

HOW TO PAY TRADERS IN INFORMATION MARKETS: RESULTS FROM A FIELD EXPERIMENT

Stefan Luckner and Christof Weinhardt[†]

Institute of Information Systems and Management, University of Karlsruhe, Germany

Abstract

The results of recent studies on prediction markets are encouraging. Prior experience demonstrates that markets with different incentive schemes predicted uncertain future events at a remarkable accuracy. In this paper, we study the impact of different monetary incentives on prediction accuracy in a field experiment. In order to do so, we compare three groups of traders, corresponding to three treatments with different payment schemes, in a prediction market for the FIFA World Cup 2006. Somewhat surprisingly, our results show that performance-related payment schemes do not necessarily increase the prediction accuracy. Due to the risk aversion of traders the competitive environment in a rank-order tournament leads to the best results in terms of prediction accuracy.

[†] Email: luckner@iism.uni-karlsruhe.de

INTRODUCTION

Information markets are a promising approach for forecasting uncertain future events. The Iowa Electronic Market (IEM) for predicting the outcome of the presidential elections in 1988 was the first political stock market (Forsythe et al. 1992). Since then, political stock markets have been widely used as an alternative to polls and initially seemed to be the miracle cure in psephology, i.e. the scientific study of elections. Apart from political stock markets, the idea behind prediction markets has also been used in various settings such as market research or business forecasting (Spann and Skiera 2003). Lately, forecasting markets have also been used in order to predict the outcome of sports events (Schmidt and Werwatz 2002; Luckner et al. 2006).

The basic idea of a prediction market is to trade virtual stocks with certain payoffs that depend on uncertain future events. According to the efficient market hypothesis (Fama 1970), prices of traded assets reflect all available information and, thus, asset prices can be used to predict the likelihood of uncertain events. Consider a share that promises a payment of one currency unit for every percentage point a party obtains at an election. If, for example, a party wins 40 percent at the election, the participants receive 40 currency units for each share of that party they have in their portfolio. An investor who believes that the party will obtain 40 percentage points might sell his shares of this party for prices above and buy additional shares for prices below 40 currency units. Thus, the market prices reflect the expectations of the traders regarding the outcome of the election (Manski 2006). Several studies have shown that the market prices of the shares prior to the election are very close to the percentage points the respective parties win at the actual election. Political stock markets beat election polls in many cases (Berg et al. 2001).

The focal point of this work is to study the impact of different payment schemes on prediction accuracy in a field experiment. We want to elaborate on the question of whether prediction markets with performance-related incentives perform better than markets with fixed payments and how these performance-related incentives should be designed. This is of special interest when traders need to be paid for taking part in an information market, e.g. in the case of an internal market for company-specific predictions. Somewhat surprisingly, our results show that performance-related payment schemes do not necessarily increase the prediction accuracy. Based on our results we will give advice on engineering payment schemes for future information markets.

The remainder of the paper is structured as follows: The next section discusses some related work on incentives schemes in the field of experimental economics and two studies on

real-money vs. play-money prediction markets. In section 3, we then describe the setup of the field experiment we conducted during the FIFA World Cup 2006 in Germany. Subsequently, in section 4, we discuss our results concerning the impact of different payment schemes on the accuracy of predictions. Thereby, we also speculate why using performance-related incentives could possibly have led to a decrease in prediction accuracy. In section 5, we finally summarize our findings and give an outlook on possible implications these results might have on designing incentive schemes for public and intra-enterprise information markets.

RELATED WORK

Previous research in the field of prediction markets has shown that play-money as well as real-money markets can predict future events to a remarkable degree of accuracy (Forsythe et al. 1992; Spann and Skiera 2003). So far, market operators have employed various kinds of incentive schemes in order to motivate people to take part in such markets and to reveal their expectations. Typical examples are prizes for the top performers of an information market, lotteries among all traders, rankings published on the World Wide Web or even real-money exchanges. We suspect that the embodiment of the incentive mechanism has a huge impact on the market quality and the accuracy of predictions. Despite this, we are aware of just two papers studying incentive schemes for prediction markets by comparing real-money and play-money markets.

