COMBINING PREDICTION MARKETS AND SURVEYS: AN EXPERIMENTAL STUDY

Research-in-Progress

Kranz, Tobias T., Karlsruhe Institute of Technology, Karlsruhe, Germany, kranz@kit.edu
Teschner, Florian, Karlsruhe Institute of Technology, Karlsruhe, Germany, teschner@kit.edu
Weinhardt, Christof, Karlsruhe Institute of Technology, Karlsruhe, Germany, weinhardt@kit.edu

Abstract

The track record of prediction markets suggests that markets may help to better foresee future developments and trends. However, looking at the range of applications, it becomes clear that there are certain limits. Complex forecasts, such as conditional or qualitative judgments are better gathered with traditional forecast methods such as survey-systems. In this study we integrate a prediction market with a survey in two distinct ways. First, trading in prediction markets indicates that participants believe to have additional information. Consequently, we randomly trigger a one-question survey after a trade. Second, participants might find it convenient to answer survey questions right on the same platform. Hence, we integrated a survey feature in a prediction market for 2013 German election. In a large scale field experiment with over 3,500 participants, we test consistency and the response rate of the integrated survey compared to a standalone version. We find that the integrated survey delivers robust responses and a 3.16 % (533.33 % relative) higher response rate. Moreover, the response rate in the trigger-based survey is above 90 %. These results highlight the great potential of integrating surveys with prediction markets.

Keywords: Crowd sourcing, prediction markets, survey design.

1 Introduction

Accurate and reliable forecasts of future short- and long-term events are a crucial competitive factor for companies, regions, and countries and an important foundation for political decision making. Advances in information systems are changing information aggregation in many contexts: political institutions increasingly open up for grassroots feedback and open discussion of societal innovation, ad-hoc communities use social media to coordinate, and companies gradually shift decisions towards a broad basis of employees and allow for user-driven innovation. An underlying theme of this trend is using the collective intelligence and wisdom of the crowd.

There are various ways to utilize the wisdom of crowds or collective intelligence such as using wikis, reputation systems, or polling mechanisms. Another way to aggregate dispersed information is by using a prediction market. In these markets, participants trade contracts whose payoff depends on the outcome of uncertain future events. For example, a market contract might reward one 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 the probability of the event in question and they sell a stock if prices overestimate the probability of an event. The track record of prediction markets suggests that markets may help to better foresee future developments and trends. Although, prediction markets have their strengths in quantitative predictions and even make conditional predictions possible – albeit complicated – (cf. Berg & Rietz, 2003), they are not well suited when it comes to qualitative predictions. The strength of prediction markets is the collaborative valuation of given contracts (i.e. the mapping between payoff, event outcome, and event date). Since all valuation is based on quantitative values (i.e. prices), qualitative information can only be induced into the market in the contract design, which falls to the market operator. For example, questions like “What will the GDP of country A be in 2015?” fit perfectly to prediction markets, whereas “How can we improve productivity?” is better suited for surveys.

Complex forecasts, such as conditional or qualitative judgments are better gathered with traditional forecast methods such as survey-systems. However, traditional survey-systems also have some known drawbacks. First of all, the success of surveys largely depends on the participant selection (Ammon, 2009; Gordon, 2007). The most common selection criterion is reputation, which is based on perceived expertise. However, Tetlock (2005) shows that perceived expertise does not correlate with individual forecast accuracy. The second drawback is the decreasing participant motivation over the study’s course. The long, rigid and tedious process leads to decreasing participant numbers (Cuhls, 2003). Ilieva (2002) conducted a literature review and found response rates for online surveys from low as 6 % (Ranchhdh & Zhou, 2001) to high as 67 % (Kiesler & Sproul, 1986). Deutsksens et al. (2004) conducted a study with different types of surveys. Evans & Mathur (2005) analysed the pros and cons of online surveys in contrast to traditional mail surveys and discussed the online surveys’ best uses. Inter alia, he found online surveys are best to use, if timeline is vital, strong methodological control is sought (e.g. order of the questions), and survey research is conducted frequently; all which applies for repeated online surveys in a prediction market context.

