Prediction Markets for Long-Term and Non-Occurring Outcome Forecasting: A Comparison of Payoff Mechanisms

Abstract

Prediction Markets (PMs) have proven to be a good forecasting instrument for elections, movie success or product sales. However, traditional PMs can only be applied to forecast short- or medium-term events, i.e. where the outcome is known and a payoff can be determined. Yet, for many practical problems, the occurrence of an event might be far in the future or not happen at all. We refer to such events as non-actual events. The literature proposes the use of different payoff mechanisms, but has never compared their accuracy. In this work, we conduct a study to compare their effectiveness. We find that, although performing slightly worse compared to traditional PMs, certain alternative payoff PMs yield competitive forecasting results. We also find that the quality of results relative to traditional PMs strongly depends on the contextual forecasting topic. Moreover, we show that by increasing the risk-aversion of the market participants, the forecasting accuracy can significantly be improved.

Keywords: prediction markets, forecasting, decision making, market design

1. Introduction

Prediction markets (PMs) have emerged as a powerful instrument for information aggregation since the late 1980ies. Various studies have demonstrated excellent forecasting performance of PMs compared to alternative instruments, e.g., in politics (Berg et al., 2001; Forsythe et al., 1992), sports (Servan-Schreiber et al., 2004; Spann and Skiera, 2008) or business (Chen and Plott, 2002; Ortner, 1998; Spann and Skiera, 2003).

The general idea of a PM is to assemble participants in an electronic marketplace where they trade virtual shares of stocks whose final values, or *payoffs*, are based on the outcomes of the stocks' underlying events. Investments in the market can either consist of real-money, which participants have to deposit prior to trading (as in the Iowa Electronic Markets (Berg et al., 2001)), or play-money, which does not correspond to any real currency and can only be used within the trading system. Both incentive schemes have shown to work comparably well (Servan-Schreiber et al., 2004; Slamka et al., 2008).

Because PMs in general rely on the determination of the final payoff, they have to-date, with the exceptions we discuss in Section 2, almost solely been applied in settings where the actual outcome of each stock is known in the short or medium term (e.g. Spann and Skiera, 2003; Wolfers and Zitzewitz, 2004). Very recently, PMs for the "long-run" have been analyzed and found to forecast well (Berg et al., 2008). However, "long-run" in this case refers to a period not more than one single year. Naturally, these limitations restrict the use of PMs. In particular, corporate-internal decision-making often involves events whose outcomes are either a) not known for a long time, such as the outcomes of strategic decisions. Or, b) the outcome will never occur. This usually happens when choosing among alternatives, where only one or more, but not all, are implemented. This for instance happens in new product development where only few of many new product ideas make it to the market.

We call such outcomes *non-actual*. Research on non-actual outcomes is scarce, yet some evidence exists that PMs are a valid tool to aggregate information in non-actual event situations. The key challenge is how to appropriately determine the payoff such that participants are properly incentivized to reveal their true expectations, despite the fact that the event may or may not occur. While there have been different suggestions in previous studies, none has established external validity, nor has any comparison been accomplished in terms of

forecasting accuracy. Thus, the goal of this paper is to first test for correlation with a benchmark and to test external validity of each payoff function. Moreover, we also test ways to improve forecast accuracy for non-actual events by analyzing risk-aversion in the markets.
In Section 2, we review existing approaches for determining non-actual event payoffs, followed by a theoretic comparison in Section 3. In Section 4, we present the design of three experiments, followed by a discussion of results in Section 5. We conclude in Section 6.

