Originally published in *Electronic Markets*, 14-3, September 2004.

Prediction Markets: Does Money Matter?

Emile Servan-Schreiber
NewsFutures, Inc.

Justin Wolfers
Wharton School, University of Pennsylvania

David M. Pennock
Yahoo! Research Labs

Brian Galebach
Probability Sports, Inc.

ABSTRACT: The accuracy of prediction markets has been documented for both markets based on real money and those based on play money. To test how much extra accuracy can be obtained by using real money versus play money, we set up a real-world on-line experiment pitting the predictions of TradeSports.com (real money) against those of NewsFutures.com (play money) regarding American Football outcomes during the fall-winter 2003-2004 NFL season. As expected, both types of markets exhibited significant predictive powers, and remarkable performance compared to individual humans. Perhaps more surprisingly, the play-money markets performed as well as the real-money markets. We speculate that this result reflects two opposing forces: real-money markets may better motivate information discovery while play-money markets may yield more efficient information aggregation.

Prediction markets—also called “idea futures” or “information markets”—are designed to aggregate information and produce predictions about future events: for example, a political candidate’s re-election, or a box-office take, or the probability that the Federal Reserve will increase interest rates at its next meeting. To elicit such predictions, contract payoffs are tied to unknown future event outcomes. For example, a contract might pay $100 if George W. Bush is re-elected in 2004, or nothing if he is not. Thus, until the outcome is decided, the trading price reflects the traders’ collective consensus about the expected value of the contract, which in this case would be proportional to the probability of Bush’s re-election.

Such markets have been available on-line to the general public since the mid-1990’s, in both real-money (gambling) and play-money (game) formats, and a few have developed large communities of regular traders. Popular play-money markets include the Hollywood Stock Exchange (http://www.hsx.com), which focuses on movie box-office returns, NewsFutures’ World News Exchange (http://us.NewsFutures.com) which covers sports, finance, politics, current events and entertainment, and the Foresight Exchange (http://www.ideosphere.com), which focuses on long term scientific discoveries and some current events. Real-money exchanges that are popular with the American public include the Iowa Electronic Markets (http://www.biz.uiowa.edu/iem), which focuses on political election returns (under a special no-action agreement with the CFTC, in part due to its university affiliation and individual investment limit of US$500), and TradeSports (http://www.TradeSports.com), a betting exchange headquartered in Ireland.

In the last few years, researchers have closely studied the predictions implied by prices in these markets, and have found them to be remarkably accurate, whether they operate with real-money or play-money. For instance, the researchers who operate the Iowa Electronic Market have found that their markets routinely outperform opinion polls in predicting the ultimate result of political elections in the U.S. and abroad (Berg et al. 2000; Forsythe et al. 1999). Pennock et al. (2001a; 2001b) looked

---

Correspondence concerning this article should be addressed to Emile Servan-Schreiber at ejss@NewsFutures.com.

at the trading prices from the Foresight Exchange and the Hollywood Stock Exchange, showing them to be closely correlated with actual outcome frequencies in the real world, in some cases outperforming expert prognostications. Prices in many sports gambling markets have shown excellent predictive accuracy while financial derivatives prices have been shown as good forecasts of the fate of their underlying instruments (Jackwerth & Rubinstein 1996; Roll 1984). In a series of experiments, researchers at Hewlett-Packard enrolled some of the company's employees as prediction traders, and found that their forecasts of product sales systematically outperform the official ones (Chen et al. 2002). Other controlled laboratory experiments have verified the power of prediction markets to aggregate information diffused across a trading population (Plott and Sunder 1988). Wolfers and Zitzewitz (2004) survey of the performance of prediction markets across these and other contexts.

Early successes have attracted the attention of corporations and policymakers, and most famously the Pentagon, eager to improve their forecasting methods by leveraging a wider base of knowledge and analysis. For example, the Pentagon agency DARPA had backed a project called the Policy Analysis Market (PAM), a futures market in Middle East related outcomes (Polk et al. 2003), until a political firestorm killed the project. Academic and policy interest in these markets remains robust, and it appears likely that private-sector firms will step into this void (Kiviat 2004; Pethokoukis 2004). Part of the allure is that whereas only so many people can be practically gathered into the same room at the same time for a coherent discussion, on-line prediction markets can easily aggregate the insights of an unlimited number of potentially knowledgeable people asynchronously.

Does money matter?