There are at least two ways surveys can benefit from an accompanying prediction market; motivation and pre-selection of experts. Prediction markets motivate participants to contribute continuously through incentives and by providing constant feedback; both on the aggregate and the individual level. More participants might be willing to participate (at least partly) in a survey if they have indicated that they have information regarding a topic. The question is how to figure out when a participant has information about a topic? This can be detected through the prediction market. If participants change the market price, they most likely have information about a certain topic and might be willing to fill out some related qualitative and possibly more complex questions. In previous studies (e.g. Chen et al. 2005), surveys run in parallel to prediction markets but in separate system. Survey participants had to leave the known platform, fill out a survey, and return to continue trading. Potentially, participants might find it convenient to answer survey questions right on the same platform. Moreover, Teschner et al. (2011a) show that individual forecast input can be measured and objectively evaluated in prediction markets. Hence, this might help to pre-select experts not based on their reputation but on their previous forecast performance.

2 Prediction Markets and Surveys

Prediction markets offer a number of advantages over surveys. 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 like 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., 2011b). 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 they 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 (e.g. Berg et al., 2008; Ledyard et al., 2009; Bennouri et al., 2011). Although, other evidence suggests that the relative performance advantage of markets may be small compared to surveys or polls (e.g. Goel et al., 2010; Erikson & Wlezien, 2008; Rothschild, 2009). Prediction markets have a long track of successful field applications, e.g., in political elections (Berg et al. 2008), sport events (Luckner & Weinhardt 2008), finance (Bennouri et al. 2011), and predicting market development (Spann & Skiera, 2003). See Wolfers and Zitzewitz (2004) and Ledyard et al. (2009) for reviews. However, to our best knowledge prediction markets and surveys have never been combined during the prediction making process.

3 Research Question

The integration opens up several research questions. First, how well (measured by response rate, total number of answered questions, response speed) does an integrated survey work compared to a standalone version? Surveys often have a quite low response rate of about 10% (e.g. Ranchhod & Zhou, 2001; Deutsksens et al., 2004). Finding a way to increase the response rate would be highly beneficial. In some applications such as ad-hoc questionnaires regarding recent events, researchers aim for a fast response speed. Moreover, how well does the trigger-based survey tool, polling participants who are stating that they have new information perform? Secondly, previous work shows that in prediction markets experts can be identify ex post by their performance. This raises the question of when to ask the experts? In this paper, we try to address the question of how to best acquire information from the market’s experts. Additionally, online surveys are sovereign in response speed and the ability to methodologically control the filling process. According to Evans & Mathur (2005), online surveys fit best for our purpose. Hence, we decided to compare different types of online surveys. Specifically, we use common standalone surveys and market integrated surveys. Our main research question is:

RQ: Are integrated surveys more accepted by participants of a prediction market than standalone surveys?

We measure the ‘acceptance’ by response rate, total number of answered questions, and response speed. Additionally, we run the first experimental study on trigger-based surveys.

4 Study Design

For our study we use a German political stock market during the 2013 federal election called PIX (short for ‘Political Indicator Exchange’). The German voting system in which each voter has two votes is rather complex. It makes use of proportional representation (PR) and method of majority
Combining Prediction Markets and Surveys

decision (MD) and works roughly like this: the Erststimme (first vote) determines in each constituency which delegate is send to the Bundestag (parliament) using MD, the Zweistimme (second vote) uses PR to determine the share of seats each party achieves. After the election, usually two parties form a coalition in the weeks following. Afterwards, the Bundestag (i.e. all its delegates) elects the chancellor. The market predicts the election outcome continuously from 2013/01/23 until the election on 2013/09/22. The design follows previous prediction market designs for political elections (Berg et al., 2008; Forsythe et al., 1992; Berg & Rietz, 2006; Berlemann & Schmidt, 2001). We use play-money named PIX € for trading. Specifically, we run two markets in order to gather predictions: The candidate market is comprised of four Winner-takes-all contracts belonging to chancellor candidates Merkel, Steinbrück, Green candidate (unnamed) and rest-of-field. The party market contains eight different Index contracts, for CDU/CSU, SPD, FDP, Grüne, DIE LINKE, Piraten, AfD and rest-of-field. (Contract AfD was introduced on 2013/04/30, as it seemed no longer appropriate to keep it included in the rest-of-field contract, since the rest-of-field contract rose over 10 %.) The best performing traders by portfolio value win prizes after the market closes. To account for percentage values, Winner-takes-all contracts pay 100 PIXEL if the respective candidate becomes the next German chancellor, and Index contracts pay their respective parties’ election result percentage in PIXEL.

