PREDICTION MARKETS:
HOW THEY WORK AND HOW WELL THEY WORK

VINCENT JACOBS
Master of Economics

Supervisor: Prof. Dr. P. VAN CAYSEELE
Tutor: D. DE SMET

-2009-
Vincent Jacobs

Prediction Markets:
How They Work and How Well They Work

Abstract:

Prediction markets are markets where participants trade contracts whose payoffs depend on unknown future events. In recent years, these markets have been used to forecast many different events such as the outcome of political elections, the sales of a company or the effect of a new government policy. Firstly, this paper looks at the theoretical literature on prediction markets with particular interest in the design considerations when setting up a market. Secondly, this paper investigates the performance of various prediction markets in the run-up to the 2008 U.S. presidential elections with particular attention paid to the performance of the market relative to polls.

Supervisor: Prof. Dr. P. Van Cayseele
Tutor: D. De Smet

-2009-
Acknowledgments

Writing this thesis has certainly been my largest undertaking of the last four years and ended up taking quite a bit longer than expected. In the end, it all came together nicely and I’m very satisfied with the result. However, I could not have done it alone.

Firstly, I would like to thank my parents, Luc Jacobs and Karine Op de Beeck, for their continued support during my studies and for correcting an earlier version of this text.

I am very grateful to my tutor, Dries De Smet, for his helpful comments and continued patience, even when I had little work to show. I would also like to thank Prof. Dr. Van Cayseele for being my supervisor.

Furthermore, I would like to thank Ken Kittlitz of the Foresight Exchange, Tom Griffiths of Hubdub.com and Emile Servan-Schreiber of NewsFutures Inc. for their help in providing data used in the second part of this paper.

Lastly, I would like to thank my friends and the residents of Junior House in Leuven for their constant questions about how my thesis was coming along.
Contents

General Introduction 1

1 Prediction Markets in Theory 2

1.1 Introduction: what is a prediction market? 2

1.2 Practical uses of prediction markets 3

1.2.1 Origin of prediction markets 3

1.2.2 Political stock markets 3

1.2.3 Decision markets 4

1.2.4 Corporate prediction markets 5

1.2.5 Sports prediction markets 6

1.2.6 Other fields of interest 7

1.3 Online prediction markets 8

1.4 How prediction markets work 9

1.4.1 Historical perspective 9

1.4.2 How prices are formed 11

1.5 Design of prediction markets 13

1.5.1 Types of contract 13

1.5.2 Trading mechanisms 16

1.5.3 How contracts are introduced to the market 20

1.5.4 The motivation to play (real-money vs. play-money) 21

1.5.5 The population of traders 28

1.5.6 Other market microstructure considerations 29
General Introduction

People have always been interested in forecasting the future, whether it is predicting which politician will win an election or which football team will win Sunday’s game. The problem is how to make an accurate prediction of what will happen in the future. There are many things that could influence the outcome of the event and no, one person, is able to know all the relevant information to make a prediction as accurate as it could potentially be. We need a way of combining the knowledge of many individuals to make the prediction.

This notion that groups together can make accurate predictions, more accurate than any of its individuals could, was popularized by James Surowiecki in the book *The Wisdom of Crowds*. One example he gives is from the scientist Francis Galton; at a livestock show, visitors could wager on the weight of an ox to win prizes, Galton found that the average guess of 787 participants was 1,197 pounds while the correct weight was 1,198 pounds, nearly exactly the same. The group combined its knowledge of the weight of the ox by each submitting a wager and together they almost exactly predicted the correct weight, much better than any of the group’s individuals could have.

In this paper we will study how the knowledge of many individuals can be aggregated to make accurate forecasts for future events using *prediction markets*. This paper proceeds as follows; in the first part it is shown how prediction markets work and various aspects of the markets are discussed in theory with particular attention paid to design considerations. The second part looks at prediction markets in practice, looking at how they performed in predicting the outcome of the 2008 U.S. presidential elections.
Part 1

Prediction Markets in Theory

1.1 Introduction: what is a prediction market?

Prediction markets are markets where participants trade contracts whose payoffs depend on unknown future events. These have also been called ‘information markets’, ‘decision markets’ or ‘idea futures’. Participants buy and sell contracts based on their current prediction of the future, they can “put their money where their mouth is” \cite{Hanson1999}. The current price of a contract can be interpreted as the market’s estimated probability of that event occurring.

The interest in prediction markets comes from the idea that a market is capable of incorporating all available information about an event into its prices. This would allow a prediction market, in theory, to provide more accurate predictions of future events than other forecasting methods. This is because participants are investing real-money and have an economic incentive to make the best possible predictions. In practice, we see that prediction markets allow groups of people to make predictions at least as accurately as the best individual in the group without having to know who that individual is \cite{Surowiecki2004}.

\cite{Wolfers2004} suggest that the power of prediction markets can be explained in three parts: they provide incentives for truthful revelation and information discovery and provide an algorithm for aggregating opinions. The rest of this part of the paper evaluates the literature on prediction markets to learn how these markets can be designed to function optimally.
1.2 Practical uses of prediction markets

This section briefly summarizes the wide range of applications for prediction markets and the evidence for their accuracy. The goal here is not to go into great detail but to give an overview. For more details, the reader is referred to the references or, for a longer list of research into prediction markets, to [Tziralis and Tatsiopoulos (2007)].

1.2.1 Origin of prediction markets

While prediction markets in their current form are a relatively new development, investments whose value depends on a future event are certainly not. The price of a share in a company that trades on the stock market is dependent on its expected future free cash flow. A futures contract provides a forecast of what the value of the underlying asset will be at the expiry date of the contract. The current price of inflation-indexed bonds effectively tells us the market’s prediction of future inflation, while sports betting tells us the expected probability that a team or individual will win a sporting event. In all these cases the market’s price shows us what its traders expect to happen in the future, this predictive quality is exactly what prediction markets attempt to foster. The main difference between most investment markets and prediction markets is that investment markets are designed as a resource allocation mechanism and provide information aggregation as a consequence of this. A prediction market, on the other hand, is designed primarily for information aggregation.

1.2.2 Political stock markets

One of the earliest, and still most popular, uses of prediction markets is forecasting the outcome of political elections.¹

The most famous and longest running political stock market is the Iowa Electronic Markets (IEM). These have been used to predict the outcome of U.S. presidential elections since 1988 [Forsythe et al., 1992]. The prices of these markets have been shown to be accurate predictors of the outcome of elections both in absolute terms and relative to polls, in both the short-term [Berg, Forsythe, Nelson and Rietz 2008] and over longer horizons [Berg, Nelson and Rietz 2008].

¹The second part of this paper looks at an empirical investigation into the performance of prediction markets in the 2008 U.S. presidential election so political stock markets will be discussed in more detail there.
Prediction markets have also shown successes in forecasting election outcomes in other countries. Forsythe et al. (1995) study a prediction market on the 1993 Canadian federal election and show that the market did well at predicting the vote-shares of the major parties. Bohm and Sonnegard (1999) show that a prediction market made more accurate predictions than polls in the 1994 Swedish referendum over whether they should join the European Union. Similar markets have been set up in other countries such as Germany (Brüggelambert, 2004; Beckmann and Werding, 1996) and the Netherlands (Jacobsen et al., 2000). The results usually show that the markets are successful but less so than the American IEM, due to differences in the electoral system.

Rhode and Strumpf point out that political stock markets are not a new phenomenon. They discuss presidential betting markets in the U.S. from 1868 to 1940 (Rhode and Strumpf, 2004) and also in 16th century Italy and 18th and 19th century Britain (Rhode and Strumpf, 2008a). For a more detailed look at political prediction markets and our analysis of the 2008 U.S. presidential election markets see the second part of this paper.

1.2.3 Decision markets

The idea of using prediction markets to help make decisions started with Hanson’s seminal article on decision markets (Hanson, 1999). He suggested that markets allow for better sharing of information and can thus be used to better predict the consequences of decisions. The example he gives is a market to predict the effect of allowing the carrying of hidden guns on the crime rate. By allowing many traders with different information sets to trade the contract, their information can be aggregated to make better informed decisions. Prediction markets can thus help governments to make better policy decisions. Hanson even goes so far as to suggest a new form of government (‘futarchy’) where prediction markets are used to determine policy (Hanson, 2000).

Berg and Rietz (2003) discuss the use of prediction markets as decision support systems. They discuss the use of conditional contracts (contracts which are conditional on other events occurring) to help the Republican Party choose the strongest candidate for the 1996 U.S. presidential election. Using contracts to predict the vote-share that the party would receive in the election, conditional on each candidate being nominated, allows them to see which candidate the market would expect to be the most successful. The market showed that Powell was a strong candidate to run against Clinton while Dole was a weak candidate to run against Clinton. The
Republicans ended up nominating Dole and losing the election. The authors posit that they might have been more successful had they chosen the strongest candidate based on the market prices.

The possibilities are endless, data from prediction markets can help political parties choose the best candidate, help governments make better policy decisions and (as we will see in the next section) help businesses make better decisions and uncover useful information about their operations.

1.2.4 Corporate prediction markets

Prediction markets also have corporate applications and have been used by many companies around the world.\textsuperscript{2} \cite{SpannSkiera2003} discuss designing prediction markets that can be used to make business forecasts. An important difference from the previously mentioned markets is that the market is designed to aggregate private knowledge instead of public knowledge. For example; in political markets, the news drives the prices, this is public information and traders are not expected to have insider knowledge to determine the prices. In a business situation this is not the case; traders who trade a contract will have inside information about the question at hand and the market will try to aggregate this information. The market encourages traders to reveal information (good or bad) that they have about a project but might be unwilling to reveal in other situations, since negative news is usually not well received.