An oft-repeated assertion in the literature as to why prediction markets work so well is that, in contrast to professional pundits and respondents to opinion polls, traders must literally "*put their money where their mouth is*" (Hanson, 1999). The clear implication, and the common belief among economists especially, is that markets where traders risk their own money should produce better forecasts than markets where traders run no financial risk. This belief pervades the experimental economics community, which largely insists that monetary risk is required in order to obtain valid conclusions about economic behaviour. However, the relative efficiency of real-money versus play-money markets is an open empirical question; we are not aware of any prior study that has directly compared the accuracy of actual- and virtual-currency markets in a real-world setting.

Roughly speaking, prediction markets perform three tasks: they provide incentives for *truthful revelation*, they provide incentives for research and *information discovery*, and they provide an algorithm for *aggregating opinions*.

In terms of this taxonomy, real-money likely yields particularly robust incentives for information discovery, and the large number of analysts on Wall Street is an example of these incentives in action. It is also likely that individuals will be willing to bet more on predictions they are more confident about, suggesting an advantage in *intrapersonal* opinion weighting. However, in a market, the weights given to participants' opinions reflect the amounts that they are willing to bet, which might be affected by their wealth levels. Thus, in real-money markets, these *interpersonal* opinion weights likely reflect the distribution of wealth which can often reflect returns to skills other than predictive ability, or luck (such as an inheritance). By contrast, the only way to amass wealth in a play-money exchange is by a history of accurate predictions. As such, it seems plausible that play-money exchanges could have a countervailing advantage in producing more efficient opinion weights.

This research question also has important implications in practice. First, the distinction between "gambling" and "trading" in prediction markets, while not well-grounded in economics, is important for both an ethical assessment of these markets (as DARPA learned), and for the legality of a specific prediction market, since gambling is outlawed or subject to a state-run monopoly in many jurisdictions. Secondly, even in those countries that offer betting licenses, setting up an operation based on real money necessarily incurs huge technical, regulatory, and fiduciary costs far in excess of those required to operate the prediction-market technology itself. When one is the custodian of other people's money, any mistake, system failure, or fraud becomes business-critical. Thirdly, it is difficult to imagine a corporation requiring its employees to risk some of their own money on producing better company forecasts.

The alternative is to operate markets where traders run no financial risk. This does not preclude, however, some material or psychological upside for the traders in the form of bragging rights, prizes, or cash. Typically, the participants in such markets are given an initial amount of play-money to invest, and a few of those with the largest net worth when markets close win something. While participants in real-money markets are likely trying to maximize wealth levels, the play-money markets typically offer incentives that are more likely to depend on rank-order. As the popularity of diverse play-money exchanges attests, such incentives are often enough to motivate intense trading (e.g., Robinson, 2001).

In view of the legal, technical, financial and ethical obstacles to implementing real-money prediction markets, it is important for someone interested in using this technology to ask: '*how much accuracy (if any) am I going to lose if I use play money instead?*' The following experiment was designed to seek some initial answers.

The Experiment

We chose to compare the predictions of two popular online sports trading exchanges, one based on real-money, the other on play-money. Some reasons for choosing sports are: (1) the sheer frequency of games can yield many data points over a short period; (2) the intense media reporting of sports events and scrutiny of sports teams and personalities insures that enough information is publicly available that traders can be considered generally knowledgeable about the issues; (3) the standardization and objectivity of sporting events and rulings insures that contracts on both exchanges are defined equivalently, and that traders on both sites are indeed trading the same contracts; and (4) two popular and liquid exchanges already exist that are largely comparable, with the primary distinction being that one operates with real-money (TradeSports.com) and the other does not (NewsFutures.com).

TradeSports.com, based in Ireland for legal reasons, but targeted at U.S. consumers nonetheless, is a real gambling site that operates with real-money. NewsFutures.com's Sports Exchange, based in the U.S., is a play-money game which, throughout this experiment, was operated in partnership with USA Today. Both exchanges propose similar contracts on sporting events valued at 100 if a team wins, and 0 if it does not, with the trading price therefore directly reflecting the traders' collective assessment of the probability that the team will prevail. On both websites, trades are conducted directly between traders, with no intermediary, although TradeSports does levy a small fee on each transaction.

To become a trader on TradeSports, one must first deposit some money to play with, using, for instance, a credit card. Winnings can similarly be charged back to one's credit card. In contrast, NewsFutures registration is free, and a small amount of play money is given to each new trader and also to each trader who falls below a certain level of net worth. Because this inflationary system has been operating for more than two years, some skilled traders have been able to accumulate enormous amounts of play money, worth up to 20,000 times the initial allowance. This play money is not entirely worthless: the richest players can use it to bid on a few real prizes—worth a few hundred dollars—offered through auctions at the end of every month. So, even though NewsFutures traders cannot lose money by playing the game (in contrast to TradeSports gamblers), a few are able to convert their play-money winnings into real prizes.