Registration is free, but every person is only allowed one trading account. Upon registration, traders receive an initial endowment of 100,000 PIXEL and 1,000 stocks of each contract. Since the contracts in each market are interdependent due to their different underlying events, this constraint dictates that prices should sum up to 100 PIXEL (which equals 100 %). To easily enforce this, traders have the possibility to buy and sell the unit portfolio, consisting of one of each contract in a market, for 100 PIXEL. Therefore, if the sum of best bids equals up to over 100 PIXEL, or the sum of all best asks equals up to less than 100 PIXEL, there is opportunity for arbitrage. Participants on the PIX can submit limit orders continuously. Short sales are not allowed. As limit orders with an extreme price can be used, there is no need for market orders. Orders are matched continuously according to the order precedence rule. The five best bids and asks for each contract are displayed in an orderbook. In order to compare the integrated to standalone surveys we set up two treatment groups (integrated vs. standalone). We randomized all traders in one of the groups. We created an extensive questionnaire with 73 items consisting of 6 parts: 1) General questions and platform feedback, 2) Election outcome, 3) Information sources, 4) Personal questions, 5) Political coalitions, and 6) Election Polls. While the integrated group could answer the questions one by one directly on the platform the standalone group was presented with a link to an external survey software (LimeSurvey) where participants had to answer all questions in one pass (see Figure 1). The integrated questionnaire allows for questions with a predefined list of answers to appear on the PIX’s main page, one at a time. Adjustable settings include the question mode, allowing only one reply selection or multiple. Equally important, we can define precisely when, where and in what order questions are asked. These properties are defined by a time window for question activity, and a priority ranking for the order of appearance. Obviously, we make sure that every user can answer every question only once. For each of our questions, we provide the option “prefer not to say”, due to the sensitive nature of the questions. Users submit their responses by selecting the radio button or checkbox and clicking submit. We use a self-made questionnaire infrastructure because this way, traders can quickly reply to a question or two without leaving the website. Hence, we do not lose willing respondents who simply not want to go to an external website.
Combining Prediction Markets and Surveys

Additionally, we ask every participant an ‘ad hoc post-trade question’ (trigger-based survey) using a self-made trigger-based survey-system. In order to gain a better understanding of the trader’s thought processes, after every submitted order the trigger-based survey-system was called in order to decide, if the current participant should be asked the trigger-based survey. The algorithm used determines on the basis of the participant’s trading history, number of already answered ad hoc post-trade questions in the last 21 days and a random component, if the question will show up immediately after the submission of an order or not. We configured the parameters in order to ask each trader at least once in three weeks per tradable product. The trigger-based survey asked the trader for the reason of his last order. Possible answers are: 1) I feel, that party/candidate was under/overvalued, 2) I can/cannot imagine to vote for that party/support that candidate, 3) I like to see a lower/higher price for that party/candidate on this prediction market, 4) I need stocks/money to sell/buy a bundle, 5) other reason. (Note, that all answer were selectable via radio-buttons; especially it was not possible to answer ‘5)’ textual.) The trigger-based survey (see Figure 2) motivates participants to rationalize their trading decision ex-post and opens an interesting area for further field-studies on decision behaviour.

5 Dataset

Participants of both treatment groups were invited to fill out either the integrated or the standalone survey from 2013/07/29 until 2013/08/25 (28 days). Both groups are nearly equally sized.
(integrated: N=1,864, N_{active}=706; standalone: N=1,731, N_{active}=657) and hardly different in their trading activity (integrated: 71.57/7/466.83, standalone: 101.78/7/685.36 (#orders mean/median/SD)). Traders are counted as „active“, if they submitted at least one order while the market was active.

6 Results

In this section we present the results of our study. First, we report differences in response rate and reaction time, followed by a closer look on the number of answered items. Last, we report response rates and present an application of the trigger-based survey.

6.1 Response Rate and Reaction Time

The integrated survey has a higher response rate; both complete responses and partial responses are higher for participants using the integrated survey. The standalone survey leads to 6 (0.91 %) complete and 38 (5.78 %) partial responses in contrast to 32 (4.53 %) complete and 124 (17.56 %) partial responses in the integrated survey. (Percentages relate to active traders in corresponding groups.) The advantage of integrated surveys over standalone surveys cannot conclusively shown here, due to the small response rates. Here, the integrated survey leads to an increase of 533.33 % in complete responses, and 326.32 % in partial responses compared to the standalone survey. Next, we compare the duration from the moment the survey was available until a participant answered his last question („reaction time“). In treatment integrated, participants that only partly answered the questionnaire have a median reaction time of 10.01 days (mean=12.21) compared to 13.26 days (mean=14.66) in treatment standalone. To completely fill the survey participants’ reaction time is on median 6.52 days (mean=7.80) in treatment integrated and 13.64 days (mean=12.48) in treatment standalone. Summing up, treatment integrated delivers completely filled questionnaires (t-stat=1.535, p=6.68 %) as well as partially filled questionnaires significantly faster (t-value=1.786, p=3.94 %).

