PUBLIC INFORMATION BIAS AND PREDICTION MARKET ACCURACY

Thomas S. Gruca† and Joyce E. Berg‡§

†Henry B. Tippie Research Professor of Marketing, Henry B. Tippie School of Business, University of Iowa, Iowa City, IA 52242-1994
‡Associate Professor of Accounting, Henry B. Tippie School of Business, University of Iowa, Iowa City, IA 52242-1994
§Tel.: 319-335-0840. Email: joyce-berg@uiowa.edu

How do prediction markets achieve high levels of accuracy? We propose that, in some situations, traders in prediction markets improve upon publicly available information. Specifically, when there is a known bias in publicly available information, markets provide an incentive for traders to “de-bias” this information. In such a situation, a prediction market will provide a more accurate forecast than the public information available to traders. We test our conjecture using real-money prediction markets for seven local elections in the United States. We find that the prediction market forecasts are significantly more accurate than those generated using the pre-election polls.

Keywords: prediction markets, information aggregation, election forecasting, public information

INTRODUCTION

In a prediction market, traders buy and sell contracts whose value is tied to an uncertain future event. Originally used for predicting election outcomes (Forsythe et al., 1992), today, prediction markets are forecasting product sales (Chen and Plott, 2001), and movie box office returns (Spann and Skiera, 2003). There are other markets seeking to predict the winners of movie awards (Pennock et al., 2000) and sporting events (Servan-Schrieber et al., 2004). Their ability to provide accurate insights into future events has raised a great deal of interest in the business community (Kambil and Van Heck, 2002; Kiviat, 2004) as well as in the popular imagination (Surowiecki, 2004). For a recent review of research on prediction markets, see Wolfers and Zitzewitz (2004).

Businesses are interested in prediction markets due to their perceived ability to tap into potentially valuable private information held by employees who can be widely scattered throughout various levels of the firm. Usually, top decision makers cannot access this vital information due to barriers imposed by hierarchical reporting systems typical of large, complex organizations. Through proper market design, suitable incentives and judicious selection of traders, an organization can use a prediction market to aggregate valuable private information from market participants (Kambil and van Heck, 2002). In fact, the ability of prediction markets to tap into private information held by very few traders was one of the key motivations for the DARPA experiments into “terrorism futures markets” (Hulse, 2003).
Prediction markets have a major advantage over other methods for aggregating information from a wide variety of sources (Plott, 2000). Rational expectations theory (Lucas, 1972; Grossman, 1981) suggests that “price summarizes and reveals all the relevant information in possession of all traders” (Sunder, 1995). That being said, it is surprising how few studies (outside of a lab setting) determine how well the information held by traders is aggregated into the prices observed in a prediction market. One exception is a study by Gruca, Berg and Cipriano (2005) suggesting that market prices accurately reflect both the consensus of traders’ private forecasts as well as their differences of opinion. In this paper, we focus on how prediction markets can aggregate and improve publicly available information to provide superior forecasts of future events.

The central argument of our research is that traders in prediction markets are able to correct known biases in publicly available information. Motivated by monetary gains, traders move prices in the proper direction resulting in more accurate forecasts of future events. We test this hypothesis using the results of seven prediction markets associated with local elections in the United States. In this setting, the public information available to traders consists of pre-election polls of potential voters. Prior research (Gruca, 1994; Panagakis, 1989) shows that the vote shares for challengers of incumbents seeking reelection are significantly biased downwards. Traders in prediction markets “de-bias” the information provided by polls resulting in a more accurate forecast than is available from public opinion polls.

In Section 1, we highlight some limitations of the current view of how prediction markets generate superior predictions. Section 2 contains a discussion of the role of biased public information in prediction markets. We document the extent of the bias in the publicly available information (i.e. political polls) with respect to local U.S. elections in Section 3. In Section 4, we examine the ability of the Iowa Electronic Market, a real money prediction market, to improve on the publicly available polling data. We close with a discussion of the results and their implications in Section 5.