The experiment started at the beginning of the US professional National Football League (NFL) season on 4 September 2003, and ran fourteen weeks until 8 December, spanning 208 NFL games (14 to 16 games per weekend). For each game, the prediction of each website was taken to be the last trade before noon (U.S. east coast time) on the day of the game. Prices were recorded automatically by a specially-designed web crawler program. Typically the game would not start until several hours after we recorded the market predictions. Traders were neither informed nor aware that their trading prices were being sampled for this experiment. With prices on both sides of each game, we have 416 observations, although only 208 are independent (the buy price of one team is, by construction, equal to 100 minus the sell price of its opponent).

On average, each NFL game on NewsFutures attracted about 100 traders, rarely less than 50, and rarely more than 200, out of a pool of about 11,000 active NewsFutures members over of the 14 weeks of the experiment. The number of traders per contract was not available for TradeSports, but we do know that there are around 10,000 registered and active TradeSports members, and that in our sample each contract attracted on average US$7,530 in trades. If one assumes a typical average bet of less than US$100 per person, we can deduce that the number of participants per contract on TradeSports is of the same order of magnitude as on NewsFutures.

To compare the forecasting ability of the markets with that of individual human (self-declared) experts, we entered the trading prices from both markets into a popular internet prediction contest called ProbabilityFootball (http://www.ProbabilityFootball.com). This contest is original and well-fitted to the purpose because, rather than asking participants to just predict who is going to win each game, it asks them to rate the probability that a team will win. So one would enter 67 per cent if one believes that the team has 67 per cent probability of winning the game. The contest then rewards or penalizes participants according to the quadratic scoring rule, one of a family of so-called proper scoring rules (Winkler 1968) that reward players such that each player maximizes his or her expected score by reporting true probability assessment. The specific scoring function employed by the contest is $+100 - 400 * lose\_prob^2$, where *lose_prob* is the probability the player assigns to the eventual losing team. The scoring rule rewards confident predictions more when they are correct, and penalizes confident predictions more when they are wrong. For example, a prediction of 90 per cent (probability 0.9) earns 96 points if the chosen team wins and loses 224 points if the chosen team loses. In contrast, a prediction of 60 per cent earns 36 points if correct and loses 44 points if incorrect. A prediction of 50 per cent earns no points, but equally, loses no points. Participants in this contest were also required to produce their probability predictions before noon (U.S. east coast time) on the day of the game. On the 14th week of the experiment, 1,947 individual human participants were competing against our two prediction markets.

The Results

Overall, 65.9 per cent of TradeSports' favorite teams actually won their games (135 out of 208), and its average pre-game trading price was 65.1 for the favorite. NewsFutures fared similarly with 66.8 per cent favorite team victories (139 out of 208), and an average pre-game trading price of 65.6 for the favorite. We observe at this level a close correspondence between the markets' trading prices and the actual frequency of victory in the field. Both types of markets also had almost exactly the same prediction accuracy.

To analyze the correspondence between trading prices and outcome frequency in finer detail, we sorted the data into buckets by rounding each home-team trading price to the nearest whole factor of 10. Figure 1 plots the frequency with which home teams within each bucket won. It shows, again, but at a finer level, significant correlation between trading prices and outcome frequencies. The points at the extremes are based on fewer data points (because most NFL games are expected to be highly competitive), yet even so, these sample are extremely accurate. For TradeSports, the correlation coefficient is 0.96, while it is 0.94 for NewsFutures. Again, neither market seems to reliably outperform the other.