In one of these studies, Servan-Schreiber et al. (2004) found that there was no statistically significant difference between the real-money market TradeSports and the play-money market NewsFutures. Rosenbloom et al. (2006), however, found TradeSports to be significantly more accurate than NewsFutures for non-sports events. In the case of NFL games, they produced conclusions consistent with those from Servan-Schreiber et al. (2004). Considering both studies, we believe that the impact of real-money vs. play-money still remains an open question in the field of prediction markets. Moreover, there exists far more than one design option only for play-money markets – and also for real-money markets. In this paper we focus on designing payment schemes.

The strength of both studies is the large data set from real-world online experiments that both papers rely on. However, both studies do not consider any other differences apart from the use of real-money or play-money in their comparison of the two markets. Although the markets they compare are quite similar, they are far from identical. We agree that a key difference between the two markets is that one uses real money while the other does not. But how did some other aspects influence the prediction accuracy? It remains an open question

how, for example, the number of traders and their trading activity influences the market and thus also the accuracy of predictions. This seems to be an interesting question, since the number of traders per contract was not available to the authors in case of TradeSports. What is more, TradeSports does also levy a small fee on each transaction. How does this impact the trading behavior and the resulting asset prices? The two markets – TradeSports and NewsFutures – were not identical and we thus claim that other influencing factors might have caused the results described by Servan-Schreiber et al. (2004) and also by Rosenbloom et al. (2006).

As already mentioned before, these two are the only papers dealing with incentive schemes that we are aware of in the field of prediction markets. In experimental economics however, there is quite a lot of research concerning payment schemes for participants in lab experiments. Many experimental economists most probably would insist that monetary risk is required in order to obtain valid conclusions about economic behavior. Payments based on the participants' performance are usually intended to provide incentives for rational – or at least well considered – decision making. On the other hand, there is evidence that performance-related payments do not necessarily increase performance (Gneezy and Rustichini 2000).

Hence, we consider studying the impact of different payment schemes on the prediction accuracy of markets an open and interesting question. We thus conducted a field experiment to analyze several monetary incentive schemes that could for instance be used in internal information markets for company-specific predictions.

AN EXPERIMENTAL SPORTS PREDICTION MARKET

In this section we describe the setup of the field experiment we conducted during the FIFA World Cup 2006 in Germany. Firstly, we present the basic setup. Secondly, we elaborate on the three payment schemes we studied in our field experiment and explain why we chose these three incentive schemes. Thirdly, we discuss our expected results for this experiment.

BASIC SETUP

In our field experiment we were operating 20 prediction markets for the last 20 matches of the FIFA World Cup 2006. As assets we traded the possible outcomes of all the matches. There were three possible outcomes for every match – either team A won or team B won or there was a draw after the second half. We introduced the third asset (“draw”) although there were no draws possible in the tournament. The reason was that we did not want to care about penalty shootouts because we considered their outcome to be more or less unpredictable. The

asset corresponding to the event that actually occurred during the World Cup was valued at 100 currency units after the match; the other two assets were worthless.

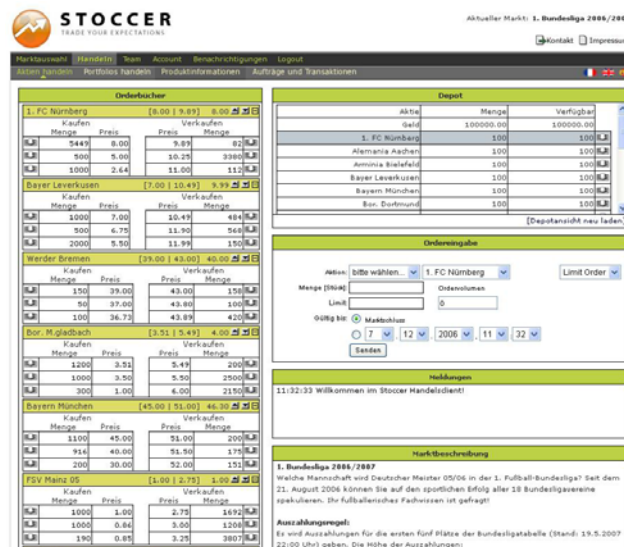


FIGURE 1. Web interface of the STOCER trading platform

In total, 60 undergraduate students from the University of Karlsruhe, Germany, were taking part in our field experiment in June and July 2006. All the markets opened about two days before the corresponding match and closed at the end of the match. We implemented a continuous double auction trading mechanism with an open order book. As a trading platform we used the system that is currently available at www.stoccer.com. A screenshot of the web interface is depicted in Figure 1. For more information on the system itself please refer to (Luckner et al. 2005).