1. BIASED + MARGINAL TRADERS = EFFICIENT PRICES?

The documented accuracy of prediction markets presents an interesting challenge to researchers trying to understand how prediction markets aggregate information. From research on individual traders in political prediction markets, there is ample evidence that the average participant is biased and prone to trading errors. Consider, for example, the traders in the original 1988 U.S. Presidential election market. Self-identified supporters of Bush were more likely to believe that their candidate won each debate. These same traders increased their holdings of Bush contracts after each debate while supporters of the opponent Dukakis reduced their holdings of Bush contracts (Forsythe et al., 1999). In a prediction market for the 1992 U.S. Presidential election, Oliven and Rietz (2004) found that traders often made mistakes in submitting limit orders or executing market orders. These errors
included posting prices that allowed other traders to execute transactions resulting in risk-free profits, i.e. arbitrage opportunities. Nevertheless, both of these markets predicted the eventual election outcome with a higher degree of accuracy than the polls. These results suggest that prices in these prediction markets were efficient in the sense that they accurately reflected the true outcomes of these two elections.

To explain how a market made up of biased traders can result in efficient prices, Forsythe et al. (1992) proposed the “marginal trader hypothesis.” They conjecture that there is a small group of “marginal traders” whose actions are not dictated by their political preferences. Furthermore, these traders are able to move the market toward the correct prices. Marginal traders are identified as those whose limit orders are near the market. Their trading behavior appears to be free of the biases of the average trader in the market. Oliven and Rietz (2004) define marginal traders as “market makers” who submit limit orders which are executed in the market. Other traders, labeled “price takers”, generally only execute market orders at prices set by the market makers. Compared to price takers, market makers are significantly less likely to make pricing mistakes. Forsythe et al. (1999) suggest that the transactions of these marginal traders enable the markets to function efficiently.

There are two potential problems with the “marginal trader hypothesis.” The first is the amount of information that a marginal trader would need to be an effective arbitrager between the two camps of biased traders, one supporting each of the opposing candidates. Marginal traders would have to know the degree of other traders’ biases towards their favored candidate and the proportion of biased traders in each camp (Forsythe et al., 1992). As noted by Forsythe et al. (1992: 1157), it is unlikely that marginal traders have such insight into the nature and numbers of other traders.

The second and, perhaps, more important limitation is the incompleteness of the concept. Marginal traders seem to be able to take advantage of the biases of other traders as well capitalize on their pricing mistakes. However, the “marginal trader hypothesis” is silent on how traders “know” the outcome of the election before it occurs. This is a crucial shortcoming. Prediction markets are usually organized as peer-to-peer trading systems wherein all trades of individual contracts occur between traders. A pricing mistake by one trader will, if acted upon, enrich another trader. By rectifying pricing errors by others, marginal traders should earn higher returns, on average. There is no requirement for such traders to have better information that would result in accurate prices in the market.

However, it does appear that marginal traders do have insight into the outcome of the future event. In the cases of political prediction markets, the marginal traders appear to know the voting results in advance of Election Day. Consequently, their trading activity moves prices in the market in the correct direction. The key question is: How do marginal traders come into possession of such valuable information?
One possible explanation is that some traders have superior private information about the election outcome. The mechanism of a real money prediction market allows traders with superior private information to benefit monetarily by trading with other, less well informed participants. In fact, the importance of traders with superior private information to the effective functioning of prediction markets is a central point in discussions of prediction markets (e.g., Forsythe et al., 1999; Kambil and van Heck, 2002). While it is very difficult to refute the claim that marginal traders possess superior information, we propose a simpler alternative explanation. We suggest that traders in prediction markets can aggregate and improve upon publicly available information to generate more accurate forecasts. Specifically, when publicly available information exhibits a known bias when used to forecast future outcomes, traders will “de-bias” the public information. The resulting prices will provide a superior forecast than is available from the public information.

2. THE ROLE OF PRIVATE AND PUBLIC INFORMATION IN PREDICTION MARKETS

The value of using prediction markets as an information aggregation mechanism lies in their expected ability to incorporate both private and public information into a single summary statistic: price. Clearly, traders in prediction markets have access to both public and private information. However, much of the discussion of the value of prediction markets centers around the private information traders bring to the market. This may be due to the nature of the seminal experimental research on the functioning of markets as information aggregation mechanisms.