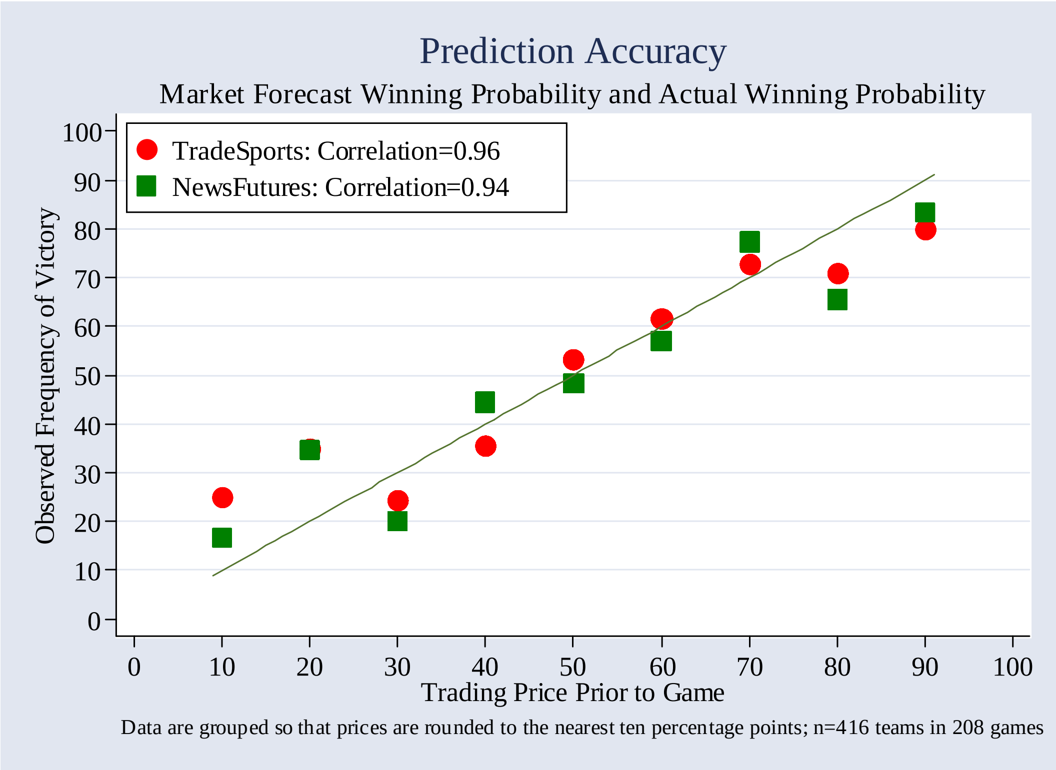

Figure 1: Pre-game home-team prices for each game are rounded to the nearest ten percentage points, and the observed frequency of victory is plotted against these prices.

We turn to assessing the relative forecast performance of each prediction market. Table 1 presents four common metrics of forecast accuracy, comparing TradeSports with NewsFutures:

One simple scoring rule is to simply consider a victory as a score of one, and a loss of a score of zero, and to assess the forecast errors as the (absolute value of the) difference between the ex post outcome and the market ex ante predicted probability of winning. As such, the forecast error is equal to the probability or price assessed for the losing team. The first row shows the average of these forecast errors, taking the prices from each prediction market as their prediction. The losing team was typically slightly more favoured on TradeSports than on NewsFutures, although the final column shows that this difference is both extremely small and statistically insignificant.

The square root of the mean squared error is a familiar measure of forecast errors, and the second row of Table 1 shows that under this scoring rule there is no statistically significant difference in the accuracy of the two prediction markets.

The ProbabilitySports contest employs a quadratic scoring rule, in which the loss function varies with the square of the prediction error, shown in the third row of Table 1.

The fourth row shows the average logarithmic score, another common proper scoring rule also appropriate for judging the accuracy of probability assessments. The logarithmic score is the logarithm of probability assigned to the winning team (in this context, the probability is the winning team’s price divided by 100); the table reports this quantity averaged over the 208 samples. Across these four measures of forecast accuracy, the advantage to NewsFutures is tiny, and in no case comes close to being statistically significant.

Table 1: Assessing the relative prediction accuracy of real-money and play-money markets

	ProbabilityFootball Average	TradeSports (real money)	NewsFutures (play money)	Difference (TS-NF)
Mean Absolute Error *=lose_price* *[lower is better]*	0.443 (0.012)	0.439 (0.011)	**0.436** (0.012)	0.003 (0.016)
Root Mean Squared Error *=√Average(lose_price^2)* *[lower is better]*	0.476 (0.025)	0.468 (0.023)	**0.467** (0.024)	0.001 (0.033)
Average Quadratic Score *=100 + 400*(lose_price^2)* *[higher is better]*	9.323 (4.75)	12.410 (4.37)	**12.427** (4.57)	-0.017 (6.32)
Average Logarithmic Score *=Log(win_price)* *[less negative is better]*	-0.649 (0.027)	**-0.631** (0.024)	**-0.631** (0.025)	0.000 (0.035)

win_price = winning team's price / 100

lose_price = losing team's price / 100

Standard error shown in parentheses.