INCENTIVE SCHEMES

We divided the 60 students randomly into three groups of 20 students each. At the end of the FIFA World Cup the traders were paid according to their group's incentive scheme. We can thus study the impact of three different payment schemes by comparing the prediction accuracy of the three groups of traders, corresponding to three treatments with different incentive schemes. The subjects of the first group were paid a fixed amount of 50 Euro (from now on referred to as fixed payment, FP). In the second group, individuals were paid according to their ordinal rank (rank-order tournament, RO). The trader ranked first was paid 500 Euro, the second 300 Euro and the third 200 Euro. All the other traders in this group did not receive any payment at all. This also results in an average payment of 50 Euro per person. To subjects in the third group we promised what we called a performance-compatible payment, also with an average amount of 50 Euro (deposit value, DV). Performance-

compatible means that the payment linearly depended on the traders' deposit value in the prediction market (deposit value divided by 10.000) and was therefore directly influenced by every transaction a trader conducted.

We chose these three payment schemes because we think they are somewhat related – although they are not the same – to incentives that we can nowadays typically observe in public information markets, namely markets without any payment, markets with rank-order tournaments, and real-money markets. Comparing these three different monetary incentives is also of interest for operators of internal markets for company-specific predictions since companies are usually willing to reward their employees' effort. In this case, the question arises which monetary incentive scheme is the most suitable.

For every group we separately ran the 20 markets on 20 soccer matches that were described in the last section. Since we did not want to pay students that were not trading at all we imposed a relatively small minimum trading volume per week on all of the traders. Especially in the case of the first group with the fixed payment we were worried that the students might otherwise consider not trading at all or simply forget to take part in the online experiment.

EXPECTED RESULTS

Before conducting our field experiment we expected the third group with the performance-compatible payment to be the best and the first group with a fixed payment to be the worst in terms of prediction accuracy. In the following we explain the intuition behind these expectations.

On the one hand, there exists no extrinsic motivation for members of the first group to reveal their expectations or to be among top performers of the group. In addition, there is no incentive for them to trade more than the minimum required trading volume per week. On the other hand, we should not forget about the traders' intrinsic motivation and perhaps their interest in soccer. Furthermore, traders don't risk any money and risk aversion does thus not come into play. Nevertheless, we suspected this payment scheme to perform worse than the others since it is known that the presence of financial incentives does affect average performance in many tasks (Camerer and Hogarth 1999).

Members of the third group receive a performance-compatible payment, meaning that every transaction directly influences their payment. Traders should consequently be motivated and try their best. Due to their risk aversion traders should try to avoid losing money and consider very carefully what and how to trade. We expected that traders in this group should for that reason trade less and at slightly lower asset prices. Their increased effort, however,

should improve their performance (Camerer and Hogarth 1999) and thus also the prediction accuracy. In short, traders with the incentive scheme DV have to “put their money where their mouth is” (Hanson 1999) and we consequently expected the predictions to be rather accurate.

For the second group we expected a result somewhere in between the other two groups. On the one hand, traders have a strong incentive to be among top three traders of their group because they will not receive any payment otherwise. This should lead to a rather high trading activity. Moreover, rank-order tournaments have also been considered as a promising payment scheme in other contexts (Lazear and Rosen 1981). On the other hand, the rank-order tournament provides an incentive to take higher risk compared to traders receiving, for example, the performance-compatible payment. Also, traders might start betting on unlikely events because they consider this the best or maybe even only way to outperform their competitors from the group. We also expected traders to stop trading as soon as they don't have any chance to be among the top three traders of their group. For these reasons, we expected that the performance-compatible payment scheme would outperform the rank-order tournament.

RESULTS

In this section we will now discuss the results from our field experiment. We will first compare the distribution of asset prices in the three treatments before discussing the impact of the three different payment schemes on the prediction accuracy.

MARKET PRICES

In total, every group traded 60 assets in 20 different markets (three assets per market). In Figure 2 we can see how many assets were traded within a certain price range in each of the three treatments. The very first column for example means that 32% of the assets were traded at prices between 0 and 20 virtual currency units in the first treatment with a fixed payment.