This research emphasizes the ability of markets to correctly disseminate private information. In one set of experiments, the focus was the dissemination of private information from informed to uninformed traders (Plott and Sunder, 1982). In another set of studies, individual traders were provided incomplete private information (Plott and Sunder, 1988). Across all traders, the aggregate of the private information revealed the true state of nature. Through trading, the market participants revealed their own private information and eventually all traders shared the same information (Plott and Sunder, 1988).

While these studies were critical building blocks to our understanding of how markets aggregate information, their primary focus was on traders’ private information. However, traders in prediction markets may have access to publicly available information in addition to their private sources. By publicly available information, we are referring to information outside of the market and its participants. Once trading occurs, some private information is revealed and becomes publicly observable. In this study, we focus on external information that may be used by traders in a prediction market.

Often, much of the information available to traders consists of publicly available information. Consider, for example, a trader seeking to predict the price of a stock in the near future. In addition to past prices, a trader has access
to analysts’ forecasts of earnings, disclosures of stock transactions by company officers, etc. For some prediction markets, traders have access to public information regarding the future outcome that is the focus of the prediction market. In a prediction market for sports contests, traders may have access to predictions from experts as well as odds information from Las Vegas sports bookies. In political prediction markets for U.S. elections, traders usually have access to pre-election polls. These publicly available sources of information are the focus of our empirical study.

In some situations, there is a substantial, directional bias in the information that is publicly available to traders. If traders know of the bias, they can use this information to their advantage. Since there are monetary rewards for traders when market prices mirror the actual outcome of a prediction market, it is in the traders’ best interest to move prices in the direction that corrects the bias. As a result, the prediction market should be consistently more accurate than the public information. This leads to our main hypothesis.

**Hypothesis:** If the public information available to traders has a substantial bias, the forecasts generated by prediction markets will be more accurate than those generated by the public information.

In the next section, we discuss the presence of a substantial bias in pre-election polls for U.S. state and local elections.

### 3. THE INCUMBENT RULE FOR U.S. STATE AND LOCAL ELECTIONS

For state and local elections, traders have public information in the form of pre-election polls that can serve as a forecast of the vote shares candidates will achieve on Election Day. In the U.S., a typical political poll in a two-candidate race asks respondents a question of the form, “If the election were held today, would you vote for candidate A or candidate B or are you undecided?” The respondents are usually likely voters (Crespi, 1988). The resulting proportions of respondents that would vote for candidates A or B are often portrayed as a forecast of the outcome at the ballot box (Kou and Sobel, 2004).

One potential source of bias in this publicly available information is the presence of undecided voters. On Election Day, all voters cast their ballot for one of the candidates. In order to use a poll as a forecast, one must decide how to handle the undecided voters. Most pollsters interviewed by Crespi (1988: 22) allocate the undecided voters proportionally to their current vote share. While easy to implement, this approach ignores actual voter behavior in U.S. state and local elections when an incumbent is running for re-election.

Panagakis (1989) examined the results of 155 state and local polls – primarily from the late 1980’s – all from incumbent re-election campaigns. He found that the majority of the undecided vote went to the challenger in
82% of the cases. The phenomenon of undecided voters splitting in favor of challengers was dubbed, “The Incumbent Rule.” The reasoning behind the Incumbent Rule is that in re-election situations, the undecided voters are undecided about the incumbent whose record is well known. Consequently, many end up voting against the current office holder. There are exceptions for short-term incumbents which have not built up a sufficient track record or well-known challengers who held comparable office.

As a forecast of vote shares on Election Day, polls for incumbents and challengers have very different properties. Panagakis (1989) found that about 50% of the challengers gained more than 10 points over their final poll vote share. In contrast, the final poll results for incumbents were within 4 points of the actual vote share in more than 60% of the cases. These results suggest pre-election polls are more accurate when predicting the incumbent’s vote share while the vote share of the challenger seems to be consistently understated.