2. Approaches to Determine Payoffs for Non-Actual Events

As the payoffs cannot be determined by the outcome of the non-actual event, basically two general possibilities to determine the payoff exist. First, payoffs can be determined *internally* by using only data from the trading activity. In that case, the trading actions themselves serve as proxy for the payoffs. Second, payoffs can be determined *externally* by a proxy measure which is independent of the trading data. This measure in a way tries to replicate or approximate the "true outcome" of the respective event.
To the best of our knowledge, only six previous studies consider forecasting of non-actual events, with payoffs either determined internally or externally. While most of these studies reveal different advantages of market-based methods compared to traditional instruments, a test of external validity, i.e. a comparison of predictions with actual outcomes, were not established. Rather, evaluations were typically based on proxy measures such as conjoint studies, surveys or Delphi methods (see Table 1).
Four studies use internal measures. LaComb et al. (2007) study an "imagination market", where company-internal participants generate and evaluate business and product ideas. The final payoff was based on the *volume-weighted average price* (*vwap*) over the last 5 trading days prior to the close of the market, while the entire market was running for a total of three weeks. The predictions revealed a high correlation of market prices with the assessment of the ideas by management. On the other hand, using the *last traded price* as payoff, the studies of Dahan et al. (2007a) and Soukhoroukova and Spann (2005) test new product concepts and find high internal consistency and high correlation using independently administered surveys and conjoint studies, respectively. Moreover, the authors point to advantages of the market mechanism such as cost-effectiveness, time-consumption, a smaller bias and the need for fewer traders compared to surveys, conjoint studies, focus groups or concept tests. Dahan et al. (2007b) also use the *last traded price* (*last-price*) as payoff, however, in contrast to the former studies, they close the market at a *random point in time* (*last-price-random-close*) to avoid last minute market movements. They find a high internal and external correlation between preferences of participants, again studied via surveys.
Two recent studies use external proxy measures to determine payoff. In a study comparable to the "imagination market" (LaComb et al., 2007), Soukhoroukova et al. (2008) create an "idea market" to generate new product ideas for a high-tech company. In contrast to aforementioned studies, they base the payoffs on the assessment of a corporate-internal expert committee. Graefe & Weinhardt (2008), in a field experiment, use a Delphi study with external experts which did not participate in the markets to determine the payoffs in markets involving a group of students and a group of experts. In that sense, both of these two payoff mechanisms are based on some type of aggregated expert opinions.
Using external proxy measures, though, suffers from substantial drawbacks. First, in the case of internal experts which are known to traders, traders predict potentially biased expert decisions, rather than submitting their own assessments, especially if the experts and their possible biases are known. In the second case (external experts), uncertainty about the true values of stocks could potentially harm forecasts by not knowing any experts and their opinions. Also, experts may not always be available or may be only be able to come by at high costs. We thus do not consider payoffs based on expert opinions in our subsequent analyses.

Study	Application	Payoff of stocks based on (internal or external measure)	Results	(Theoretic) comparison to alternative instruments	Proof of external validity	Comparison of different payoff mechanisms
LaComb et al. (2007)	"Imagination market", creating and evaluating ideas	*Volume-weighted average trading price over last trading days (internal)*	- High correlation of ideas of market evaluations and management	- More ideas and more participants compared to traditional methods - immediate feedback, visibility of ideas, fun mechanism - ranking not necessarily better than with other methods	*No*	*No*
Dahan et al. (2007a)	Consumer preferences of new product concepts	*Last trading price before close of market (internal)*	- High internal consistency - high correlation with independent survey	- Cheaper, less time-consuming and less biased compared to e.g. surveys, conjoint studies, focus groups or concept tests	*No*	*No*
Soukhoroukova and Spann (2005)	Consumer preferences of new product concepts	*Last trading price before close of market (internal)*	- High internal consistency - high correlation with conjoint study	- Cheaper, need for less subjects than conjoint study	*No*	*No*
Dahan et al. (2007b)	Consumer preferences of new product concepts with high number of product features	*Last trading price before random close of market (internal)*	- High internal and external correlation with preferences	- High scalability w.r.t. number of features - engaging and fun task - but: no individual preferences	*No*	*No*
Soukhoroukova et al. (2008)	Creating and evaluating new products with a company-internal idea market	*Expert committee (external)*	- Many ideas - new ideas - original ideas	- Only method which involves large number of ideas and creators, group decisions and combination of idea creation and combination	*No*	*No*
Graefe & Weinhardt (2008)	Long-term forecasting of future trends	*Delphi study (external)*	- High correlation with Delphi study	- Delphi study	*No*	*No*

Table 1: Studies of prediction markets with non-actual outcome

3. Theoretical Comparison of Payoff Mechanisms

Wolfers & Zitzewitz (2004) cite three reasons why traditional PMs with actual outcomes as payoffs are expected to yield good result: they provide a) incentives to seek information, b) incentives for truthful information revelation, and c) an algorithm for aggregating diverse opinions. The essential idea of PMs is that the updating of existing information is *rewarded* when the new information is more accurate than existing information, but *penalized*, when the new information is less accurate. Ultimately, rewards or penalties are decided when the event occurs (or fails to do so). As a consequence, *public* as well as *private* information is incorporated into prices using this reward mechanism.