When comparing the three treatments we can observe that a relatively high number of assets are traded at prices between 60 and 100 currency units in the second treatment. This is exactly what we expected because people are obviously willing to take the risk in case of the rank-order tournament and buy assets even at rather high prices. Students in the third group with the performance-compatible payment, in contrast, do not trade any asset at a price between 80 and 100 currency units and almost no asset in the price range from 60 to 80. Obviously, traders with payment scheme DV are not willing to take the risk of buying assets at such high prices although there is no reason why their expectations should differ that much from the traders' expectations in the other two treatments.

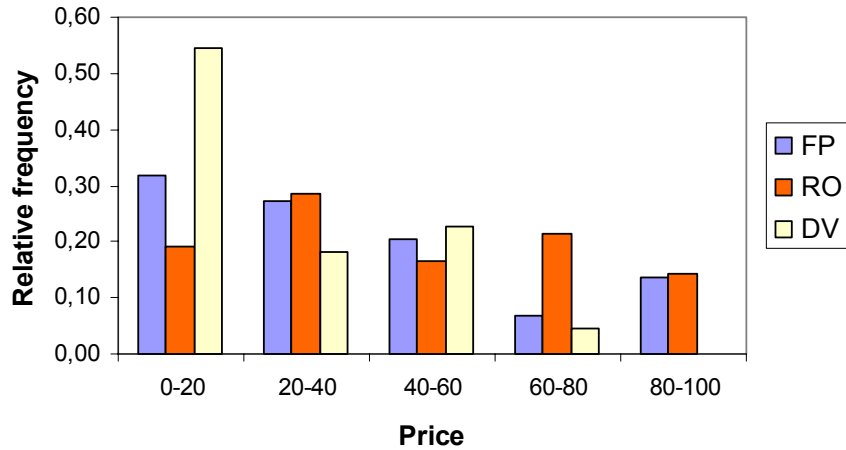


FIGURE 2. Distribution of asset prices in the three treatments

One again, “people are typically willing to pay less for almost anything if the money is real than if it is hypothetical” (Read 2005). One explanation for this behavior of traders in the third treatment could be their risk aversion. Due to their risk aversion traders seem to trade assets at lower prices compared to the other two treatments. More than 50% of the assets are even traded for less than 20 currency units. In the next subsection, we will discuss how this impacts the prediction accuracy of our treatments.

PREDICTION ACCURACY

Overall, 35% of the assets with the highest share price out of the three assets per match actually corresponded to the observed outcome in case of the fixed payment and the average pre-game trading price of the asset corresponding to the outcome was 40.83 virtual currency units. In the rank-order tournament, the most likely outcome according to the asset prices actually occurred in 45% of the cases and the average pre-game trading price of the asset corresponding to the outcome was 51.65 currency units. Finally, in case of the performance-compatible payment, the most likely outcome according to the asset prices occurred in merely 20% of the cases and the average pre-game trading price of the asset corresponding to the outcome was 26.64 currency units. This means, when interpreting the asset prices as probabilities the third treatment predicted the outcome of a match worse than randomly drawing one of the three possible events. This was indeed rather surprising to us. The rank-order tournament, however, seems to work quite well. The difference between the average pre-game trading prices of RO and DV is significant ($P < 0.05$, Mann-Whitney U test).

In the last subsection we have already learned that asset prices seemed to be rather low in case of the performance-compatible payment. This can also be seen when calculating the sum

of the three asset prices corresponding to the three possible outcomes of a match. These prices should sum up to about 100 virtual currency units since the probability that one of the three events occurs is 100%. In case of the performance-related incentive scheme the average price of such a so called portfolio is only 53.30 virtual currency units while it is indeed very close to 100 in the other two treatments (97.72 in the first treatment and 102.83 in the second treatment).