There was a follow-up study by Gruca (1994) that analyzed a separate set of 138 polls from gubernatorial, U.S. senate and U.S. house contests in 1990 and 1992. He found that the majority of the undecided vote went to the challenger in 67% of the cases. Like the earlier results, 62% of the challengers gained 10 or more points between the last poll and Election Day. Also, a majority of the poll projections for the incumbents were within ± 4 points of the actual vote share.

For this present study, we augmented the Gruca (1994) data set with an additional set of 43 polls for a total of 181 observations from 1990 and 1992. Only final media-sponsored polls were included in the sample. For each poll in this sample, the forecasting error was computed as the difference between the vote share on Election day and the vote share of the candidate in the final poll. If the forecasting error is positive, it implies that the candidate received a higher proportion of votes on Election Day than in the final poll. The cumulative distribution of the forecasting errors is presented separately for incumbents and challengers in Figure 1.

This analysis of the polls from these two years (1990 and 1992) for state and local elections shows that for 84% of the observations, the challenger gained four or more points between the final poll and Election Day. For incumbents, the actual vote share for incumbents was within 4 points of the final poll vote share in 48% of the cases. Both of these results are consistent with Panagakis’ (1989) results.

The distribution of forecasting errors suggests that the publicly available information from pre-election polls regarding the vote shares of challengers is substantially biased downwards. Therefore, we would expect that the forecasts for challengers from prediction markets in U.S. local and state elections would be more accurate than the publicly available information from polls. At the same time, the publicly available information regarding the vote shares of incumbent candidates has comparatively little bias in either direction. Therefore, we expect that forecasts from prediction markets will probably be no more accurate than the forecasts from polls.
We tested these conjectures using the results of seven prediction markets conducted by the Iowa Electronic Market (IEM) in 1994. The IEM is a small scale, real money futures market operated by business school faculty at the University of Iowa since 1988. In contrast to the original Iowa Political Stock Market in 1988 which was limited to academic traders within the state of Iowa, trading in the 1994 election markets was open to traders world-wide.

**Prediction market description**

In these markets, traders bought and sold contracts whose payoff was determined by the share of the popular vote in a given election for a particular candidate. The payoff structure for these contracts was linear, i.e. $1 times the share attained by the respective candidate. The total payoff across all contracts sums to $1. For example, in the market for the 1994 Pennsylvania Senate race, there were three contracts. One contract was associated with incumbent Senator Wofford, one with the challenger Rick Santorum and one contract was designated for the rest of the field (all other candidates). Each contract had a payoff value equal to $1 times the candidate’s share of the popular vote in the November 1994 election.

Participants in IEM prediction markets act as both buyers and sellers. The exchange of contracts occurs via a computerized, anonymous double auction. To buy a contract, a trader can execute a market order and buy at the current best price available (lowest ask from another trader) in the market. Alternatively, the trader can submit a limit order. This would include an offer to buy (bid) at a higher price and a time limit on the offer. [The corresponding process may be used to sell or offer to sell a contract.] The limit orders (bids/asks)
are queued by price and submission times. Traders can view the best prices in each queue. Prices in these markets are solely the result of trading activity between individual traders who remain anonymous to each other.

Traders may acquire contracts from the market in a bundle consisting of one of each of the contracts in the market. Traders can purchase bundles from or sell bundles to the IEM exchange at any time for the guaranteed liquidation value of $1. Selling bundles allows the supply of contracts to vary without contaminating the individual contract prices.

At the end of the market, contracts are liquidated based on the official election returns. For a contract associated with a given candidate, there is an exchange of $1 * popular vote share for that candidate. All money invested in the market is returned to traders and there are no transaction fees. Individual trading accounts are limited to $500.

Results

Our study consists of seven different prediction markets. Some details of the contests are listed in Table 1.¹

To determine the accuracy of these prediction markets, we used the last trade prices in the market on the night before the election (Berg et al., 2007; Forsythe et al., 1992). As in prior research on political prediction markets, we normalized these prices to sum to one.

For each political contest, we obtained the final pre-election polling data for each media organization that sponsored a unique poll. We computed the mean absolute deviation for each incumbent and challenger using the prediction market vote shares and the data from the final polls. The results for incumbents are presented in Figure 2.