Payoff mechanism	**Actual outcome payoff**	**Alternative payoff**
Traders' performance dependent on actual outcome	yes	no
Aggregation of public/private information	public and private	public
Probability of occurrence of information cascades	rather low	rather high

Table 2: Comparison of payoff mechanisms

In contrast, the alternative payoffs described in Section 2 do not depend on the actual outcome, as the event may never occur or may only occur far in the future. Thus, traders' portfolio valuations are completely independent of a "true" state. In theory, this should change trading strategies in the markets. While in the traditional *actual-outcome* markets, traders' investment decisions are based on the expected actual outcomes, a trader in alternative payoff markets must predict the *vwap* or the *last traded price*, respectively. Here, traders are not incentivized to reveal their private information because doing so is not necessarily rewarded by the mechanism. This might lead to a form of information cascades (Bikhchandani et al., 1992), where own private information is underweighted and choices are made depending on choices made by other market participants. Although this informational inefficiency might occur in traditional PMs as well (Anderson and Holt, 1997), the effect is likely larger in alternative PMs in general. This is because one's own portfolio performance essentially depends not only on one's own assessments, but to a large extent on the assessment of others, i.e. the assessments of the majority of the "trading crowd". Summarizing, we expect actual-outcome markets to be superior in terms of prediction accuracy compared to alternative payoff markets.

While we expect actual-outcome markets to perform best overall, it is harder to assess which among the alternative payoff mechanisms will perform best. From a trader's perspective, the vwap payoff format is not very intuitive, especially for non-experienced traders. In contrast, the concept of a last price is quickly comprehensible. Another aspect about alternative payoffs is the potential of gaming the market. That is, one or more traders may try to manipulate the price to their advantage. While this risk should especially be prevalent in last-price markets, and, to the extent, in last-price-random-close markets, it has also been found in markets with vwap benchmarks (Berkowitz et al., 1988). Moreover, trading should be different in vwap markets. In these markets, with each trade that a participant executes, s/he is likely to make a loss from a myopic point of view. This is because, when e.g. buying shares, the mean price a trader pays for each share will likely be above the vwap of the considered time frame in which the vwap is calculated. Although a purchase will increase the vwap and therewith, the payoff, the price paid is very likely to be below the vwap. In order to avoid this situation, traders might expose a form of herd behavior (Smith et al., 1988), creating price bubbles as a special form of information cascades and driving prices towards a certain direction in order to make profit with their trades. Thus, extreme over- and/or under-pricing might occur. This particular phenomenon, however, should not occur in last-price (-random-close) markets. This is due to the fact that from a myopic point of view, each trade a trader executes would not harm her or his performance, as this fixed price would be the final payoff price of the stock.

Another aspect that can affect the performance of alternative payoff mechanisms is the degree of risk-aversion of the traders and their relationship to manipulation and gaming. While in actual-outcome markets, manipulation such as artificially moving stock prices up or down will eventually penalize the manipulator (Hanson et al., 2006), this is not the case in alternative payoff markets where the true event may never occur. However, a manipulator may be penalized by subsequent traders which "correct" the market price and consequently adjust the vwap or the last traded price. In contrast, manipulation and gaming should not occur if the fear of being penalized is high, i.e., if traders are risk-averse. In that case, the best prediction a trader can input in market prices is the publicly known information.

4. Experimental Design

In the following, we describe a field experiment which we conducted to compare the different alternative payoff mechanisms. The main aim of the experiments is to assess external validity for each payoff, as well as to compare the payoffs among themselves. Because we cannot assess true performance in markets with non-actual events, we base our experiment on forecasting events that *do occur*. Therewith, we are able to fully analyze the different payoffs' external validity and obtain an objective benchmark. Also, prior research has indicated that the type of topic to be predicted may have an effect on the market outcome (Rosenbloom and Notz, 2006). Moreover, different topics are likely to be different in their relation between publicly and privately held information (see Section 3). Therefore, unlike most studies which focus only on one topic, we base our analyses on three different topics to obtain validity across a broad set of issues, namely politics, sports and general economic issues.