An alternative accuracy test computes how much profit could theoretically be made in one market by trading according to the probabilities given in the other market. Note that this is a hypothetical test only, since the precise availability of trades was not recorded, only the last traded price. A strategy of buying exactly one contract at the TradeSports price if the NewsFutures price is greater (or selling exactly one contract at the TradeSports price if the NewsFutures price is smaller) yields a positive rate of return of 4.8 per cent. A strategy of buying exactly one contract at the NewsFutures price if the TradeSports price is greater (or selling exactly one contract at the NewsFutures price if the TradeSports price is smaller) yields a slightly greater return of 8.0 per cent, suggesting a slight edge for the TradeSports predictions according to this measure. The fact that both strategies yield a positive profit suggests that a more efficient estimator of the likely outcome lies somewhere between the two prices.

This leads us to our third approach, which is to run a simple linear regression of the winning team against the prices in each market:

Team i wins =	*-0.004 +*	*0.50 * TradeSports +*	*0.51 * NewsFutures*
	(.092)	*(.75)*	*(.72)*

n=416 teams; R^2*=.12* *(Standard errors in parentheses adjusted to reflect 208 independent games)*

The regression puts equal weight on the TradeSports and NewsFutures prices, thus treating them as equally accurate. Across all of our tests the differences in predictive power are quite small and we conclude that the predictive accuracies of the two markets are statistically indistinguishable.

To further investigate the statistical significance of our results, we employed the so-called *randomization test* (Fisher 1966; Noreen 1989). Results are reported in Table 2. We describe the testing procedure for determining the statistical significance of the difference between the mean absolute error of TradeSports' predictions and the mean absolute error of NewsFutures' predictions;

the remaining tests are analogous. First we record the difference between the mean absolute error of TradeSports' predictions and the mean absolute error of NewsFutures' predictions. Call this quantity *OrigDiff*. In this case, *OrigDiff* = 0.003, as reported in Table 1. Next, we randomly swap NewsFutures' and TradeSports' predictions, creating two new groups of randomly re-shuffled predictions. We then compute the new difference in mean absolute error of the two (randomized) groups. Call this quantity *RandDiff*. The statistical confidence values reported in Table 2 are the percentage of times (out of 10,000 trials) that |*OrigDiff*| > |*RandDiff*|. If the TradeSports and NewsFutures predictions arose from the same distribution, the confidence value would not be very high. On the other hand, a high confidence value means that, with high probability, the differences reported in Table 1 are statistically significant. Table 2 shows that , with high confidence (>95%), we can say that NewsFutures' predictions are better than ProbabilityFootball average predictions. With not quite as high confidence (>90%) we can say that TradeSports' predictions are better than ProbabilityFootball average predictions (except for the mean absolute error metric). Finally, in agreement with all our previous tests, the difference between NewsFutures' and TradeSports' predictions is not statistically significant to any reasonable degree.

Table 2: Assessing the statistical confidence of the differences in prediction accuracy of real-money markets, play-money markets, and opinion averages. For example, the upper-left entry in the table should be interpreted as saying that "with 97.7% confidence, the mean absolute error of NewsFutures' predictions is statistically significantly lower than the mean absolute error of ProbabilityFootball average predictions."

	NewsFutures vs TradeSports (%)	**TradeSports vs ProbabilityFootball Avg (%)**	**NewsFutures vs ProbabilityFootball Avg (%)**
Statistical confidence of difference in mean absolute error	62.3	65.3	**97.7**
Statistical confidence of difference in root mean squared error	1.3	91.9	**99.0**
Statistical confidence of difference in average quadratic score	1.2	91.8	**99.0**
Statistical confidence of difference in average logarithmic score	2.9	92.8	**99.1**

Confidence scores are computed using the randomization test (Fisher 1966; Noreen 1989).

Confidences above 95% shown in bold.

Were there differences in prediction behaviour even if there was little difference in predictive performance? FIGURE 2 plots the NewsFutures prices against the corresponding TradeSports prices for all 208 games. We observe a tendency for NewsFutures prices to be somewhat more dispersed (standard deviation = 18.1 percentage points) than TradeSports (standard deviation = 17.4 percentage points), meaning that the chosen favorite tended to be given a greater chance to win on NewsFutures than on TradeSports, though the distinction is slight. On average, NewsFutures and TradeSports prices differed by 3.4 per cent, with a standard deviation of 2.8 per cent. These reasonably large differences in forecasts are not surprising, because real-money markets and play-money markets are not directly linked by arbitrage.