Result 1: The integrated survey delivers results significantly faster (24.51 % – 52.20 %).

6.2 Number of Answered items

As shown, the difference between reaction times for partial responses is lower than for complete responses. As we use a rather long questionnaire, we also take a look on the number of items answered in total and per survey participant. Those measures can help to estimate participants’ acceptance of the duration of this particular survey and the possible response rate a shorter survey would have. In treatment integrated participants answered 3,522 items and 28.40 items per participant (median=11). Treatment standalone leads to (2,497) answered items and 65.71 per participant (median=64.71). Altogether, participants in treatment integrated answered on average significantly (Wilcoxon rank sum test, W=1214, p < 0.01%) less items than in treatment standalone. We assume, the reason lies in the „entry barrier“ of standalone surveys; i.e. in contrast to an integrated survey, which a participant might start, suspend, and continue as he pleases, an invitation link to a standalone survey represents a certain barrier. Participants might think twice before leaving the main web site to take part in a survey of unknown length and cognitive effort.

Result 2: The integrated survey lead to 41.05 % more answered items.

6.3 Trigger-based Survey

Last, we analyse the response rate for the trigger-based survey. A total of 3,691 questions were triggered to 699 different traders resulting in 3,388 responses from 681 traders. As we cannot
Combining Prediction Markets and Surveys

distinguish if a participant answered ‘other reason’ or denied to answer, the resulting response rate of 91.79 \% is a lower bound. Participants decided to not answer a trigger-based survey 303 times (152 unique traders). 547 participants or 78.25 \% never declined to answer a trigger-based survey.

**Result 3:**  
*Trigger-based survey: high response rate (91.79 \%) and widely accepted (78.25 \%).*

In order to illustrate how the two survey types can be used in conjunction, we will show some preliminary data matching both surveys on a per participant basis. Results of the trigger-based survey are shown in Table 1. Column ‘Answer’ lists the possible responses as described in section 4; column ‘Total’ contains all data gathered with the trigger-based survey; column ‘Party/Candidate’ shows only results of the trigger-based survey for traders who stated their preferred candidate/party via the integrated survey.

<table>
<thead>
<tr>
<th>Answer</th>
<th>Total Count</th>
<th>Total Percentage</th>
<th>Party/Candidate Preferred Count</th>
<th>Party/Candidate Preferred Percentage</th>
<th>Party/Candidate Other Count</th>
<th>Party/Candidate Other Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) under/overvaluation</td>
<td>2,458</td>
<td>66.59 %</td>
<td>206</td>
<td>62.05 %</td>
<td>1,107</td>
<td>65.35 %</td>
</tr>
<tr>
<td>2) potential support/vote</td>
<td>511</td>
<td>13.84 %</td>
<td>75</td>
<td>22.59 %</td>
<td>267</td>
<td>15.76 %</td>
</tr>
<tr>
<td>3) like to see lower/higher price</td>
<td>175</td>
<td>4.74 %</td>
<td>22</td>
<td>6.63 %</td>
<td>76</td>
<td>4.49 %</td>
</tr>
<tr>
<td>4) bundle</td>
<td>244</td>
<td>6.61 %</td>
<td>18</td>
<td>5.42 %</td>
<td>112</td>
<td>6.61 %</td>
</tr>
<tr>
<td>5) other reason/no answer</td>
<td>303</td>
<td>8.20 %</td>
<td>11</td>
<td>3.31 %</td>
<td>132</td>
<td>7.79 %</td>
</tr>
<tr>
<td>Sum (answered)</td>
<td>3,388</td>
<td>91.79 %</td>
<td>321</td>
<td>96.69 %</td>
<td>1,562</td>
<td>92.21 %</td>
</tr>
<tr>
<td>Sum (total)</td>
<td>3,691</td>
<td>100.00 %</td>
<td>332</td>
<td>100.00 %</td>
<td>1,694</td>
<td>100.00 %</td>
</tr>
</tbody>
</table>

Table 1:  
*Result of trigger-based survey (left: total; right: by traders voting preference).*