\cite{Ortner1997, Ortner1998} used prediction markets at Siemens Austria asking the question of whether a software development project would be finished on time and (if not) how late it would be. In the end, the project was not finished on time. The market had predicted this while traditional project management techniques continued to show that they would be able to meet the deadline.

\cite{ChenPlott2002} discuss the use of prediction markets inside the Hewlett-Packard Corporation (HP) to forecast the sales of specific products. They found that the market predictions outperformed the official HP forecasts by being closer to the actual outcome in 6 out of 8 events. Also interesting is that the market’s predictions about the direction that the actual outcome would be relative to the

\footnote{Firms whose prediction markets have been mentioned in the public domain include: Abbott Labs, Arcelor Mittal, Best Buy, Chrysler, Corning, Electronic Arts, Eli Lilly, Frito Lay, General Electric, Google, Hewlett-Packard, Intel, InterContinental Hotels, Masterfoods, Microsoft, Motorola, Nokia, Pfizer, Qualcomm, Siemens, and TNT (\cite{Cowgill2009} p.1).}
official forecast (the official forecast would be too high or the official forecast would be too low), were also accurate.

A different use of corporate prediction markets is discussed by Cowgill et al. (2009). They report on the internal markets of Google which have been used to make forecasts relevant to their business; for forecasting demand (such as “How many users will Gmail have?” by a certain date) or internal performance (such as “Will this new product be ready on time?”). They report that the markets were reasonably efficient (with some biases) but more interestingly they use the information from the markets to study the information flow within the business. They find that the trades of employees who are in physical proximity are correlated and that the work history of an employee and also their organizational proximity play a role. They thus use data from their prediction market to confirm the significant role of micro-geography in corporate information flows.

Staying with Google, another interesting application is shown by Berg et al. (2009). They use a prediction market to predict the market capitalization of the company at the close of the first day after its initial public offering (IPO). This is important for a company since a higher price the day after the IPO implies that the IPO price was set too low and that the company could have made more money. They find that the market’s forecast exceeded the actual value by 4.0% while the actual IPO price was 15.3% lower than the market capitalization at the close of the first day. They conclude that prediction markets could be used to help gather information before an IPO to help set its price.

1.2.5 Sports prediction markets

Betting on sports events has always been a popular pastime. Traditionally, this has functioned with a market maker; someone who sets the prices in order to have equal amounts of bets on each outcome. But, with the increased popularity of prediction markets, the concept of a market has also been applied to sports betting, allowing traders to trade contracts dependent on the outcome of a sporting event with each other. The difference is that prices are no longer set by a market maker but instead by the prices at which individual traders are willing to buy or sell a contract. An added advantage is that traders can sell contracts, which they have previously bought, to make a profit or a loss before the event has even occurred. Since these markets tread the fine line between investing and gambling we will not spend too much time discussing them.
A study by Smith et al. (2006) shows that ‘betting exchanges’ (prediction markets for sports, such as Betfair\(^3\) or the now defunct TradeSports\(^4\)) where traders trade with each other instead of with a market maker, have lower transaction costs for consumers and that the favorite-longshot bias tends to be lower, allowing us to assume that the prices will be set more efficiently.

Chen et al. (2005) look at prediction market forecasts for 210 National Football League (NFL) games in 2003 from two different markets (TradeSports and NewsFutures\(^5\)) and compare these to the aggregated predictions of multiple experts. They find that the predictions of the markets were as accurate as those of the pooled experts when compared at the same point in time ahead of the games.

Likewise, Pennock et al. (2001) discuss the prediction market Formula One Pick Six\(^6\) where traders simply pick the top six drivers in each Formula One race and receive points depending on how accurate their predictions were. They find that the market’s predictions are equally good or better than the official betting odds for the same race.

### 1.2.6 Other fields of interest

To show how wide the range of applications for prediction markets is, we now briefly mention a few more, less conventional, uses.

Spann and Skiera (2003) look to forecast the gross box-office returns of U.S. movies on the Hollywood Stock Exchange (HSX)\(^7\). This is an important forecasting situation since there are high costs in developing a new movie and a high uncertainty about how successful it will be. This high risk makes accurate predictions very useful for deciding which movies to produce and which not to produce. Their market data from 152 movies tries to forecast the total U.S. box-office revenue during the first four weeks that the movie is released and does a relatively good job. The market has a mean absolute percentage error of only 30.96%. Comparing the results of the market to the judgements of two self-proclaimed experts, they find that the market (and models based on market data) perform as well or better than these experts.

Pennock et al. (2001) also, independently, find that the HSX forecasts are accurate at predicing box-office success and also look at the market’s ability to predict

---

\(^3\) [http://www.betfair.com/](http://www.betfair.com/)

\(^4\) TradeSports was the sister website of Intrade and closed in 2008: [http://tradesports.com/](http://tradesports.com/)

\(^5\) [http://us.newsfutures.com/](http://us.newsfutures.com/)

\(^6\) [http://motorsport.com/compete/p6](http://motorsport.com/compete/p6)

\(^7\) [http://www.hsx.com/](http://www.hsx.com/)
Academy Award winners in 2000. The market offered contracts for every nominee in each category and the authors found that every nominee with the highest final price in their category on the HSX also ended up winning the award, showing again the accuracy of the market’s predictions.

Researchers at Pennsylvania State University have even set up prediction markets to forecast the weather (daily high and low temperatures) and found that the market’s predictions were as accurate as those of major forecasting services. Furthermore, Nelson et al. (2006) discuss the use of prediction markets to predict influenza outbreaks.

Lastly, we mention a prediction market which received less than favorable coverage. The Policy Analysis Market (PAM) was a prediction market proposed by DARPA which was designed to allow trade in contracts based on various geopolitical risks such as the economic health of a country or the stability of a government (see Polk et al. (2003) for details of the market). However, it was perceived that the market would encourage trading in events such as assassinations and terrorist attacks so was quickly denounced as selling “terrorism futures”. The project was cancelled the day after it was announced and lead to the resignation of the head of the unit responsible for the market. It was also feared that interested parties could profit from the market by influencing the events or mislead the researchers by manipulating the market in the wrong direction, in hindsight this was shown to be unlikely (see Looney (2004) for more information about the markets suggested design and the events that led to its cancellation). Here we see a market that could have been very useful for the decision makers but that was closed down for ethical reasons.

1.3 Online prediction markets

In practice, there are many prediction markets functioning on the internet with great success. Table I gives an overview of eight popular online prediction markets which offer contracts on a variety of events: the Iowa Electronic Markets, Intrade, Betfair, Bet2Give, NewsFutures, Hubdub, the Foresight Exchange and Inkling Markets. Data from these eight markets will be used in the second part of this paper when looking at prediction markets in the run-up to the U.S. presidential elections of 2008.

Nelson et al. (2006)

Polk et al. (2003)

Looney (2004)

The Defense Advanced Research Projects Agency (DARPA) is a think tank of the United States Department of Defense: http://www.darpa.mil/
As can be seen from the table, there are significant differences between these markets: whether they use real-money or not, the purpose of the market, the incentives offered to trade, the trading mechanism and various other factors. The theoretical effects of these differences will be discussed in this part of the paper and it will be tested if these are true in practice in the second part.

1.4 How prediction markets work

1.4.1 Historical perspective

The idea that markets can efficiently aggregate large amounts of information, that is distributed over a large number of individuals, goes back to Hayek (1945). He argued that the price of a market performs the task of aggregation much better than any central planner could since the sum of the information available to individuals is always greater than the information that one expert can possess. He set out the idea that the market, through price signals, was the best way to aggregate the information that all the individuals have and transmit this information to the decision makers.

The ideas of Hayek were formalized by Fama (1970), in what we now know as the efficient market hypothesis which states that markets are informationally efficient meaning that prices at any time reflect all the information available about future events. It also states that prices react quickly when new information becomes available. This means that prices at any time are equal to the expected value of the asset and that no new information can be combined with the market prices to improve on the predictive accuracy of the prices. However, Grossman and Stiglitz (1980) point out the paradox in this; if prices already reflect all relevant information then there is no incentive for individual traders to collect information on which the prices are set. In experimental settings however, Smith (1982) finds that Hayek's hypothesis tends to hold.

Muth (1961) modeled the expectations of individuals as being rational, this assumes that their expectations are (on average) correct. They use all the information available to them and their predictions are not systematically biased.