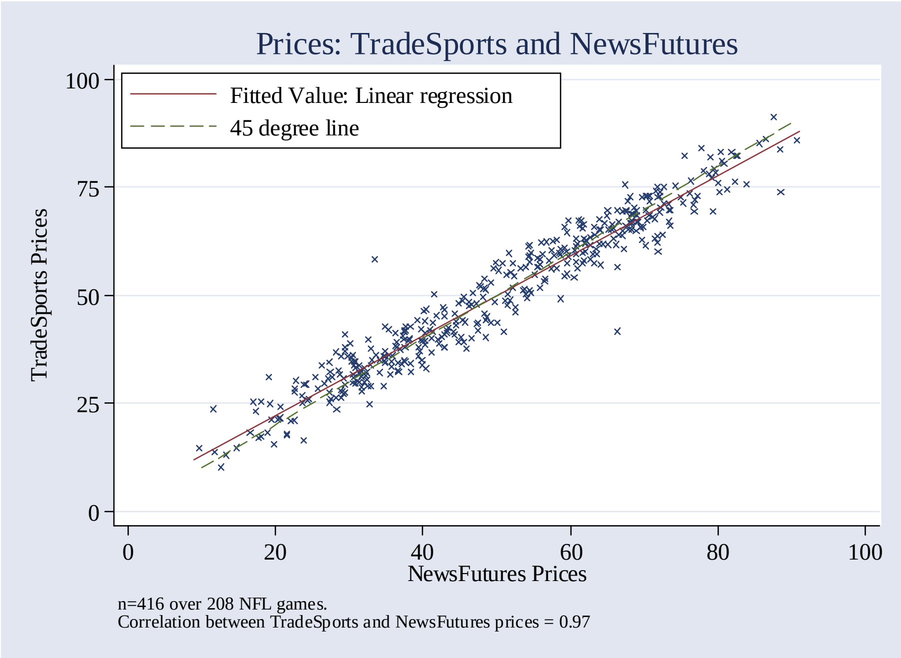

Figure 2: NewsFutures prices plotted against the corresponding TradeSports prices for each of 208 NFL games. The correlation coefficient between the two sets of data is .90. NewsFutures prices are slightly more disperse than TradeSports.

Finally, let us look at the how well the markets performed against the 1,947 individual contestants in the ProbabilityFootball forecasting contest. FIGURE 3 plots the progression of both TradeSports and NewsFutures in the contest rankings. Both real and play-money prediction markets have quickly and steadily closed in on the top ranks. At the end of the 14th week of the NFL season, NewsFutures (play-money) was ranked 11th, and TradeSports (real-money) was ranked 12th, comfortably within the top 1 per cent of the participants. (By the end of the 2003-2004 NFL season, which covered a total of 21 weeks, NewsFutures was ranked 6th and TradeSports 8th.) Alternatively phrased, for both markets we can reject the hypothesis that they yield forecasts that are only as accurate as the average individual. For comparison, the ProbabilityFootball averages ranked 39th, performing better than the vast majority of individuals, but not as well as the two markets.

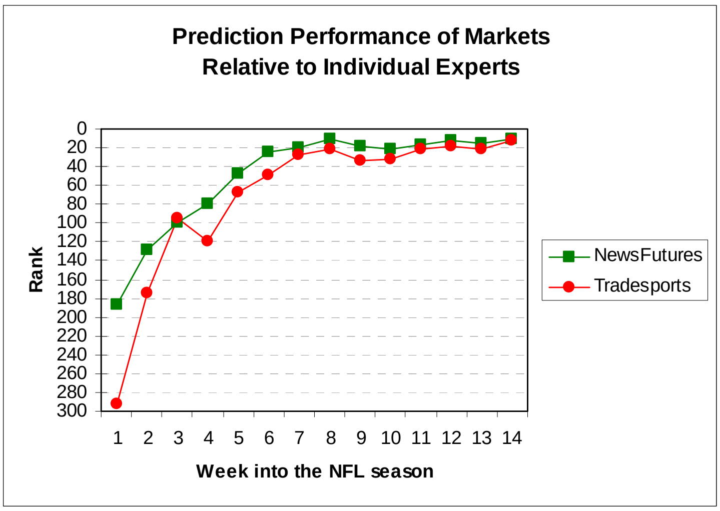

Figure 3: The markets competed against 1,947 individual self-declared experts in the ProbabilityFootball forecasting contest. They quickly and steadily closed in on the top ranks. By week 14, they were ranked 11th (NewsFutures / play money) and 12th (TradeSports / real money), comfortably within the top 1 per cent of all participants. By the end of the 2003-2004 NFL season, which covered a total 21 weeks, NewsFutures ranked 6th and TradeSports ranked 8th, both comfortably within the top 10 of nearly 2,000 participants.