Obviously, answer 1 was the major reason for trades (66.59 \%), regardless if traders submitted an order for their preferred party/candidate (62.04 \%) or for any other party/candidate (65.35 \%). Hence, the majority of orders were reportedly submitted based on (subjectively rational) economic considerations and thus might be more ‘sensibly priced’ as orders that were submitted for reasons 2 and 3. Similar to answer 1, answer 4 indicates subjectively rational trading behaviour, as trading bundles is profit neutral. Altogether, over 80 \% of participants reported to make their decisions based on rational considerations, which seems pretty reasonable. Although, there is a considerable proportion (18.58 \%) where participants report, they want to see higher/lower prices or (not) support a certain candidate/party. At first glimpse, this seems to indicate irrational trading behaviour. Nevertheless, this answer does not strictly exclude rational consideration (e.g. one might ‘know’ a certain candidate is overvalued, but prefer answer 3 over 1 anyway.). We therefore have no reason to doubt participants answered truthfully on a large scale.

When we compare the percentual responses of all participants with those of participants that reported their party/candidate preferences, we see two salient contrasts. First, the differences between ‘Total’ and ‘Preferred’ are on average higher than for ‘Total’ and ‘Other’. Second, the biggest difference is present between ‘Total’ and ‘Preferred’ for answer 2 (8,75 \%). Both observations might indicate, that participants’ political preferences do affect their trading decisions. This might complement findings like the one that traders tend to buy more stocks of their preferred party (e.g. Kranz et al., 2014). Nevertheless, most differences between ‘Total’ and ‘Preferred’/’Other’ are rather small and –at first glance– the major tendency that roughly 2/3 chose answer 1 looks consistent.

Scratching the surface, it seems that trading behaviour matches stated behaviour in both surveys and survey responses are most widely consistent between both surveys. Summing up, the combination of integrated and trigger-based surveys provides a promising way to analyse individual trading behaviour more deeply in further research.

**Result 4:**  
The integrated survey and the trigger-based survey seem to deliver consistent results.
7 Conclusions

In the age of near-ubiquitous internet access through an expanding variety of connected devices, it has not only become possible to conduct a greater number of polls, surveys, and preference elicitation tasks that involve a greater number of people; it is also possible to obtain more and richer information from each respondent. Nowadays, a popular way to gather that information on a continuous and repeated way is to run prediction markets. Their track record suggests that these markets may help to better foresee future developments and trends. Markets are powerful instruments 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 (e.g. Wolters & Zitzewitz, 2004; Wolters & Zitzewitz, 2006).

In this study we integrate a large-scale prediction market with a survey in two distinct ways. First, trading in prediction markets indicates that participants believe to have additional information. Consequently, we randomly trigger a one-question survey after a trade. The response rate for these types of question is with 91.79% extremely high compared to typical online surveys. This approach reveals two advantages: (i) those participants who trade have information and (ii) they are actually interested to share that information. Both, having information and willingness to share information are usually out of scope of an online survey. Second, participants might find it convenient to answer survey questions right on the same platform. Hence, we integrated a survey feature. In a large-scale field experiment with over 3,500 participants, we test consistency and the response rate of the integrated survey compared to a standalone version. We find that the integrated survey delivers robust responses and a 3.16% (533.33% relative) higher response rate. Although we could not prove the higher response rate conclusively, both findings highlight the great possibilities for surveys to integrate with prediction markets. However, we have to be aware of structural differences between those survey types (e.g. participants might change their mind during an integrated survey).

Online surveys allow for qualitative responses and more complex question design. Especially given the fact that simple fill-in-the-blank and multiple choice questions give way to enhanced graphical interfaces that can capture probability distributions over response categories, even from people not familiar with distributions (Goldstein et al., 2013). Hence, combined forecasting with surveys and prediction markets can handle both; continuous, incentive-compatible forecasting as well as complex, quantitative question design.

As a direction for future research, it seems fruitful to develop adaptive survey systems that provide to ask participants only if the participant is expected to respond to it. In order to achieve that, we have to evaluate whether we are able to model the likelihood of a response. Moreover the trigger-based question, in its present form leads participants to rationalize their trading behaviour. This leads to the question whether the trigger-based questionnaire leads to a different trading behaviour. The limitations of the present work are straightforward: Most importantly, we explore one instance of an integrated system of surveys and markets in a political context. In order to increase external validity, the next step is to explore other information settings and implement surveys in prediction markets with other topics and populations.

References

Combining Prediction Markets and Surveys


Combining Prediction Markets and Surveys