To analyze the correlation between asset prices and outcome frequency in more detail, we sorted the data into buckets by assigning all of the assets to one of five price ranges according to their pre-game trading price. The size of the circles and triangles indicates how many asset prices fall into the corresponding price range. The larger the circle or triangle is, the more assets were assigned to this bucket. Figure 3 plots the relative frequency of outcome against the prices observed before the match started. We look at the correlation between the relative frequency of outcome and the asset prices as an indicator for the accuracy of predictions. For the rank-order tournament the correlation coefficient is 0.84, while it is only 0.34 for the fixed payment and with 0.19 even worse for the performance-compatible payment scheme. Thus, the prediction accuracy is – in contrast to our expected results – quite poor in the third treatment DV. Somewhat surprisingly, the rank-order tournament clearly outperforms the other two incentive schemes and even the group with the fixed payment beats the group with the performance-compatible payment.

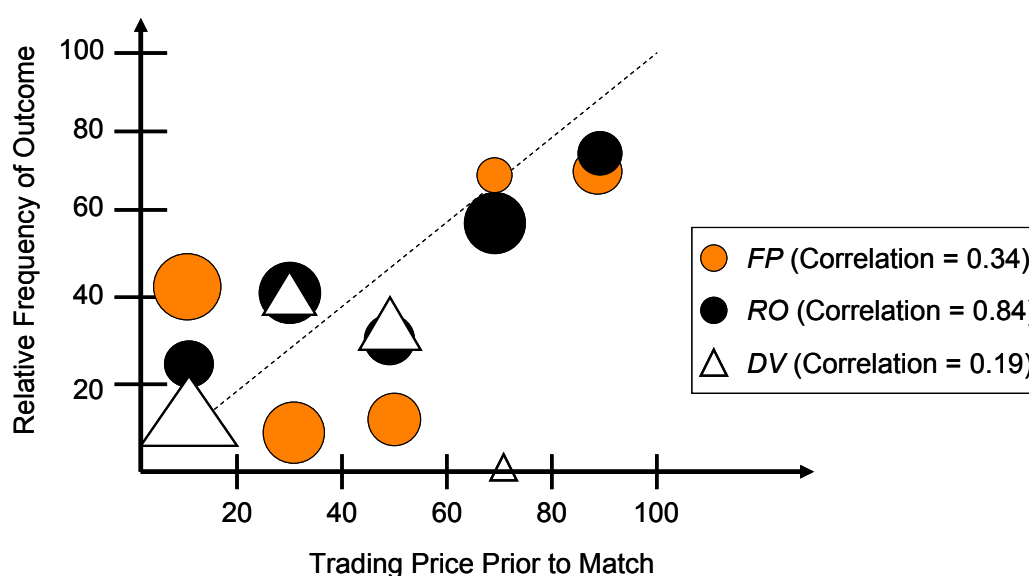


FIGURE 3. Prediction Accuracy: Market forecast probability and actual probability

As we have mentioned earlier, on average the sum of the three asset prices corresponding to the three possible outcomes of a match was only 53.30 virtual currency units in case of performance-compatible payment scheme. This might explain why the prediction accuracy is quite poor in case of this incentive scheme. To analyze this in more detail we divided all the asset prices by the average price of a portfolio and then once more plotted the relative frequency of outcome against the prices observed before the match started. Even then the rank-order tournament still performs much better than the performance-compatible incentive scheme.

DISCUSSION OF OUR RESULTS

We can only speculate about possible reasons for this result. Besides extrinsic motivation traders might also be intrinsically motivated. This could also help to explain why even the fixed payment scheme seems to work to some extent. Traders are obviously not only driven by monetary incentives. This would explain why they do not stop trading as soon as they reach the minimum weekly trading volume in the FP treatment. Also in case of the rank-order tournament traders continue to trade even if winning becomes extremely unlikely for them.

However, we think that the traders' risk aversion is most likely the main reason for our results. We conducted a lottery choice experiment as known from Holt and Laury (Holt and Laury 2002) in order to measure the traders' degree of risk aversion before we started our field experiment. The choices involved large cash prizes that were paid to the participants. Nearly 75% of our subjects exhibit risk aversion.

In case of the fixed payment, traders can neither win nor lose money and risk aversion does as a consequence not matter. Moreover, traders will take quite a lot of risk in the rank-order tournament because they have to be among the top performers within their group to receive the relatively high payment. Only in the third treatment, the performance-compatible incentive scheme, traders receive an endowment of 50 Euro and could potentially lose money with every transaction they make. As a result, buyers are obviously very careful and not willing to spend too much money on any asset. Sellers on the other hand are probably willing to sell at rather low prices to avoid the risk of holding shares of an event that does in the end not occur. Asset prices are thus much lower than in case of the two other treatments. Evidently, the performance-compatible payment scheme is not suitable to reveal the traders' expectations about the likelihood of events.