The idea of (and enthusiasm for) prediction markets relies on this idea that markets can be informationally efficient and thus accurately reflect all available information. If this is true, the market price (the market’s prediction) will be the most accurate predictor of the future event. Under which conditions the market is efficient, its
<table>
<thead>
<tr>
<th>Market Name</th>
<th>Iowa Electronic Markets</th>
<th>Intrade</th>
<th>Betfair</th>
<th>Bet2Give</th>
<th>NewsFutures</th>
<th>Hubdub</th>
<th>Foresight Exchange</th>
<th>Inkling Markets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real Money?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Purpose of the market</td>
<td>Research</td>
<td>Financial market</td>
<td>Betting exchange</td>
<td>Encourage charitable donations</td>
<td>Entertainment</td>
<td>Entertainment</td>
<td>Entertainment</td>
<td>Entertainment</td>
</tr>
<tr>
<td>Incentives to trade</td>
<td>Financial gain</td>
<td>Financial gain</td>
<td>Financial gain</td>
<td>Financial gain</td>
<td>leaderboards, donations list, giving to charity</td>
<td>leaderboards, social networking features</td>
<td>leaderboards, player levels, social networking features</td>
<td>leaderboards, leaderboards, social networking features</td>
</tr>
<tr>
<td>How are contracts issued?</td>
<td>traders can buy unit portfolios from the market</td>
<td>matching of buyers and sellers in order book</td>
<td>a back bet is matched with a lay bet in the order book</td>
<td>matching of buyers and sellers in order book</td>
<td>each contract has an opposite: matching of buyers of one contract with buyers of its opposite contract</td>
<td>market scoring rule acts as market maker</td>
<td>matching of buyers and sellers in order book</td>
<td>market scoring rule acts as market maker</td>
</tr>
<tr>
<td>Trading mechanism</td>
<td>continuous double auction</td>
<td>continuous double auction</td>
<td>continuous double auction</td>
<td>continuous double auction</td>
<td>continuous double auction</td>
<td>Hanson’s market scoring rule</td>
<td>continuous double auction</td>
<td>Hanson’s market scoring rule</td>
</tr>
<tr>
<td>Information available to traders</td>
<td>current prices, last prices, historical data</td>
<td>complete order book, historical data</td>
<td>complete order book, historical data</td>
<td>complete order book, historical data</td>
<td>complete order book, historical data</td>
<td>current probabilities, prediction history, names of traders, historical data</td>
<td>complete order book, historical data</td>
<td>current probabilities, prediction history, names of traders</td>
</tr>
<tr>
<td>Contract price range</td>
<td>$0 to $1</td>
<td>$0 to $10 (0 to 100 points)</td>
<td>decimal odds</td>
<td>$0 to $1</td>
<td>$X0 to $X100</td>
<td>0% to 100%</td>
<td>0 FX-cent to 100</td>
<td>0 to $100</td>
</tr>
<tr>
<td>Min/Max investment</td>
<td>min $5, max $500</td>
<td>min $25, max –</td>
<td>min £5, max –</td>
<td>min $5, max $300</td>
<td>receive X$1,000 at registration, no maximum</td>
<td>receive $FX2,000 at registration, no maximum</td>
<td>receive $5,000 (play-money) at registration, no maximum</td>
<td></td>
</tr>
<tr>
<td>Fees</td>
<td>activation charge ($5), no trading fees</td>
<td>variable per trade + expiry fee</td>
<td>commission on winnings</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>2008 U.S. presidential election contracts</td>
<td>WTA Democrat, WTA Republican, VS Democrat, VS Republican</td>
<td>WTA Democrat, WTA Republican, WTA Obama, WTA McCain</td>
<td>WTA Democrat, WTA Republican, WTA Obama, WTA McCain</td>
<td>WTA Democrat</td>
<td>WTA Democrat</td>
<td>WTA Democrat</td>
<td>WTA Democrat</td>
<td></td>
</tr>
</tbody>
</table>

$ = U.S. dollar, £ = Pound sterling, WTA = winner-take-all, VS = vote-share

**Table I: Comparison of various online prediction markets**
prices are efficient, and to what extent this is true in practice will be explored in this paper.

1.4.2 How prices are formed

If the market is efficient, the current price of a contract will be equal to the expected value of the contract at expiration. [Berg, Nelson and Rietz (2008, p.289) show that contracts can be priced using the Capital Asset Pricing Model (CAPM) and/or the Arbitrage Pricing Theory (APT) model. In both cases:

\[ P_t = \frac{E(P_{t+h})}{(1+k)^h} \]  

(1.1)

Where \( P_t \) is the price of the contract on the current date (at time \( t \)), \( E(P_{t+h}) \) is the expected price of the contract on any future date \( (t+h) \) up to and including the day that the contract expires and \( k \) is the required expected return which is the sum of the risk-free rate and the compensation for aggregate risk factors. By design, the risk-free rate is zero. For example, the Iowa Electronic Markets introduce contracts to the market in unit portfolios, these are portfolios which contain all the contracts traded in the market and will always be traded for the sum of the liquidation values. The closing price of a unit portfolio is the same as its purchase price regardless of the outcome of the event. Buying and holding a unit portfolio therefore involves no risk and will earn a return of zero. The holding of cash in the prediction market account will also earn zero interest and carry no risk. There are also no aggregate risk factors since the individual payoffs of the contracts depend on the outcome of the event but the overall sum of payoffs does not. This means that the required expected return is zero \( (k = 0) \) which allows for the simplification of the equation as:

\[ P_t = E(P_{t+h}) \]  

(1.2)

The current price of the contract is equal to the expected price of the contract on any future date. The best forecast for the price at any time in the future is the current price and therefore the current price is also the best prediction for the closing price on the day that the contract expires.

To illustrate, imagine a contract that has a payoff of 100 if an event occurs and 0 if it does not. If the current price of the contract is 60 \( (P_t = 60) \) then the expected price for any date in the future (including the expiry date) is \( E(P_{t+h}) = \text{Probability(event occurs)} \times \text{Payoff(event occurs)} = 0.6 \times 100 = 60 \). The current price can thus be interpreted as the probability of the event occurring.
However, Kou and Sobel (2004) argue that these standard approaches (the CAPM and APT) cannot be used for prediction markets. Firstly, when using these approaches there is usually an underlying asset that is being traded (for example, a stock in a company or a bond) which has an observed current value which is assumed to follow a log-normal distribution. This is not true in a prediction market where the payoff does not follow a log-normal distribution and does not have an observed current value. Their second argument is based on political prediction markets and states that there is a correlation between the results of an election (to which the contract payoffs are linked) and movements in other asset markets such as the stock market. This would imply that equilibrium prices in the prediction market are correlated with prices on the other markets. So, if the standard approaches cannot be used, some assumptions need to be made before the equilibrium prices of a prediction market can be derived.

To derive the equilibrium prices for a vote-share contract in a political election with multiple candidates Kou and Sobel (2004, pp.284-286) make six assumptions about the behavior of the traders and the structure of the market:

1. Investments in the election market are not part of a participant’s overall investment strategy.

2. The utility function $U$ is continuously differentiable and $U' \neq 0$.

3. Participants have access to the same information set $\mathcal{F}_t$ at time $t$ and form identical expectations.

4. The market is in equilibrium.

5. (A) No one participant can influence the market. (B) There are no transaction costs involved in trading. (C) The bid-ask spread is zero.

6. Expectations are rational.

The first assumption is the most important and is also, in the current state of prediction markets, realistic. Making this assumption solves the second problem of correlation mentioned above. This is realistic because there are limits to how much each trader can invest (for example, a maximum of $500 for the IEM, as shown in Table I) and, as mentioned above, the risk-free rate is zero while in general in stock, or other asset markets the risk-free rate is positive, meaning there is little incentive to use prediction markets as part of an overall investment strategy.
If all the assumptions above are met, the market price $P_c(t)$ of a vote-share contract for candidate $c$ at time $t$ can be written as:

$$P_c(t) = E(\theta_c(T) | F_t)$$

(1.3)

Where the market uses all the information available to traders at time $t$ ($F_t$) to make a prediction for the outcome of the election ($\theta_c(T)$) for candidate $c$ on the day of the election ($T$). This is exactly the same conclusion as was reached with the standard approach above but now we know the assumptions that need to be made for it to be valid.

### 1.5 Design of prediction markets

As has been seen previously, prediction markets can be used to make predictions about a wide range of future events and have been successful in doing so in practice. It is known that the design of a prediction market impacts its prediction making ability. This section discusses various aspects of designing a prediction market, focusing on the most important design decisions such as the type of contract, the trading mechanism and whether real-money or play-money is used but also briefly mentioning a few other design considerations. This should give an idea of when a market can be expected to perform well and when it would be expected to perform less well.

#### 1.5.1 Types of contract

On a prediction market, traders buy and sell contracts tied to a particular event occurring. The market writes these contracts, which can then be traded by the traders. The contracts must be specific and must clearly state when the contract ends and what happens then. The main difference in the type of contracts is how the payoff is linked to the event that is being predicted when the contract expires. Wolfers and Zitzewitz (2004) distinguish between three types of contract that are often used: winner-take-all contracts, index contracts and spread contracts.

In winner-take-all contracts, the event that the contract is based on either occurs or it does not occur and the contract that was correct gets ‘everything’ while the contract that was incorrect gets nothing. For example; a sports team wins a game or does not, a politician wins an election or does not, or an economic index is above a certain value at a certain time or it is not. Once the event occurs, the contract
that was correct is worth a certain, pre-agreed, value (usually 100) at which it can then be sold back to the market while the contract that was incorrect is worth zero. The use of winner-take-all contracts is popular in practice because if the prices are set to be between 0 and 100, the current price of the contract can be interpreted as the market’s expected probability of the event occurring.

The second group of contracts are index contracts. With these, the closing value of the contract is indexed to the outcome of an event, for example; the percentage of the vote that a political party receives in an election. The price that the contract closes at will depend on the percentage of the vote the party received. A contract can, for example, trade between 0 and 100 and if the party wins 60% of the vote in the election, the contract will close at 60 (in other words, at one unit of currency times the percentage of the vote). The current trading price can be interpreted as the mean value that the market assigns to the outcome.