We ran three experiments on the above three topic in the spring term of 2008 at a major east coast university. The subjects were 78 MBA students. With three alternative payoffs to test and the standard prediction market mechanism as benchmark, we obtain four different types of markets. We randomly assigned each student to one the four markets. Students did not participate in the same market-type more than once in order to eliminate possible learning effects. Moreover, to achieve a higher robustness of results, each market was run in two replications with a student assigned to either the first or second replication. This results in a total of eight different markets for each experiment, each experiment consisting of nine to eleven stocks. Consequently, we obtained a total of $4 \cdot 2 \cdot (11 + 10 + 9) = 240$ single stocks (see Table 3). We used winner-takes-all stocks (17 in total), which cash-out at either \$0 or \$100, and 13 linear stocks, which can cash-out at any value, depending on the definition of the stock (Spann and Skiera, 2003). The initial endowment of each trader consisted of \$10,000 in virtual currency. After each experiment, we re-set traders' portfolio values to the initial values to avoid the occurrence of endowment effects across experiments. We added up the final portfolio values after each experiment and cashed-out the stocks and thus obtained a list of students ordered by performance. The top 10% of students received 110% of extra course credit, the top 90% to 60% received a 100% of extra credit, the top 60% to 20% received 90% and the lowest 20% received 80%. Previous research (Luckner and Weinhardt, 2007) has shown that this rank-order tournament incentive scheme leads to the best results in terms of prediction accuracy in play-money markets. Additionally, we gave the top four traders gift certificates in values of up to \$50 (real-money) to create a second incentive to perform well. This information was given to all students before the start of the markets. As a consequence, it was transparent to all students that in order to obtain a high standing in the final rankings, students had to maximize their portfolio values. Participants were also instructed to be especially aware of the individual payoff mechanisms orally, by e-mail and within the trading system at different places, on the main screen and in the descriptions of the stocks.

Factor	**Number of levels**	**Specification**
Experiments (with different topics)	3	• Primaries on March 3rd, 2008 • "Final Four" NCAA basketball games on April 5th, 2008 • Economic events in or at the end of April
Payoffs	4	• Based on actual outcome • Based on vwap of last 48 hours • Based on last traded price • Based on last traded price with random close of market (within 4 hours of close of all markets)
Replications	2	• Each market with 9-10 traders
Stocks	10 (on avg. per experiment)	• Overall 17 winner-takes-all and 13 linear stocks
Total number of stocks/questions	$3 \cdot 4 \cdot 2 \cdot 10 = 240$	

Table 3: Study design

Vwap's were calculated over the last half of the trading period, i.e. over the last 48 hours. We found this period of time appropriate for two reasons. It is a) *short enough* to let the market prices move away from the initial starting prices and b) *long enough* for prices not to be moved easily by possible manipulation shocks. For the last-price-random-close markets, markets would close at a random point in time within the last 4 hours of trading. Both time spans were made transparent on the website to all traders participating in the corresponding markets. In order to test the risk-averseness of participants, we applied a common procedure from Holt & Laury (2002) via a survey. Details about the testing procedure can be found in the Holt & Laury paper.

Additionally, in order to compare the small actual-outcome markets' validity as benchmark, we set up a larger, self-contained market for the first experiment (primaries) with 11 stocks, consisting of 24 experts from political consultancy firms across the United States. Although the only extrinsic incentive for participants was a $100 gift certificate which was given to the winner, we believe that the display of traders in the ranking was a high incentive to perform well due to peer pressure. The payoff was based, as in traditional PMs, on the actual outcomes of the events.

As software, we used a self-developed trading platform which was tested during several previous experiments and field studies. Due to the expected low liquidity of the markets, which is the case for most PMs in general, we implemented Hanson's market maker algorithm (Hanson, 2003).

We also allowed for the short-selling of stocks, which, in conjunction with the market mechanism, enabled participants to move stock prices in their desired direction at any point in time. Among the different payoff groups, markets were completely identical, except for the descriptions of the stocks' final payoffs, the trading rules where the payoffs where explained, and the descriptions of the stocks. For example, the stock "A margin greater than 10 percent by either Clinton or Obama in Ohio" was described as follows:

- *In the traditional markets (actual-outcome):* "The price of this stock denotes the probability that the margin of votes by either Obama or Clinton is greater than 10 percentage points in the Ohio primaries […] After the elections, the stock will cash out after the primaries at $100 if the margin is more than 10 points, else at 0$.[…] The market closes Monday, March 3rd, 8 PM."
- *In the last-price markets:* "The price of this stock denotes the probability that the margin by either Obama or Clinton is greater than 10 percentage points in the Ohio primaries. […] The stock will cash-out at the **last fixed price** before the close of the markets on Monday, March 3rd, 8 PM."

5. Results

In the following, we discuss the results of our experiments. Because we have both winner-takes-all stocks, which range from 0 to 100, as well as linear stocks, ranging from a lower to an upper bound in our data set, we linearly normalize linear stocks to the range 0 and 100. E.g., a (linear) stock ranging from 40 to 60 with a price of $55 is mapped to ($55 - $40) / ($60 – $40) = 75, etc.