The average MAD for the seven incumbents was 2.52 for the IEM forecasts and 4.25 for the final polls. This difference is not statistically significant (paired comparison t = −1.24, p < 0.23 for a two-tailed test). This should be expected since there is comparatively little bias in the polling information about the vote shares of incumbents.

<table>
<thead>
<tr>
<th>Year</th>
<th>State</th>
<th>Office</th>
<th>Incumbent</th>
<th>Vote Share on Election Day</th>
<th>Challenger(s)</th>
<th>Vote Share on Election Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>1994</td>
<td>New Jersey</td>
<td>Senator</td>
<td>Lautenberg</td>
<td>0.50</td>
<td>Haytaian</td>
<td>0.47</td>
</tr>
<tr>
<td>1994</td>
<td>New York</td>
<td>Governor</td>
<td>Cuomo</td>
<td>0.45</td>
<td>Pataki</td>
<td>0.49</td>
</tr>
<tr>
<td>1994</td>
<td>Pennsylvania</td>
<td>Senate</td>
<td>Wofford</td>
<td>0.47</td>
<td>Santorum</td>
<td>0.49</td>
</tr>
<tr>
<td>1994</td>
<td>Texas</td>
<td>Governor</td>
<td>Richards</td>
<td>0.46</td>
<td>Bush</td>
<td>0.53</td>
</tr>
<tr>
<td>1994</td>
<td>Texas</td>
<td>Senate</td>
<td>Hutchinson</td>
<td>0.61</td>
<td>Fisher</td>
<td>0.38</td>
</tr>
<tr>
<td>1994</td>
<td>Utah</td>
<td>2nd CD</td>
<td>Shepard</td>
<td>0.36</td>
<td>Wadoltz</td>
<td>0.46</td>
</tr>
<tr>
<td>1994</td>
<td>Utah</td>
<td></td>
<td>Cook</td>
<td></td>
<td></td>
<td>0.18</td>
</tr>
<tr>
<td>1994</td>
<td>Virginia</td>
<td>Senate</td>
<td>Robb</td>
<td>0.46</td>
<td>North</td>
<td>0.43</td>
</tr>
<tr>
<td>1994</td>
<td>Virginia</td>
<td></td>
<td>Coleman</td>
<td></td>
<td></td>
<td>0.11</td>
</tr>
</tbody>
</table>

Table 1
Political Contests Associated with 1994 IEM Prediction Markets

226
The results for challengers are presented in Figure 3.
For challengers, the results are clear. The prediction markets provided a more accurate forecast of vote share for all 9 challengers (two contests had multiple challengers). The MAD for the IEM forecasts was 2.7 compared to 7.6 for the final polls. This difference is significant (paired comparison $t = -5.07, p < 0.001$ for a two-tailed test). These results suggest traders

![Figure 2. Comparison of forecast errors.](image1)

![Figure 3. Comparison of forecast errors.](image2)
were able to use the substantial downward bias in the publicly available information on the vote shares of challengers (i.e. pre-election polls) to drive prices in the prediction market in the correct direction. Therefore, we find strong support for our study hypothesis.

**Comparisons with Proportional Allocation of Undecided Voters**

To better understand how traders in these prediction markets improved upon the publicly available information, we generated an alternative set of election predictions from the polling data for our seven races. Researchers in forecasting (e.g Clemen, 1989) suggest that combining multiple forecasts should result in more accurate results than relying on a single forecast. As noted above, the typical method to generate vote forecasts from polling data is to allocate the undecided vote proportionately among the major candidates. In addition to its use by polling firms, this method of allocating undecided voters is employed in prior research on a prediction market for 1988 U.S. Presidential election (Forsythe et al., 1992) and more recent research on using combinations of forecasts to predict the 2004 U.S. Presidential election (Cruzan et al., 2005).

We applied this normalization to all of the polls for each contest. We then computed an average vote share for the incumbent and challenger(s). A comparison of the absolute deviations from the normalized poll results and the IEM forecasts is presented in Figures 4 and 5.