The third type of contract uses the techniques of spread betting. Here traders bet on whether the outcome will be above or below a certain value, called the spread. It is best to clarify this with an example: imagine a sports event with two teams and a market maker who sets the spread at 5 points. A trader can then buy a contract for the ‘underdog’ (the team that is least likely to win the event) to win the event or the ‘favorite’ (the team most likely to win the event) to win the event. The contract for the underdog will ‘win’ if the underdog’s score plus the spread is greater than the favorites score, so the contract can win even if the underdog loses the event as long as they lose by less than 5 points. The contract for the favorite will ‘win’ as long as the favorite’s score minus the spread is greater than the underdog’s score, so the contract can lose even if the favorite wins the event as long as they win by less than 5 points. The spread can then be adapted to current information to make each contract equally likely to win. The bet and the payoff are fixed, but the spread changes depending on the current trading situation. The current trading price can be interpreted as the market’s expectation of the median outcome of the event. In the above example, the spread implies that with a spread of 5 points both contracts are equally likely to win.

The use of different contracts allows the same prediction market to predict different aspects of the same event. For an election; a winner-take-all contract can be used to predict who will win the election while an index contract can be used to predict which percentage of the vote a party will receive. A spread contract can be used to determine what the median expectation is of the percentage of the vote that a party will receive. Imagine an (unrealistic) situation where all three contracts for a
political party are currently priced at 60. The winner-take-all contract implies that the party has a 60% chance of winning the election, the index contract implies that the party is expected to receive 60% of the vote while the spread contract implies that the party is equally likely to receive more than 60% of the vote as they are to receive less than 60% of the vote.

**Specification of contracts**

Another aspect of the design of the contract is the exact specification of when the contract expires and at which value. All the possible outcomes have to be accounted for when writing the contract, otherwise contracts may lead to disputes between traders and the market. An example would be a corporate prediction market set up to predict the sales of a certain product in one year. It would be important to specify when the contract expires, how the number of sales will be measured and how the expiration value of the contract is linked to the number of sales. But, it is also important that other, less likely, events are accounted for such as what happens to the contract if the company stops selling the product before the end of the year?

For real-money markets the specification of contracts is of much greater concern than for play-money markets. While for most contracts the outcome is easy to determine, such as sports contracts (team A will win) or financial contracts (an index will be above a certain level), contracts related to current events can often be problematic. A real world example was in 2006 when TradeSports had a contract that would expire at 100 if North Korea would “launch a test missile and it leaves North Korean air space on/before 11:59:59pm ET on 31st July 2006” and 0 otherwise. The source to be used to settle the contract was the U.S. Department of Defense. In early July 2006, the North Korean government claimed to have done such a test and many news sources reported it. However, since the Department of Defense did not confirm the action, the contract closed at 0. Having chosen a different source to judge the contract by, it would possibly have closed at 100 instead.

**User defined contracts**

On some prediction markets, traders are allowed to define their own contracts which can then be traded. The Foresight Exchange allows traders to make their own “claims” which need to be written according to certain guidelines and then approved by a “judge” before appearing on the website and being open to trading. On Hubdub, traders are also able to submit their own “questions” which can then
The questions are settled by the administrators of the market and can be edited or deleted if they do not meet the set guidelines of the market. Both are play-money markets, it would be unlikely to see such user-created contracts on real-money markets since the market could be held responsible if there was a dispute.

### 1.5.2 Trading mechanisms

Once a contract has been defined, it is still very important how this contract will be traded; it is up to the market to determine the trading mechanism used.

[ Luckner (2008) ] identifies four different trading mechanisms: the continuous double auction (CDA), the call auction (CA), dynamic pari-mutuel markets (DPM) and market scoring rules (MSR). These are compared in Table II and are explained individually below. There are three main criteria on which trading mechanisms can be compared: whether they allow for the continuous incorporation of information into their prices, whether they guarantee liquidity and whether there is risk for the market operator in using this mechanism.

### Continuous double auction (CDA)

First is the continuous double auction (CDA), this is the trading mechanism that is used most often in prediction markets. It involves traders wanting to buy a contract being matched to traders wanting to sell a contract using a limit order book. Any trader can submit a market order at any time (the market is continuous) which allows them to buy or sell the contract at the current best price offered by other

---

**Table II:** Comparison of various trading mechanisms (adapted from [Luckner (2008), p.239])

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous information incorporation</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Guaranteed liquidity</td>
<td>No</td>
<td>No</td>
<td>Buying: Yes</td>
</tr>
<tr>
<td>Risk for market operator</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Examples</td>
<td>IEM, NewsFutures, Yahoo! Tech, Foresight Exchange, Buzz Game, Intrade, Bet2Give, Betfair</td>
<td>Passauer</td>
<td>Wahlborse</td>
</tr>
</tbody>
</table>
traders (they can buy at the ask price and sell at the bid price). Traders can also submit limit orders where they specify the price at which they want to buy (sell) at, if this price is lower (higher) than the current best price available, the order is placed in the order book until it can be matched with an incoming market order or expires. Orders are matched according to price and then time priority.

This trading mechanism allows for the continuous incorporation of information. Since orders can be placed and matched at anytime, the current best ask and bid prices will quickly adapt to new information as it becomes available. Orders which are mispriced according to the new information will be cancelled or picked off by other traders. This is only possible if the market is liquid.

One disadvantage of this mechanism is that liquidity depends solely on the number of limit orders that are placed and therefore on the number of traders in the market. This can mean that, if there are few traders, the order book can be empty or if they have very different views on the value of the asset, the bid-ask spread (the difference between the best buy and sell prices) can be large. In practice this does occur. On real-money prediction markets there are popular contracts which have enough liquidity to work well as a CDA but there are also other contracts which are much less popular and where liquidity is thus a concern.

One of the main reasons CDAs are used in practice is that there is no financial risk for the market. Buyers are matched directly to sellers, the market’s only role is in bringing them together. If the market is for-profit they can then charge a fee on each trade. Traders are also comfortable with this mechanism since it is what is commonly used in financial markets and easy to explain to new traders.

**Call auction (CA)**

The second market mechanism is the call auction (CA), this is a periodic market where trading occurs only at specific points in time. Buy and sell limit orders are accepted by the market at any time but are not executed until a predetermined time when the orders are simultaneously executed using a set execution rule. An example would be to use a rule to execute the highest number of trades possible, the market would collect all the orders and at the specified time decide on a price that enables the highest number of orders to be executed.

These have not often been used in prediction markets but have been used in financial markets as an alternative trading mechanism. They are also often used to set the opening and closing prices at stock exchanges by having a call auction at the start.
and end of the day. Another example is crossing networks, were unpriced buy and sell orders are matched at a certain time at a price that is derived from another market.

A disadvantage of a CA is that there is no continuous incorporation of information into prices. Since orders are only matched at certain times, new information will not be incorporated into the prices until the next auction is held. Again, liquidity depends solely on the number of orders placed but the advantage over CDAs is that orders are accumulated until the execution time which can be advantageous for a low liquidity market. It is also more difficult to influence the prices with small trades than in CDAs. Just like CDAs, there is no financial risk for the market.

**Dynamic pari-mutuel markets (DPM)**

The third type of mechanism is used in dynamic pari-mutuel markets (DPM) developed by Pennock (2004); these are a combination between the CDA mentioned previously and pari-mutuel markets. Pari-mutuel wagering has traditionally been used in betting on horse races; a better could wager a certain amount on the horse that they predict to win the race. All the bets on a certain race are then pooled and after the race all the money in the pool (minus the commission) is divided amongst those who had bet on the winning horse. The payoff odds are thus dependent on how much money is in the pool (the amount bet on all the horses) and the amount of money bet on the winning horse so they are not known until after the race. How much is bet on each horse is known at all times before the race.

The advantage of this system is that it provides infinite liquidity, solving the problem of the previous two trading mechanisms. New bets increase the size of the pool and do not require anyone willing to ‘sell’ the bet. The main disadvantage of this system is that there is an advantage to waiting till the last minute to place a wager. Firstly, since new information can become available, the trader’s information set is largest just before the race. Secondly, placing a wager informs other traders that you have done so, making private information public. If a large bet is placed on a specific outcome it might encourage others to do the same since they assume these trades were made with private information. This revelation decreases the expected profits of the trader with private information.

Because of this incentive to wait, pari-mutuel markets are not a particularly good choice for prediction markets since prices before the close of the market will not be accurate predictions using all the information available. Traders can also not
sell the contract again before the expiration date. To solve this problem, Pennock (2004) combines a parti-mutuel market (which provides infinite liquidity) with a CDA (where prices react quickly to new information) to make it more appropriate for use in a prediction market; he calls this a dynamic pari-mutuel market (DPM). The way a DPM works is that traders can always purchase shares in any outcome, at any time, from the market (infinite liquidity for those wanting to buy) and the prices at which they can buy each contract are set continuously by the market. The market sets the price for each contract depending on the current state of wagering; if a lot of traders want to buy a certain contract they will increase the price of it, if a contract is relatively unpopular they will decrease its price. This is done automatically by a predetermined price function and in this way the current prices reflect all current information. In more traditional finance terms; this can be interpreted as a market maker who is always willing to sell each contract to any buyer at a price that they set.