5.1 Correlation of Markets and Overpricing

We first investigate the correlation of the predictions of the actual outcome markets with the alternative payoff markets (Figure 1).

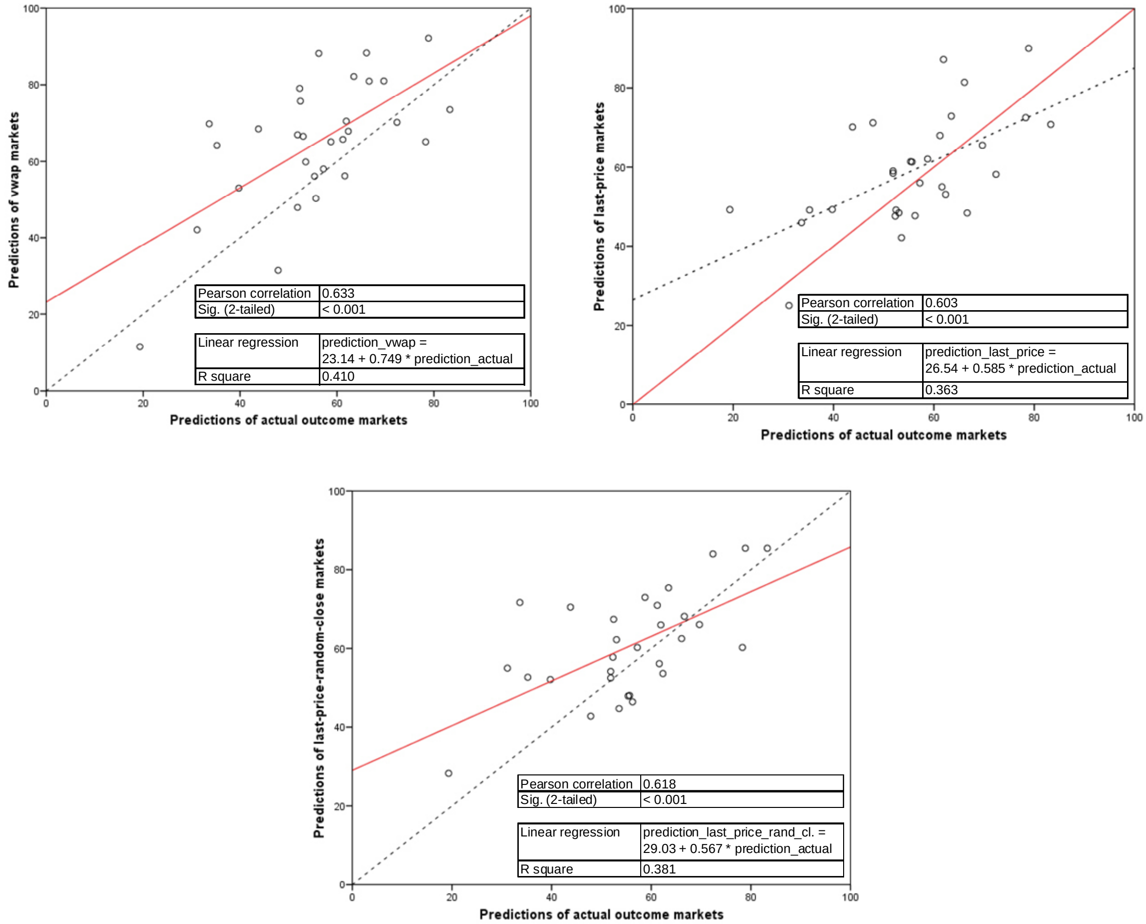

Figure 1: Correlations of average predictions of each stock of actual outcome markets with alternative payoff markets (dashed line signifies perfect correlation, solid line best linear correlation) and linear regression results

With each alternative payoff, we can see a significant ($p < .001$) correlation of predictions, implying a non-randomness of predictions by the alternative payoffs. Considering that payoffs are not tied to actual outcomes, but only on the trading of participants, this observation is not self-evident. The Pearson correlation in the vwap markets is the highest with a coefficient of 0.633 compared to the last-price-random-close markets with .618 and the last-price markets with 0.603 (Figure 1).