FIGURE 4 plots the actual accumulation of contest points from week to week for both NewsFutures and TradeSports. The difference is visibly negligible.

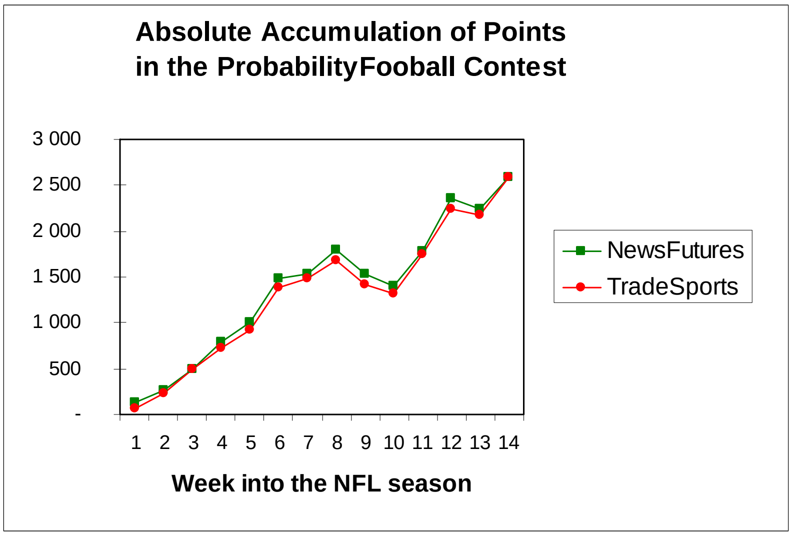

Figure 4: Participants in the ProbabilityFootball contest earn points for each victory predicted with over 50 per cent probability, and earn more points when they assign a higher probability to a victory than when they assign it a lower probability. Conversely, predicted victories that end in defeat subtract points in proportion to the strength of the failed prediction. Both types of markets ran neck and neck throughout the season.

Discussion

Both types of markets performed remarkably well compared to individual human probability estimators, ranking 11th and 12th in a competition against 1,947 humans and covering 208 NFL games. Their trading prices also correlated well with actual outcome frequencies, which suggests that the trading prices can indeed be read as probability estimations of real-world outcomes. Both of these results confirm earlier findings in the literature (e.g., Pennock et al, 2001).

The original research question we tried to address with our experiment was whether one type of market (real money) performs better than the other type (play money). The answer from this experiment appears to be "no": We found no significant difference in predictive accuracy. The differences in trading prices seems to suggest that the two markets did not simply both align their prices on publicly available bookmaker odds; similar accuracies are not purely a function of equivalent prices.

The two websites we chose to compare are quite similar in that they offer mechanically and conceptually equivalent markets, and they are both populated by traders recruited primarily from the general U.S. population. The primary difference between them is that one uses real money whereas the other uses play money. This likely has some impact as well on the kind of person that registers to trade on one or the other. But, other than that, traders on both websites are obviously motivated and, at least in general, knowledgeable about the issues being traded.

In declaring a draw between real-money and play-money prediction markets, it is worth reiterating the context of this experiment. The presence of deep instrinsic interest in NFL football and the existence of large betting markets already serves to motivate substantial information discovery in these markets, with team abilities already scrutinized in the daily press, on ESPN, and around the water cooler. This is also a context in which there is little reason to believe that forecasters will not truthfully report their views (except perhaps when team bias gets in the way). Thus perhaps the most important factor in generating an efficient forecast is weighting the relative opinions of many forecasters. On this score, it appears that real-money and play-money markets perform around as well as each other.

In light of our results, we would argue that knowledge and motivation are the essential factors responsible for the accuracy of prediction markets, and that the use of real money is just one among many ways of motivating knowledgeable traders to participate. In the case of play money, knowledgeable traders can be motivated, for example, by community bragging rights, or by prizes awarded to the best forecasters. In practice, the problem of recruiting knowledgeable traders to a play-money market can be reduced to the matter of expending some marketing effort.

Conclusion

The question we tried to address was: how much prediction accuracy is lost when one operates prediction markets with play money rather than real money, the big difference being whether one requires traders to take a personal financial risk or not. Besides its intrinsic scientific merit regarding the psychological importance of hard currency, this question is also very much of practical importance in view of the geographical, financial, technical, fiduciary, regulatory, and, perhaps, ethical obstacles to the establishment of real-money predictions markets, which, in most parts of the world, are viewed as just a fancy kind of betting shop. If the play-money alternative doesn't force one to compromise too much accuracy, then the ease of implementing them should help prediction market technology find wider uses in public policy, corporate forecasting, and product research. Theory suggests that real money may better motivate information discovery, while in play money markets those with substantial wealth are those with a history of successful prediction, suggesting potential for more efficient weighting of individual opinions.