However, the market maker does not buy back the contracts of traders who want to sell, this occurs via a standard CDA mechanism. As a consequence of this, on the sell side all information is also incorporated into prices but liquidity is limited. As with the previous two trading mechanisms there is no risk for the market, it only redistributes money from those making incorrect predictions to those making correct predictions.

Market scoring rules (MSR)

The last important market mechanism is the market scoring rule (MSR) introduced to prediction markets by Hanson (2003). He wrote that, in combinatorial information markets, where one wants to aggregate information on the entire probability distribution over many variables, a scoring rule can be used to reward forecasters for incremental improvements made to the forecast. He suggests the use of a market scoring rule to act as a market maker. Any trader can see the current probability distribution and if they believe that it is incorrect they can pay off the last trader (the trader who made the last prediction) according to the scoring rule and replace this probability distribution with their own new prediction. In this way each new predictor will receive money for improving the prediction of the previous predictor and lose money if they make a worse prediction. The final trader is paid off by the market depending on how far their prediction is from the actual outcome. Hanson (2007) suggests the use of a logarithmic market scoring rule to act as the market maker.
As with the previous trading mechanisms, there is continuous information incorporation since trades can be made at all times. There is also guaranteed liquidity at all times and on both sides of the market due to the market scoring rule acting as a market maker. Here, however, there is a risk for the operator which depends on how informative the last prediction was relative to the initial price of the contract (so is bounded). This cost is independent of the number of trades that take place and can be minimized by the market by setting the initial price to the initial beliefs.

Another disadvantage is its apparent complexity; an amateur trader will find changing the probability distribution a rather complex way of trading. However, it is possible to apply the same trading mechanism but make it appear to traders as if it were a limit order book where trades are shown in terms of buying and selling. This is what is done most often in practice when market scoring rules are used.

1.5.3 How contracts are introduced to the market

Once contracts have been defined and the trading mechanism has been chosen, there is still the question of how contracts are to be introduced to the market. This is only of real concern when using a continuous double auction or a call auction since in the other trading mechanisms there is infinite liquidity on the buy side. In practice we see two methods used.

Firstly, the most common method (such as is used by Intrade for example) is to open an order book for each new market and contracts are created when trades occur. For example, starting from a new market with no contracts traded; a first trader wanting to buy a contract at a certain price can enter their order into the order book. If a second trader believes this price is too high, they can place a sell order (without having to own any of the contracts), the orders will be matched and the first trader will have bought a contract and the second trader will have sold a contract. The reason that this works is that when the contract expires, all the traders need to have a balance of zero contracts; meaning, if they purchased a contract during trading, they need to sell it before or at the expiration date while a trader who has sold a contract needs to buy one before or at the expiration date. This means that the first trader is betting that the contract will close at 100 and will make a profit if this happens since he can sell his contract at 100. The second trader is betting that the contract will close at 0 since they will be able to buy back the contract, that they previously sold, at 0. This method means there is no risk for the market.

A second method is the use of unit portfolios (for example, at the IEM). This means
that any trader can purchase a portfolio of contracts from the market at any time. This portfolio contains all the contracts traded in a certain market so that the price of the portfolio is equal to the expiration price of the contracts, usually 100. For example, a unit portfolio on an election with two candidates would contain one contract for candidate A to win which will close at 100 if they win and 0 otherwise and one contract for candidate B to win which will also close at 100 if they win and 0 otherwise. With only two candidates, the portfolio will be worth 100 regardless of the outcome of the election. Traders who have purchased a unit portfolio can then trade one of the contracts depending on their predictions. Unit portfolios can also be sold back to the market at any time. This method also involves no risk for the market and means that there will always be an equal number of all contracts in the market.

1.5.4 The motivation to play (real-money vs. play-money)

In practice, we see prediction markets that use real-money such as Intrade, Betfair, Bet2Give and the Iowa Electronic Markets but also markets which do not use real-money such as the Foresight Exchange, NewsFutures, Hubdub and Inkling Markets. Instead, traders are given a certain sum of play-money when they open their trading account and can buy or sell contracts using this fake money. Here the motivation to play is different since there is no direct financial reward for making correct predictions. Often traders are rewarded psychologically; by having rankings of the best traders listed on the website (introducing an element of competition) or materially; by allowing traders who have performed very well, to trade in their play-money for real world prizes.

We would expect that there is an advantage to using real-money, making traders “put their money where their mouth is” (Hanson 1999, p.17). Real-money encourages the making of rational decisions and encourages information discovery. Financial gain encourages traders to actively trade the contract and reveal their preferences. In order for a market to make accurate predictions, the active participation of traders who are knowledgeable about the event in needed, financial incentives are a good way of encouraging them to join the market. Also, the more certain a trader is about their predictions, the more money they will be willing to invest, potentially allowing the market’s price to be weighted by the certainty of its traders.

One often mentioned advantage to using play-money markets is that the money only exists on the market and so the only way of acquiring a large capital is by making correct predictions. Since predictions are weighted by the size of the investment,
the predictions of those with large amounts of play-money will have a larger weight in the setting of prices than those with small amounts of play-money. In real-money markets, the money can come from outside the market and the ability to make money in the real world is not directly linked to the prediction making ability of the traders. Play-money markets could thus lead to a more efficient weighing of predictions.

**Financial considerations**

When deciding on whether a market should use real- or play-money it is not just the predictive accuracy of the market that needs to be taken into account, real-money markets are more difficult to set up and have higher costs which might outweigh any possible gains in predictive accuracy.

In many countries, prediction markets are seen as a form of gambling which is either illegal or controlled by a state monopoly, making it illegal to set up a real-money prediction market. We see that real-money markets usually have a large proportion of traders from the United States but are set up in other countries such as Intrade in Ireland or Betfair in the UK. The IEM function in the U.S. using real-money but is offered an exception to the law because the amount of money traders can invest is limited and because they are designed for educational purposes. There has recently been a call from many academics for U.S. regulators to lower the barriers to creating prediction markets ([Arrow et al. 2007](#)).

Setting up the trading platform for a real-money market or a play-money market is similar but the real-money money will need approval from governments and has other costs that come about from looking after other people’s money.

There are also many situations in which it is difficult to use real-money; a company making forecasts cannot ask employees to stake their own money to improve its forecasts.

**Research from experimental economics**

The use of monetary incentives has been a popular subject in experimental economics. When setting up an experiment studying the actions of economic agents, the organizers have to decide whether the experiment will use real-money or hypothetical money. The traditional view is that subjects in an experiment will act differently when being asked to make decisions using real-money or hypothetical-money, the only way to gain accurate results is to use real monetary incentives.
However, Read (2005) discusses the use of monetary incentives in economic experiments and concludes that “there is no basis for requiring the use of real incentives to do experimental economics” (p.265). He sees the advantages of monetary incentives, pointing out that monetary incentives compensate participants in experiments for the effort they have to exert and that the higher the compensation is, the more willing they are to put effort and consideration into their actions and decisions but does not consider them necessary to reach accurate results. However, he also points out that there are situations which must be experienced and can thus not be predicted based on a hypothetical scenario.

Interesting for our study of prediction markets, Read writes that “people are typically willing to pay less for almost anything if the money is real than if it is hypothetical” (p.266). In one study (Cummings et al., 1995), participants were asked if they were hypothetically willing to pay a certain amount for various objects and were then given the opportunity to purchase the same objects with (their own) real money. The experiments showed that the percentage willing to hypothetically buy the object was always much larger than the percentage that actually purchased the object when offered. If this conclusion can be generalized to prediction markets, we would expect the predictions of play-money markets to be higher than those of real-money markets, since traders would be willing to pay more for the contracts.

Read’s paper however points out that while money can be a good motivator, non-monetary motives can be used to have the same effect in experimental settings. His main argument for non-monetary incentives is that there are great costs (in time and funding) required to run real-money experiments and so there has to be a trade-off between the added realism and the added costs of using real-money. This is the same as has already been mentioned above with respect to the costs of running a real-money prediction market.

Gneezy and Rustichini (2000) discuss a set of experiments testing whether monetary incentives improve performance. The traditional belief of economists is that monetary incentives will improve performance in an activity. The authors point out that an action has an intrinsic motivation (this is the inherent motivation which is not dependent on the person being rewarded for their action) which encourages the person to do this action, which can be replaced by an extrinsic motivation; a reward (for example, monetary compensation). A common example is that of donating blood; when donors are not being compensated, they donate blood due to the intrinsic motivation of societal duty and feeling good about one’s self. If monetary compensation is then offered for the donating of blood, the compensation replaces
the intrinsic motivation and people see the compensation as the reason for donating blood. The offering of monetary compensation decreases the individual’s utility of the action since the gain in money does not fully compensate the loss in intrinsic motivation. This is interesting for prediction markets where the intrinsic motivation is the enjoyment of trading; if a payment were introduced in a play-money market, predictions could worsen.

The authors run various experiments to explore this. In one such experiment, pairs of students went door-to-door collecting donations for charities. The amount that each pair collects depends mainly on the amount of effort they put into it. The students were divided into three groups: the first group was given a speech about the importance of the money they were collecting and told that the results of the collection would be published (how much each pair collected would become public knowledge), the second group was given the same speech and promised 1% of the total value of donations they collected, the third pair was promised 10% of the donations they collected (it was made clear this payment would come from the researchers and not from the charities). The results show that the pairs in the first group collected, on average, more than the pairs in the other groups. The pairs in the third group collected on average less than those in the first group and those in the second group collected on average the least.