However, we remind that a high linear correlation does not imply that the predictions from both markets are identical. Observing the first plot in Figure 1, we notice that for the vwap, predictions are almost constantly above those of the actual outcome markets (for vwap, in 22 out of 30 cases), indicated by the dots above the dashed line. The average overpricing, i.e. the difference of the predictions of the alternative payoff markets, can be inferred from Table 4. All markets have an average overpricing, with the vwap markets being the largest with 9.11, followed by last-price-random-close-markets with 4.88 and the last-price markets with 3.37. These averages are significantly different from zero for the vwap markets (t-test, $p = 0.001$), as well as for the last-price-random-close markets ($p < 0.05$), but not statistically significant for the last price markets ($p > 0.1$). Thus, we can conclude that vwap markets result in the most overpriced market mechanisms; while overpricing is also present in the last-price-random-close markets, it is not significant in last-price markets, where under- und overpricing is roughly balanced. This is in line with research by LaComb et al. (2007, Table 2), where most of the prices traded above the starting price of $50. This is an indicator for the existence of price bubbles, discussed in Section 3, where traders exhibit a form of herd behavior in order to be better off with their trades. Another indication for the existence of this phenomenon is the trading activity, which is significantly higher in the vwap markets than in all other markets, including actual outcome markets.

	Vwap	**Last-price**	**Last-price-random-close**
Mean difference of alternative payoff prediction to actual outcome prediction	9.11	3.37	4.88
Std. error	2.51	2.32	2.23
N	30	30	30
Sig. that mean different from 0, t-test	.001	.158	.037

Table 4: Overpricing in alternative payoff markets

5.2 Forecasting accuracy

The mean absolute forecast errors are depicted in Table 5.

	Actual outcome with students	**Actual outcome with experts**	**Vwap**	**Last-price**	**Last-price -random-close**
Experiment 1					
Mean abs. error	18.15	19.72	30.70	23.39	31.66
Std. error	3.62	7.27	5.03	4.21	5.10
N	22	11	22	22	22
Experiment 2					
Mean abs. error	31.22		27.77	30.49	29.30
Std. error	6.50		6.31	6.66	5.71
N	20		20	20	20
Experiment 3					
Mean abs. error	39.28		46.05	48.37	41.83
Std. error	6.34		6.93	5.85	5.70
N	18		18	18	18
All experiments					
Mean abs. error	28.85		34.33	33.25	33.92
Std. error	3.37		3.63	3.49	3.24
N	60		60	60	60

Table 5: Mean absolute errors across experiments

First, in order to investigate the performance of the small actual-outcome student markets and to confirm them as a valid benchmark, we compare their performance to the expert markets with the (eleven) stocks in the first experiment. The actual-outcome student markets performed no worse than the expert markets; in fact, they performed slightly better, with an absolute error of 18.15 compared to an error of 19.72 for the expert markets, however, the difference is not significant (paired t-test, paired on each stock for the expert markets and each the average of the two actual outcome market stocks). We can thus confirm the actual-outcome student markets to be a legitimate benchmark.

When analyzing alternative payoff markets, we see that the actual-outcome markets performed better (in terms of absolute accuracy). Overall, the absolute error is 28.85 across all experiments; the second lowest error was obtained by the last-price markets with an error of 33.25 (absolute difference 4.40); the vwap and last-price-random-close markets performed only marginally worse with an error of 34.33 and 33.92, respectively.

Interestingly and in line with results obtained by Rosenbloom & Notz (2006), we find that prediction accuracy varies between different topics. As was found by these authors when comparing real- to play-money markets, for the second experiment (sports) the accuracy of the alternative payoff markets was better than that of the actual-outcome markets (27.77 / 30.49 / 29.30 to 31.22). On the other hand, for the two other experiments, the actual-outcome markets outperformed the alternative markets, with an absolute difference of at least of 3.45 in the first experiment and at least 2.55 in the third one. One possible explanation could be that there is more public information in sports compared to politics and the economy. As it was theoretically laid out in chapter 3, information revelation and aggregation is limited to public knowledge for the alternative payoff markets.

	Linear stocks				*Winner-takes-all stocks*			
	Actual	**Vwap**	**Last-Price**	**Last-price-random close**	**Actual**	**Vwap**	**Last-price**	**Last-price-random close**
Mean abs. error	14.01	19.78	18.77	17.71	48.25	53.36	52.19	55.12
Std. error	2.69	3.54	3.24	2.50	4.74	4.90	4.74	3.84
N	34	34	34	34	26	26	26	26

Table 6: Mean absolute errors across stock types

As we have two distinct types of stock in our experiments, we analyze them separately, as they are expected to produce different errors (Table 6). The reason is that the error of a winner-takes-all stock can be inflated, even when forecasting correctly. For instance, if the true probability of an event was 50:50 and the market correctly predicted a 50% probability, then the error has to be as large as 50 (i.e. it cannot be zero). In our experiments, the actual outcome markets perform better in both cases with an average error of 14.01 for linear stocks and an error of 48.25 for winner-takes-all stocks. However, the results for the alternative payoffs are less consistent. The last-price-random-close markets perform best for linear stocks with an error of 17.71, while they perform worst with winner-takes-all stocks with an error of 55.12. Among vwap and last-price markets, the last price markets outperform the vwap markets in both stock type cases with an error of 18.77 versus 19.78 for winner-takes-all and 52.19 versus 53.36.