To find some answers, we compared the predictions of two popular sports trading websites, one that operates play-money markets of the type that can be easily implemented in corporate settings or in accordance with strict anti-gambling legislation (NewsFutures.com), and another that operates as a sophisticated betting operation (TradeSports.com).

We found that neither type of market was systematically more accurate than the other across 208 experiments. In other words, prediction markets based on play money can be just as accurate as those based on real money. In this case, (real) money does not matter. The essential ingredient seems to be a motivated and knowledgeable community of traders, and money is just one among many practical ways of attracting such traders.

References

Berg, JE, Forsythe, R, Nelson, F, & Rietz, TA (2000) Results from a dozen years of election futures markets research. Technical Report, University of Iowa.

Chen, KY, Plott, CR (2002) Information Aggregation Mechanisms: Concept, design, and implementation for a sales forecasting problem. *Lee Center Workshop*.

Debnath, S, Pennock, DM, Giles, CL, & Lawrence, S. (2003) Information incorporation in online in-game sports betting markets. *Fourth ACM Conference on Electronic Commerce,* pp. 258–259

Forsythe, R, Rietz, TA & Ross, TW (1999) Wishes, expectations, and actions: A survey on price formation in election stock markets. Journal of Economic Behavior and Organization, 39:83–110.

Fisher, RA (1966) *Design of Experiments*, 8th edn, New York: Hafner (Macmillan).

Hanson, R, (1999) Decision Markets. *IEEE Intelligent Systems*, 14(3):16–19.

Jackwerth, JC, & Rubinstein, M. (1996) Recovering probability distributions from options prices. *Journal of Finance*, 51(5): 1611–1631.

Kiviat, B (2004) The End of Management? TIME, Inside Business, July 12, 2004.

Noreen, EW (1989) *Computer-Intensive Methods for Testing Hypotheses: An Introduction*, New York: Wiley.

Pennock, DM, Lawrence, S, Giles, CL, & Nielsen, FA (2001a) The real power of artificial markets. *Science*, 291 (5506): 987-988.

Pennock, DM, Lawrence, S., Nielsen, FA, & Giles, CL (2001b) Extracting collective probabilistic forecasts from web games. *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 174–183.

Pethokoukis, J (2004) All seeing all knowing. US News & World Report, August 22, 2004.

Plott, CR & Sunder, S (1988) Rational expectations and the aggregation of diverse information in laboratory security markets. *Econometrica*, 56(5):1085–1118.

Polk, C, Hanson, R, Ledyard, J, & Ishikida, T, Policy analysis market: An electronic commerce application of a combinatorial information market. *Fourth ACM Conference on Electronic Commerce*, pp. 272–273.

Roll, R (1984) Orange juice and weather. *American Economic Review*, 74(5): 861–880.

Winkler, RL & Murphy, AH (1968) Good probability assessors. *Journal of Applied Meteorology*, 7:751–758.

Wolfers, J, Leigh, A, Zitzewitz, E (2003) What do financial markets think of war in Iraq? *NBER Working Paper 9587*

Wolfers, J and E. Zitzewitz (2004) Prediction Markets. *Journal of Economic Perspectives* Spring 2004.

About the Authors

Emile Servan-Schreiber (ejss@NewsFutures.com) is cofounder and CEO of NewsFutures, Inc., a leading provider of prediction markets software and consulting services. He is also a consultant for the OECD's Brain Research and Learning Sciences project. As a cognitive scientist, he is primarily interested in learning systems and emergent phenomena.

Justin Wolfers (jwolfers@wharton.upenn.edu) is an Assistant Professor of Economics at the Wharton School, University of Pennsylvania, and a faculty research fellow of the National Bureau of Economic Research. An economist, his research interests include behavioural finance, prediction markets, labour economics and macroeconomics.

David M. Pennock (david.pennock@overture.com) is a Senior Research Scientist at Yahoo! Research Labs. His research interests include prediction markets, auctions, Web analysis and modelling, recommender systems, machine learning, and artificial intelligence.

Brian Galebach (research@ProbabilitySports.com) is the founder of ProbabilitySports.com, a company that operates online sports contests. He has combined his passion of mathematics and sports by creating and operating these free contests designed to objectively determine the accuracy of participants' probability estimations of sports outcomes.