The conclusion that the authors draw is that when compensation for an activity is offered, a higher compensation increases performance. However, when compensation is not being offered, introducing a small monetary incentive can have a negative effect on performance. The relation to prediction markets is that on play-money prediction markets, the traders trade for entertainment and enjoy making trades. If they were given a small compensation for each trade they may see trading as a way of earning money, instead of pleasure (extrinsic motivation replaces intrinsic motivation) and they may trade less. However, we can assume that the monetary incentive on real-money markets is large enough to compensate the loss in intrinsic motivation, this is comparable to sports gambling or investing on the stock market.

**Research from prediction markets**

The previous research comes from experimental economics and shows that the use of monetary incentives can have both advantages and disadvantages. This section looks at the research directly related to prediction markets.
Servan-Schreiber et al. (2004) ran an experiment to compare the predictions of a real-money market (TradeSports) and a play-money market (NewsFutures\textsuperscript{11}) to see how much (if any) predictive accuracy is lost when using play- instead of real-money. Their experiment uses data from 208 National Football League (NFL) games in 2003 and for each game took the last trade of the contract before noon on the day of the game and compared it to the outcome. They find that 64.9% of the real-money market’s favorite teams actually won their games (135 out of 208) and an average pre-game trading price of 65.1 for the favorite. The play-money market saw 66.8% of their favorite teams win (139 out of 208) with an average pre-game trading price of 65.6 for the favorite. Other tests of the predictive qualities of the two markets showed that the predictions of the markets were extremely similar. They also compared the markets to individual human experts by entering them in an online football forecasting contest. In competition with 1,947 contestants, the play-money market ranked 11\textsuperscript{th} at the end of the 14 week study and the real-money market ranked 12\textsuperscript{th}. The authors conclude from their experiment that there is no statistically significant difference in predictive accuracy between the two markets so prediction markets using play-money can be just as accurate as those using real-money.

Rosenbloom and Notz (2006) compare the same two markets (TradeSports and NewsFutures) using a SPRT-like statistical test\textsuperscript{12} to determine whether there are statistically significant differences in accuracy between the two markets. Their data set consisted of an observation for both markets each time both offered contracts on the same event, this included; sports, financial markets and political events. They find (after 522 observations) that, although the differences in predictions of the two markets in absolute terms were slight, the predictions of the real-money market were more accurate than those of the play-money market at the 1% significance level.

They also investigated whether the results were market specific by looking at three different types of contracts separately. Firstly, looking at just markets related to NFL games they found the same results as Servan-Schreiber et al., both markets performed equally well. Secondly, looking at just contracts predicting the daily direction of the Dow Jones Industrial Average, they found that the real-money market significantly outperformed the play-money market. The third group included only

\textsuperscript{11}NewsFutures is a play-money market but allows traders to bid on real prizes using their play-money in auctions at the end of every month.

\textsuperscript{12}A sequential probability ratio test (SPRT) is a sequential test for a simple hypothesis $H_1$ against an alternative simple hypothesis $H_2$. The actual calculation is not explained here and only the results are presented, for more details about the test see Rosenbloom and Notz (2006).
contracts about North American team sports (college basketball, NBA basketball and Major League baseball) and they found that the real-money market performed slightly better than the play-money market but despite the large data set the results were insignificant. They conclude that real-money markets are significantly more accurate than play-money markets for non-sports events (such as the direction of the Dow Jones Industrial Average) than for sports events. They speculate that this might be because sports events have many sources of information about the expected outcome (for example, odds at sports betting shops) while for non-sports events the traders are more likely to rely on their own information aggregation.

A similar real-world experiment was conducted by Luckner et al. (2006) in Germany, comparing a real-money market (BLUEVEX\textsuperscript{13}) to a play-money market (STOC-CER\textsuperscript{14}). Both markets tried to predict the top team in the German Soccer League after the first half of the 2005-2006 season using a winner-take-all contract for each team in the league. They found that the play-money market had a lower average absolute prediction error than the real-money market on 37 out of the 50 days, meaning that the play-money market made better predictions. However, they mention various other differences between the markets (such as the number of users, transaction fees and portfolio trading) which could explain this difference, the previous two studies (Servan-Schreiber et al., 2004; Rosenbloom and Notz, 2006) had the same weaknesses.

Lastly, we look at the research of Luckner (2006) who set up an experiment during the FIFA World Cup of 2006. He operated 20 prediction markets for the last 20 matches of the tournament and used 60 undergraduate students as traders. They were split into three groups and each group was given a different incentive scheme with an average payment of €50 per person. Members of the first group were paid a fixed amount of €50 for taking part in the experiment (there was a minimum trading volume per week imposed to make sure there was some trading). Members of the second group were promised a “performance-compatible payment”, meaning the payment they received depended linearly on their performance on the market. Members of the third group were rewarded according to their ordinal rank in the group, the student making the best predictions ranked first at the end of the game and was paid €500, the second €300 and the third €200 with all the rest getting no payment. The author expected the performance-compatible payment group to perform the best and the fixed payment group the worst. The real results were the opposite, with the rank order group performing the best with the favorite outcome.
predicted by the group occurring in 45% of the cases, the fixed payment group was
correct in 35% of the cases and the performance-compatible group was only correct
in 20% of the cases (since there are three possible outcomes, this is lower than
choosing at random). The author speculates that risk aversion was the main reason
for these results; in the fixed payment tournament, traders can neither win or lose
money so are free to take risks (as in a play-money market) and in the ordinal group
traders are encouraged to take risks since these lead to high payments which are
needed to rank in the top 3. The performance-compatible group however could lose
money so were less prone to risk taking.

Non-monetary incentives

As the cases above show, just because a market uses play-money does not mean
that the predictions are any worse than if it were to use real-money, the market
provides other incentives for traders to trade. The main incentive is that prediction
markets can be fun and traders will happily trade for purely entertainment reasons.
Play-money markets are often set up in such a way as to encourage competition
between the traders; for example by giving leaderboards showing how accurate each
trader is at making predictions, the ambition to move up the leaderboards motivates
players to make accurate predictions.

An interesting case in encouraging trading without the use of real-money is Hubdub.
As described earlier, this is one of the more recent markets and has created an online
community for the traders. They provide leaderboards with all the traders ranked
by their net worth and clearly show where each trader ranks. They also include
social networking features allowing traders to become ‘friends’ with other traders on
the market and challenge them and compete with them. There are active forums
allowing members to discuss events with each other. Hubdub also gives each trader a
‘star’ next to their username depending on their net worth, allowing traders to show
the other traders that they are performing well on the market. Another way they
encourage active participation is by allowing traders to submit their own “questions”
(contracts which are specified by the trader).

Other play-money markets (such as NewsFutures) award real prizes to the top
traders, in this case, play-money can be either ‘spent’ in an online shop on real
prizes or used in an auction for prizes. Usually the prices are high so that only the
best performing traders can reach them and this also encourages traders to get more
play-money by making better predictions.
1.5.5 The population of traders

An important question that needs to be answered when designing a prediction market is who the population of traders will be. For a corporate market it might be useful to run the market internally since the people who have the information that you are looking to aggregate all work within the company. The people running the market can then individually choose the traders so that they include anyone who might have useful information. The market is designed to aggregate information, if the selected traders do not possess any information there is no use to running the market. Usually, other ‘uninformed’ traders are also included in the market to increase its liquidity and so make the market more efficient. Markets on topics that rely more on the interpretation of public information instead of revelation of private information are usually open to the public on websites to allow anyone who would like to trade to do so.

Number of traders

The traditional view is that the larger the number of traders and the higher the amount of money in the market, the more efficient the market will be and so the better the predictions it will make. The larger the number of traders, the thicker the market so the more accurate the prices are (Berg et al., 1996). In practice, some markets (such as Intrade) have a much larger number of traders than others (for example, the IEM), we would expect the market with the largest number to make the most accurate predictions.

However, experiments have shown that accurate forecasts can also be made using only a small number of participants (Surowiecki, 2004). In the HP internal markets discussed previously, Chen and Plott (2002) found that the market’s predictions are more accurate than company forecasts despite often involving less than 25 traders.

Research from the IEM (Forsythe et al., 1992; Forsythe et al., 1999; Oliven and Rietz, 2004) has shown that only a small core of rational traders is necessary to reach efficient prices in a prediction market. While the average trader tends to make mistakes and is biased in their valuation of the contract, they are not the ones who set the prices. Prices are set by the marginal traders who are more rational and make less mistakes than the average traders. The effects of traders not being rational will be discussed later in Section 1.6.2.
Diversity of information

It is safe to assume that the number of traders alone is not of great importance. As long as there is enough liquidity in the market, even a small group can make accurate predictions. What is of importance though, is the information that the traders possess. For trades to be made, the traders need to disagree on the price of the contract, if all have the same information and agree on the same price then there will be no trading (Wolfers and Zitzewitz, 2004). Surowiecki (2004) also points out, with many examples, that disagreement in a group allows them to make better decisions than if everyone agrees.

While diverse information is useful, the traders themselves do not need to be diverse. In predicting election outcomes, polls try to find a representative sample of the voting population to ask how they intend to vote in the election and then extrapolate the results. If the polled sample is not representative of the population of voters, the poll is unlikely to have accurate results. For a prediction market this is not the case. Berg, Nelson and Rietz (2008) point out that traders on the IEM are typically young, white, well-educated, men with high family incomes. In the 2000 IEM presidential election markets, 20% of the traders were from the state of Iowa while the state only accounted for 1% of the U.S. population. While these traders were unrepresentative of the voting population, they were still able to make better predictions than if you were to ask the voters themselves.