SUMMARY

In this paper, we have analyzed the impact of three payment schemes on the prediction accuracy of information markets. The results from our field experiment show that performance-compatible payment schemes seem to perform worse than fixed payments and the rank-order tournament. Due to the risk aversion of traders, the competitive environment in case of the rank-order tournament seems to lead to the best results in terms of prediction accuracy.

But what are the implications for designing future prediction markets? Well, out of the three incentive schemes we looked at one should probably choose the rank-order tournament when, for example, setting up an internal market for company-specific predictions where employees want to be paid for trading.

We also argued in this paper that performance-compatible payment schemes are somewhat similar to real-money markets. But can we now draw the conclusion that play-money markets e.g. with prizes for the top performers will outperform real-money markets although the latter raise numerous legal and technical difficulties? Probably yes, but we would rather be careful when answering this question based on our results because the situation might be somewhat different in prediction markets that are open to the public. In this case, there is a self-selection of traders and we would thus expect many risk-seeking traders in a public real-money market. In such a situation a performance-compatible payment scheme might potentially produce much better predictions than in the case of our field experiment.

ACKNOWLEDGEMENTS

This work is based on research funded by the German Federal Ministry for Education and Research under grant number 01HQ0522 and by the German Research Foundation (DFG) within the scope of the Graduate School Information Management and Market Engineering (IME). The authors are responsible for the content of this publication.

REFERENCES

- J Berg, R Forsythe, F Nelson and T A Rietz 'Results from a Dozen Years of Election Futures Markets Research' in C Plott and V L Smith (eds) *Handbook of Experimental Economic Results* (2001).
- C F Camerer and R M Hogarth 'The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework' *Journal of Risk and Uncertainty* (1999) 19 7-42.

- E F Fama 'Efficient capital markets: A review of theory and empirical work' *Journal of Finance* (1970) 25 383-417.
- R Forsythe, F Nelson, G Neumann and J Wright 'Anatomy of an Experimental Political Stock Market' *American Economic Review* (1992) 82 1142-1161.
- U Gneezy and A Rustichini 'Pay Enough Or Don't Pay At All' *The Quarterly Journal of Economics* (2000) 115(3) 791-810.
- R Hanson 'Decision Markets' *IEEE Intelligent Systems* (1999) 14(3) 16-19.
- C A Holt and S K Laury 'Risk Aversion and Incentive Effects' *American Economic Review* (2002) 92(5) 1644-1655.
- E P Lazear and S Rosen 'Rank-Order Tournaments as Optimum Labor Contracts' *The Journal of Political Economy* (1981) 89(5) 841-864.
- S Luckner, F Kratzer and C Weinhardt 'STOCCER - A Forecasting Market for the FIFA World Cup 2006' *4th Workshop on e-Business (WeB 2005)*, Las Vegas, USA (2005).
- S Luckner, C Weinhardt and R Studer 'Predictive Power of Markets: A Comparison of Two Sports Forecasting Exchanges' in T Dreier, R Studer and C Weinhardt (eds) *Information Management and Market Engineering* (Karlsruhe, Universitätsverlag Karlsruhe, 2006).
- C F Manski 'Interpreting the Predictions of Prediction Markets' *Economics Letters* (2006) 91(3) 425-429.
- D Read 'Monetary incentives, what are they good for?' *Journal of Economic Methodology* (2005) 12(2) 265-276.
- E S Rosenbloom and W W Notz 'Statistical Tests of Real-Money versus Play-Money Prediction Markets' *Electronic Markets - The International Journal* (2006) 16(1).
- C Schmidt and A Werwatz 'How well do markets predict the outcome of an event? The Euro 2000 soccer championships experiment' Max Planck Institute for Research into Economic Systems (2002) Discussion Paper on Strategic Interaction.
- E J Servan-Schreiber, J Wolfers, D Pennock and B Galebach 'Prediction Markets: Does Money Matter?' *Electronic Markets - The International Journal* (2004) 14(13).
- M Spann and B Skiera 'Internet-Based Virtual Stock Markets for Business Forecasting' *Management Science* (2003) 49 1310-1326.