Using the normalized poll results improved the forecasts for seven of the nine challengers. At the same time, the normalized poll results provide worse forecasts than the raw polls for four of the seven incumbents. Overall, the IEM prediction markets provided a more accurate forecast than the normalized poll results for
five of the seven incumbents and eight of the nine challengers. To ascertain the overall effect of using the normalized poll results as forecasts, we computed the MAD by contest across the seven markets. The MAD for the forecasts using a proportionate split of the undecided voters (by contest) is 4.2 compared to 2.5 for the IEM prediction markets. This difference is statistically significant \( t = -3.19, p < 0.02 \) for a two-tailed paired comparison t-test. These results suggest traders do more than merely mechanically transform (e.g. a proportionate allocation of undecided voters) the publicly available data and embed that information into prices in the market. By understanding and correcting the bias in pre-election polls for local elections in the U.S. ("the Incumbent Rule"), traders in these markets improve the publicly available information.

5. DISCUSSION AND CONCLUSIONS

The research reported here makes important contributions to our understanding of prediction markets. First, we provide a testable conjecture – public information bias – that determines when prediction markets will provide more accurate forecasts than the public information available to traders. We confirm that the presence of the bias in the public information available to traders influences the relative accuracy of prediction market forecasts.

Second, our study of bias in the prediction of vote shares for challengers in local and state elections in the U.S. provides an interesting alternative explanation for the relatively poor performance of political prediction markets outside the United States. In a review of 25 such prediction markets,
Bruggelambert (1999) finds, in many cases, the forecasts generated by prediction markets were not more accurate than those available from pre-election polls. While researchers have examined such explanations as the type of market, number of traders, number of low share candidates (or political parties), etc., we offer a testable alternative based on the biases present in the publicly available pre-election polling data.

In the U.S., it is very rare that a media polling organization reports an allocation of undecided voters. Consequently, the known bias in state and local election polls allows prediction market traders to move contract prices for challengers in the correct direction. It is not known if a similar bias exists in polling data outside of the United States. In addition, the methodologies used in other countries with respect to undecided voters may be very different than those in the U.S., resulting in more accurate predictions by polling organizations. A useful area for future research would be to test for the presence of biases in polling data in other countries.

One of the limitations to this study is that public information bias does not explain why prediction markets succeed (or fail) in every circumstance. For example, there is considerable disagreement whether the “Incumbent Rule” applies in Presidential elections. In addition, there has been little research on the disposition of undecided voters in elections that do not feature an incumbent. Unfortunately, a comparable test of our hypothesis using data from IEM forecasts of Presidential markets is quite a way away. While the IEM has been used to predict the popular vote shares in U.S. presidential elections since 1988, there are only 5 sets of observations. Of these, only 3 featured an incumbent facing re-election. The next opportunity to collect such data may come in 2012.

More and more organizations are becoming interested in using prediction markets to aggregate information from public and private sources. This rising interest comes with a challenge for researchers. In order to determine whether a prediction market would be effective in a given situation, decision makers need a comprehensive and predictive theory elucidating the conditions under which prediction markets will perform better than the second best options whether these are public polls, Delphi forecasts from experts, econometric models, etc. Without such a framework, businesses may be hesitant to undertake the experimental applications of prediction markets that will expand beyond their current focus on elections, movies and sporting events. We hope that this study on the influence of public information bias is a first step on this important path to a general theory of prediction market efficiency.

ACKNOWLEDGEMENTS

Previous versions of this paper were presented at the Workshop on The Growth of Gambling and Prediction Markets, the ISBM Conference on Prediction Markets in Marketing: Issues, Challenges and Research Opportunities, the DIMACS Workshop on Markets as Predictive Devices, and the 17th Annual AMA Advanced Research Techniques Forum. The authors thank the participants for their insights which have helped improve this research.
NOTES

1. The IEM also organized a prediction market for the 1994 Arizona Senate contest. However, this contest did not include an incumbent candidate.

2. For a recent comparison between forecasts generated by polls and prediction markets for U.S. presidential contests, see Erikson and Wlezien (2007).

REFERENCES


Irving Crespi Pre-Election Polling (New York, Russell Sage Foundation, 1988).