5.3 Impact of Risk Aversion

As explained in Section 3, we expect that a trader's risk aversion might influence the prediction accuracy. Now, by determining if a participant is either risk-averse, risk-neutral or risk-seeking, we can determine the proportion of risk-averse traders in each market and thereby, analyze the impact of this effect on the forecasting accuracy. In total, we obtained 54[1] out of 78 traders of which we could identify as either risk-averse (22), risk-seeking (17) or risk-neutral (15). With this information, we are able to quantify the proportion of risk-averse traders who traded a certain stock. We only use the data of stocks whose traders were all classified by their risk-averseness. Also, some stocks were not traded at all, which we also excluded[2]. This leaves a total of 61 of all alternative payoff stocks with complete risk-averseness information[3].

We investigate the role of risk aversion in the following model:

[1] Some traders gave irrational answers, who we also excluded. Some traders did not complete the questionnaire or gave inconsistent answers.

[2] We believe that the non-trading of stocks arose due to a good initialization of stock prices.

[3] We do not believe that the robustness of our results is affected by the lower number of analyzed stocks due to, e.g. self-selection effects of participants who correctly answered the survey. In this case, no significant effects should be detected in the model.

$$e_i = \beta_0 + \beta_1 \cdot \text{DV\_linear}_{\text{i}} + \beta_2 \cdot \text{proportion\_risk\_aversion}_{\text{i,p,r}} + \mu_{\text{i,p,r}} \quad (1),$$

where the proportion of risk aversion is always between 0 and 1. e_i is the logarithm of the error plus one of stock $i \in \{1, \ldots, 30\}$, $p \in \{actual, vwap, last-price, last-price-rand-cl.\}$ is the index of the payoff and $r \in \{1, 2\}$ the index of the replication.

	Model 1		
Constant	4.149	(0.199)	***
DV_linear	-1.014	(0.193)	***
proportion_risk_aversion	-0.643	(0.334)	*
R^2 / adj.	0.354 / 0.332		
F-value	15.926 ***		
N	61		

Table 7: Regression results of impact of risk-aversion on forecast accuracy with coefficients (std. error) (*//***: significant at 0.1/0.05/0.01 level)**

Table 7 shows the results of the regression. The model fit is quite well with an R^2 of 35.4 %. As it can be inferred, risk-aversion has a significant effect on the forecast accuracy (p = 0.060). Thus, the more risk-aversion is in the market, the better the forecasting accuracy is for alternative payoffs, indicated by the negative sign of β_2.

6. Conclusion

In this paper, we showed that PMs can be used to determine forecasts of long-term or even non-occurring events to determine the stocks' payoffs, even without external proxy measures such as expert opinions. Regarding accuracy, PMs whose stocks' payoffs are based on the last traded price only perform 4.4 percentage points worse on average than traditional PMs which have shown superior accuracy in the past. The other two payoff mechanisms, vwap and last-price-random-close, only performed slightly worse. However, we could detect heavy overpricing with these two payoffs, especially in the vwap markets, which we could attribute to information cascades. Additionally, we could detect strong differences in forecasting accuracy between different topics among traditional PMs and PMs based on alternative payoffs. With topics with high public information, such as sports, alternative payoffs seem to work well, supporting theory. However, with less public and more privately held information, forecast accuracy substantially decreases compared to the benchmark. Finally, we could show that risk-aversion, which can be induced by appropriate incentives, significantly reduces the forecast error as a result of less gaming and manipulation in the market.

References

1. Anderson, L.R., and Holt, C.A. 1997. "Information Cascades in the Laboratory," *American Economic Review* (87:5), pp 847-862.
2. Berg, J., Forsythe, R., Nelson, F., and Rietz, T. 2001. "Results from a Dozen Years of Election Futures Markets Research," in: *Handbook of Experimental Economic Results,* C. Plott and V. Smith (eds.). Amsterdam: Elsevier, p. forthcoming.