1.5.6 Other market microstructure considerations

This section briefly summarizes other factors that need to be taken into account when setting up a prediction market. These considerations are not of great importance while prediction markets are still small volume markets but will become more and more important when they grow. There has also not been much specific research on these factors with respect to prediction markets so only a brief overview is given. For more information, the reader is referred to the literature on market microstructure such as O’Hara (1995) and Hasbrouck (2007).

Transaction costs

Trading implies both an explicit cost (the fees charged for trading) and an implicit cost (the price of immediacy and the price impact), here we look only at the explicit trading fees. A fee would be expected to harm the accuracy of the market since it
prevents traders from making trades where the expected profit is positive but less than the trading fee. We can imagine many trades which could make a small profit without the existence of fees but make a loss with fees.

However, it is clear that a for-profit market needs to charge some kind of fee to cover its operating costs and to potentially make a profit. Instead of charging a flat fee, fees can be set in such a way to minimize the effect on price formation. Intrade, for example, does not charge a fee for price makers (orders that are not immediately matched) and only charge fees to price takers (orders that are immediately matched), these are $0.03 for ‘extreme’ prices (prices that are between 0 and 5 or 95 and 100) and $0.05 otherwise. They also charge an expiry fee of $0.10 on profitable positions and nothing on positions that lost money.

Neither Bet2Give nor the IEM charge trading fees (the IEM do have a $5.00 activation charge) but both these markets have trading limits as described in the next section.

Play-money markets do not charge fees (since the money is not real anyway), they usually try to cover their operating costs by providing market data to researchers and the media or by using advertising on their websites.

Trading limits

Another interesting aspect is the limit to trading that we see in practice used for real-money markets. As explained previously, this is due to the illegality of gambling in the U.S., the IEM have an exception to this law only because it is for educational purposes and because they limit each trader’s investment to $500. Markets not based in the U.S. (for example, Intrade and Betfair) do not have significant limits to how much each trader can invest. We would expect that the higher the limit, the higher the potential profit for investors and the more it would encourage information discovery. There has, however, been little research on this; the removal of limits might make prediction markets more like traditional financial markets where large amounts of money are spent on researching profitable investments. Although, if there are no limits on how much can be invested, the markets could more easily be used for the hedging of other positions which might decrease the predictive value of market prices.

In play-money markets we do not see limits since the money is not real and can only be acquired by making accurate predictions on the market.
Information available to traders

Another important issue when designing a prediction market is the information that will be made available to the traders; how transparent the market is. By transparency we mean how well traders can observe information about the trading process such as prices and the order book. We can distinguish between pre-trade transparency which involves information about the bid and ask prices, the depth, and the order book in general and post-trade transparency which involves the publication of information about completed trades including the time, price, size and who placed them. The historical prices are needed to allow traders to analyze them and devise profitable trading strategies. A more transparent market would be expected to form prices more accurately than a less transparent one. In practice, prediction markets tend to be very transparent both pre- and post-trade. Intrade in particular gives traders a lot of pre-trade information and data from expired contracts is available on the website for a short while after the contracts expire. The IEM also allow the data from all their closed markets to be downloaded for study.

Anonymity

Anonymity is the degree to which the identity of traders is revealed to other traders. When trading is anonymous, insiders or traders with private information, would be more willing to trade, since it would be much more difficult for other traders to distinguish between a trade from an informed and an uninformed trader. This would encourage traders with private information to trade and improve the market’s predictions. In practice we see that both Intrade and the IEM do not reveal the identity of traders, there is no way of knowing who is placing which orders. Unlike other asset markets there is no explicit ban on insider trading since the information from insiders helps to improve the market forecast.

Usually on play-money markets the usernames of the traders are clearly visible, this is part of their strategy to encourage competition between individuals and to allow traders to see how well they are doing compared to other traders. If the market was anonymous one of the large motivations to trade, the ‘bragging rights’ of doing well in the competition, would be removed.

Trading hours

Most prediction markets, unlike most financial markets, are open all day, every day. This allows them to incorporate information into prices at all times and encourages
traders from all over the world to use them. However, the disadvantage is that it might prevent the use of predication markets as part of investment portfolios since if a trader is not following the news and able to trade all day, they could make a loss if new information becomes available, for example, at night. If prediction markets move into the ‘real’ investment world, it would be realistic to assume that the trading hours would be limited.

In corporate prediction markets we see that this is often done. In the HP internal markets discussed previously (Chen and Plott, 2002), the markets were only open during lunchtime and after work hours so as to not distract traders from their work.

**Competition between trading venues**

Since there is no underlying asset for prediction markets, each market can effectively offer contracts on any event they choose. For large and important events, we see that a lot of markets offer the same contract (a U.S. presidential election for example) but in general there is not much competition between the markets. The IEM are mainly for educational purposes and offer contracts on mainly political events, Intrade offers contracts mainly on current affairs and Betfair offers contracts on mainly sporting events. While there is often overlap in contracts offered (as we will see in the next part of this paper) they do not seem to directly compete for traders in terms of trading costs or liquidity. A potential trader will choose the market based mainly on the types of contract they are interested in. In the future, as more real-money markets are created we can expect more competition between them.

Often play-money markets are specialized in a certain type of event (for example, the Hollywood Stock Exchange) however, we also see markets that offer very similar contracts (for example, Hubdub and Inkling Markets), these then compete mainly on the incentives they offer to trade and the design of their websites.

**The trading interface**

Lastly, for online markets the usability of their website is of great importance. Traders, especially in play-money markets, are attracted to well designed, easy-to-use websites and this is of great importance for attracting a large trading population and encouraging them to keep trading. For example, the reader is invited to compare the design of the Foresight Exchange founded in 1994 (http://www.ideosphere.com/) to that of Hubdub founded in 2008 (http://www.hubdub.com/). At the time of

---

15For more information of the usability of markets see: http://www.usablemarkets.com/
writing NewsFutures is going through a redesign to improve their website and the Hollywood Stock Exchange has just completed a redesign.

1.6 Other questions about prediction markets

1.6.1 Do arbitrage opportunities exist?

The nonexistence of long-term arbitrage opportunities is important for a market to make accurate predictions. We can divide this into two separate conditions; first, *arbitrage free pricing* means that the sum of the contracts on a market should be 1 at all times (for winner-take-all contracts minus any commission). For example, if there is a political election with only two candidates and only one will win, then the sum of the two contracts has to be equal to 1 at all times. Otherwise, if the sum is greater than 1, guaranteed profit can be made by buying a unit portfolio and selling the parts individually. If the sum is less than one, both contracts can be bought individually and sold back to the market as a portfolio. Within a market with only a few contracts this is easy to control and we would expect any discrepancies to be quickly removed.

Secondly, the *law of one price* states that the same contract has to have the same price on different markets. Unlike a real asset, a contract from one prediction market cannot be sold on a different market but we can still expect this law to hold in practice. For example, in a political market with two candidates if the prices are not the same on all markets it could be possible to buy one contract on one market and the other on a different market for a total cost of less than 1. Holding both contracts until expiration will lead to a guaranteed profit since the sum of the expiration prices of the two contracts will always be 1. This law is more difficult to control since the markets are very different in design but we would also expect if there are guaranteed profits to be made then there will be traders doing so, until these opportunities are removed.

If arbitrage opportunities exist, traders are not behaving fully rationally but we would expect that when they do exist they are quickly removed since they offer a profit opportunity for other traders. In fact, this is what we see in practice; in their summary paper, [Wolfers and Zitzewitz (2006c, p.5)](footnote) write that “the law of one price appears to (roughly) hold, and the few arbitrage opportunities that arise in these markets are fleeting, and involve only small potential profits”.

In corporate or experimental markets however, we do sometimes see arbitrage opportunities that are not taken advantage of. In the internal prediction markets of Google, Cowgill et al. (2009) find that in practice there were arbitrage opportunities which were not exploited. They find 1,747 instances where the bid prices in a market summed to more than 1 (certain profit could be made by buying a unit portfolio and selling the individual contracts on the market) and 495 instances where the ask prices in the market summed to less than 1 (certain profit could be made by buying all the contracts in the market and selling the unit portfolio). They explain this difference by an aversion to selling contracts short so, in practice not all traders act rationally. But, we would expect that as soon as it becomes clear that there are guaranteed profits to be made through exploiting arbitrage opportunities, that some traders will take advantage of these until these opportunities are removed.

1.6.2 Are traders rational?

This section summarizes the literature that shows that traders on prediction markets make mistakes and are biased by their preferences. It also mentions that both the favorite-longshot bias and speculative bubbles have been observed on prediction markets. If traders are biased, prices will be too. Despite all these problems, prediction markets still provide remarkably accurate forecasts of future events. The explanation often offered in literature is that while average traders have all these shortcomings, these are not the ones who set the prices. Prices are set by the marginal traders who are much more rational and tend to make less errors than the average trader. Those who set prices are far more rational than those who accept prices.

Do traders make mistakes?

