16949.

Speirs-Bridge, A.; Fidler, F.; McBride, M.; Flander, L.; Cumming, G. & Burgman, M. (2010). 'Reducing Overconfidence in the Interval Judgments of Experts'. *Risk Analysis* 30(3). 512--523.

Teigen, K. H. and JØrgensen, M. (2005). 'When 90% confidence intervals are 50% certain: on the credibility of credible intervals'. *Applied Cognitive Psychology* 19(4). 455--475.

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

Teschner, F.; Mazarakis, A.; Riordan, R. & Weinhardt, C. (2011). Participation, Feedback & Incentives in a Competitive Forecasting Community. in 'Proceedings of the International Conference on Information Systems (ICIS)', pp. 1-14.

Vul, E. and Pashler, H. (2008). 'Measuring the Crowd Within'. *Psychological Science* **19**(7),.645.

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). “Prediction markets.” *Journal of Economic Perspectives* 18(2), 107–126.

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