3. Berg, J.E., Nelson, F.D., and Rietz, T.A. 2008. "Prediction Market Accuracy in the Long Run," *International Journal of Forecasting* (24:2), pp 285-300.
4. Berkowitz, S.A., Logue, D.E., and Eugene A. Noser, J. 1988. "The Total Cost of Transactions on the Nyse," *Journal of Finance* (43:1), pp 97-112.
5. Bikhchandani, S., Hirshleifer, D., and Welch, I. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy* (100:5), pp 992-1026.
6. Chen, K.-Y., and Plott, C.R. 2002. "Information Aggregation Mechanisms: Concept, Design and Implementation for a Sales Forecasting Problem," in: *Working Paper*. California Institute of Technology.
7. Dahan, E., Lo, A.W., Poggio, T., Chan, N.T., and Kim, A. 2007a. "Securities Trading of Concepts (Stoc)," in: *Working Paper*. University of California, Los Angeles, p. 33.
8. Dahan, E., Soukhoroukova, A., and Spann, M. 2007b. "Preference Markets: Organizing Securities Markets for Opinion Surveys with Infinite Scalability," in: *Working Paper, University of California Los Angeles*.
9. Forsythe, R., Nelson, F., Neumann, G.R., and Wright, J. 1992. "Anatomy of an Experimental Political Stock Market," *American Economic Review* (82:5), pp 1142-1161.
10. Graefe, A., and Weinhardt, C. 2008. "Long-Term Forecasting with Prediction Markets - a Field Experiment on Applicability and Expert Confidence," *Journal of Prediction Markets* (2:2), pp 71-92.
11. Hanson, R. 2003. "Combinatorial Information Market Design," *Information Systems Frontiers* (5:1), pp 107-119.
12. Hanson, R., Oprea, R., and Porter, D. 2006. "Information Aggregation and Manipulation in an Experimental Market," *Journal of Ecomomic Behavior and Organization* (60:4), pp 449-459.
13. Holt, C.A., and Laury, S.K. 2002. "Risk Aversion and Incentive Effects," *American Economic Review* (92:5), pp 1644-1655.
14. LaComb, C.A., Barnett, J.A., and Pan, Q. 2007. "The Imagination Market," *Information Systems Frontiers* (9:2-3), pp 245-256.
15. Luckner, S., and Weinhardt, C. 2007. "How to Pay Traders in Information Markets? Results from a Field Experiment," *Journal of Prediction Markets* (1:2), pp 1-10.
16. Ortner, G. 1998. "Forecasting Markets - an Industrial Application," Working Paper, Technical University of Vienna.
17. Rosenbloom, E.S., and Notz, W.W. 2006. "Statistical Tests of Real-Money Versus Play-Money Prediction Markets," *Electronic Markets* (16:1), February, 2006, pp 63-69.
18. Servan-Schreiber, E., Wolfers, J., Pennock, D.M., and Galebach, B. 2004. "Prediction Markets: Does Money Matter?," *Electronic Markets* (14:3), pp 243-251.
19. Slamka, C., Soukhoroukova, A., and Spann, M. 2008. "Event Studies in Play- and Real Money Prediction Market," *Journal of Prediction Markets* (2:2), pp 53-70.
20. Smith, V.L., Suchanek, G.L., and Williams, A.W. 1988. "Bubbles, Crashes and Endogenous Expectations in Experimental Spot Asset Markets," *Econometrica* (56), pp 1119-1151.
21. Soukhoroukova, A., and Spann, M. 2005. "New Product Development with Internet-Based Information Markets: Theory and Empirical Application," *13th European Conference on Information Systems (ECIS),* R.F. Bartmann D, Kallinikos J, Avison D, Winter R, Ein-Dor P, Becker J, Bodendorf F, Weinhardt C (ed.), Regensburg.
22. Soukhoroukova, A., Spann, M., and Skiera, B. 2008. "Creating and Evaluating New Product Ideas with Idea Markets," in: *Working Paper, University of Passau*.
23. Spann, M., and Skiera, B. 2003. "Internet-Based Virtual Stock Markets for Business Forecasting," *Management Science* (49:10), pp 1310-1326.
24. Spann, M., and Skiera, B. 2008. "Sports Forecasting: A Comparison of the Forecast Accuracy of Prediction Markets, Betting Odds and Tipsters," *Journal of Forecasting* (forthcoming).
25. Wolfers, J., and Zitzewitz, E. 2004. "Prediction Markets," *Journal of Economic Perspectives* (18:2), pp 107-126.