Traders make mistakes by making trades which are sub-optimal regardless of what the trader believes. These are caused by a lack of understanding of how the market works or a lack of care when placing orders. Oliven and Rietz (2004) find that traders often make mistakes; by not taking advantage of the best available prices, they lose out on money. Looking at the data from the IEM vote-share market for the 1992 U.S. presidential election they discuss two types of mistakes. Firstly, price taking violations of the law of one price where traders accept trade prices that are not in their best interest; they buy for a higher price or sell for a lower price than they needed to. Secondly, market making violations of no-arbitrage pricing restrictions
where they submit a market order that is not in their best interest. The authors document frequent violations of both types, leading to large unnecessary costs for traders who make mistakes.

**Are traders biased by their preferences?**

If traders are biased in their judgement, they make trades that are not rational but are not mistakes either. This is because they take decisions optimally but their subjective probability of the event occurring is different from the objective probability.

[Forsythe et al. (1999)](#) discuss three effects that bias the judgement of traders in political prediction markets. Firstly, they discuss the *wishful thinking effect*, this means that traders are overly optimistic about the probability that their preferred party or candidate will win the election and also interpret news about that party more favorably than they should if they were objective. In other words, they overvalue contracts on events that they would like to occur. Secondly, they discuss the *false consensus effect* where a trader believes their preferences are more representative of the preferences of the whole population than they really are, again leading to them expecting a larger proportion of the vote for their preferred candidate, and buying more of their contracts than would be rational. Lastly, there is the *assimilation-contrast effect* which occurs in political prediction markets after debates between the candidates. A trader is more likely to feel that their preferred candidate performed well in the debates than the population at large. They find that all three biases exist in prediction markets and discuss laboratory experiments where these biases can be isolated and studied.

[Berg and Rietz (2006)](#) report on a survey of IEM traders from 2004. Asking the question “Regardless of your preferences, who do you think will receive the most popular votes in the upcoming U.S. presidential election?” they found that 68% of self-reported Democrats responded that they believed Kerry would win the election while only 5% of self-reported Republicans responded that they believed Kerry would win. This shows that their predictions for the election are skewed by their own political preferences. Another question; “Relative to other traders, how informed do you believe you were about the 2004 presidential election?” found that some 89% of respondents believed that they were better informed than the other traders in the market.

Similar evidence was found by [Cowgill et al. (2009)](#) in Google’s internal markets
where traders who were newly hired were overly optimistic and optimistic biases were largest on days where the company’s stock price increased.

**Is there a favorite-longshot bias in prediction markets?**

Another bias of prediction markets is that they can suffer from a *favorite-longshot bias*, this means that the market prices overestimate the probability of very unlikely events and underestimate the probability of very likely events. Thaler and Ziemba (1988) discuss the favorite-longshot bias in horse racing and find that bettors tend to overvalue extreme longshots and will as a consequence receive lower than average returns by betting on the longshot. Snowberg and Wolfers (2007) study this and find evidence that this is driven by the misperceptions of the bettors and not by risk-loving utility functions, people tend to overvalue events with very small probabilities and undervalue events with very high probabilities. Wolfers and Zitzewitz (2006a, p.23) suggest that in horse racing, the overestimation is greatest for events priced below 5% while for the IEM it is present for events priced between 0.2 and 0.3 (so between 20% and 30%).

An example from prediction markets is given by Wolfers and Zitzewitz (2004, pp.117-118). They look at contracts on the TradeSports prediction market on how high the Standard and Poor’s 500 index would be at a certain date and find that extremely unlikely (high and low) outcomes are relatively overpriced compared to similar options traded on a financial market.

**Can speculative bubbles form on prediction markets?**

A speculative bubble forms when large trades keep the price of an asset increasing without this being justified by fundamentals. Bubbles distort the prices of the market and thus make the predictions of the market less accurate. We see this occurs often on other asset markets and there is no reason to expect that prediction markets would be exempt from them. However, there has been little research into this directly related to prediction markets. Wolfers and Zitzewitz (2004, pp.118-119) report on a possible bubble in the TradeSports contract in September of 2003 on whether Hillary Clinton would win the Democratic nomination to run for president. They suggest that prices were much higher than they should have been based on the information available and that prices were driven up by traders who were betting on large future price movements.
1.6.3 Is market manipulation possible?

One of the fears with the use of prediction markets is that they can easily be manipulated. Traders may be interested in distorting the prices of a market away from its fundamentals to influence the actions and expectations of the other traders or other observers of the market. This is of great concern when the prices of the market are used for other purposes; if the market price is used to help make policy decisions, a manipulator may be willing to take a loss in the market in order to distort prices to indirectly control policy. This section looks at speculative attacks which are large trades to change prices and manipulation which is a long term change in the price, this is not possible unless the manipulator changes the beliefs of other traders.

An example of potential manipulation is on political prediction markets where the political reward of manipulating a market could be greater than the financial cost of the manipulation. By increasing the price of the preferred candidate it might make them seem a stronger candidate than they really are which can encourage people to vote for them. Especially if a candidate is quite far behind, moving the odds closer to 50-50 will encourage people who wouldn’t normally vote, to vote since their vote becomes more significant. A candidate can also be made to look further ahead in the race than they really are which is discouraging for the supporters of the other candidates. Small stakes in the market could have a much larger effect on the election as a whole. A real world example of this is the accusation that there was manipulation on the Intrade U.S. presidential election markets of 2008 to increase the price of the McCain contract [Rothschild and Wolfers 2008].

Another problem is that prediction markets do not prohibit insiders from trading. This has the advantage that it could increase the predictive value of the market, allowing more information to be incorporated into prices. However, this means that if there is a change in price, traders must decide whether it was due to a manipulator or the introduction of new private information.

This was one of the fears with the DARPA Policy Analysis Market (as described in Section 1.2.6) where it was feared that terrorists could either commit terrorist acts to win the bets that they had placed on those acts or distort the market by buying certain contracts and selling other contracts to deceive the decision makers using the market to decide policy. However, it was shown that in practice this would be unlikely.

In their overview paper, Wolfers and Zitzewitz (2004, p.199) conclude that attempts to manipulate the prices on prediction markets do not “have a discernible effect on
prices, except during a short transition phase”. This section will look at the research on this topic.

**Experimental economics**

In experimental markets, it is easy to study the effects of manipulation since the information of all the traders in the market and the incentives offered to manipulators can be controlled and adjusted by the experimenter. Hanson et al. (2006) design a laboratory experiment to see if an experimental market can be manipulated. Firstly they replicate the design of Plott and Sunder (1988), where subjects trade an asset that can take three possible values: 0, 40 or 100. Each subject is told one of the values that the asset would not take, so individually they do not know the true value but as a group they have enough information to determine it. They then trade the asset in a double auction. In a second treatment, the authors gave half the subjects an incentive to manipulate the price upwards by offering a payoff dependent on the median contract price in the market at the end of trading. All subjects were aware that half the traders were trying to manipulate the price upwards but did not know who these subjects were. The results of the experiment show that, while the manipulators clearly attempted to manipulate the price upwards, they did not reduce the accuracy of the price as an estimate for the asset’s value. This was because the non-manipulators knew that there were manipulators driving up the price so were willing to accept systematically lower offers which undid the effect of the manipulators. From this experiment, the authors conclude that the manipulators have no systematic effect on the price but this experiment does assume that all traders know how many manipulators there are, what their incentive to manipulate is and in which direction they wish to manipulate the price, this might not be realistic in practice.

Oprea et al. (2007) run laboratory experiments to test a slightly different hypothesis; testing whether manipulators can mislead market observers, the third parties who will be making decisions based on market prices. As in the previous experiment, traders were asked to trade an asset in a double auction and were given noisy signals as to its real value. After the trading session, a group of third parties were given no signals about the asset’s value but were able to observe the price series from the market and were asked to forecast the asset’s value. The authors found that the information from the market allows them to significantly improve their forecasts. In a second treatment, half of the traders were given an incentive to manipulate the price, by offering them a payoff dependent on how close the observer’s prediction was
to their desired price. The difference with the previous paper is that the traders and
observers had no information about the desired price of the manipulators, making it
more difficult to counteract their manipulation. The results show that attempts to
manipulate the price upwards were successful but attempts to manipulate the price
downwards were unsuccessful. However, they find that the manipulation does not
reduce the accuracy of the observer’s forecasts.

Real world evidence

Camerer (1998) ran a field experiment to see if prices on a pari-mutuel racetrack
betting market could be manipulated. In each race, two horses with similar fea-
tures (based on the predicted odds by the handicapper) were chosen. By flipping
a coin, one was chosen to try to manipulate and the other was left as a control.
Bets of $500 and $1,000 were made on the horse at some time between 17 and 22
minutes before the race and cancelled at some time between 5 and 8 minutes before
the race. Using an event study, he found that the odds in the market could not
be systematically manipulated; other bettors did not systematically respond to his
temporary bets.

Hansen et al. (2004) argue that if the prices of the market are reported in the media
they can impact the voting behavior of the population. Voters are more likely to
vote in a close race since their vote becomes more important. By increasing the
prices of a losing candidate it might motivate people who wouldn’t normally vote,
to vote. They look at data from two separate, but similar, German prediction
markets (WahlStreet and Wahlboerse) during the Berlin state elections of 1999.
One party (the Free Democratic Party) needed a minimum of 5% of the vote to be
represented in the parliament. The party leaders sent an e-mail to all of its members
encouraging them to buy their party’s contracts to increase the price. The authors
found that prices on both markets were manipulated to keep the party close to 5%
and that very little capital was required to do this. This shows that the market was
very vulnerable to manipulation, however the party only received 2.2% of the vote
on election day so the manipulation did not have the desired results.

Rhode and Strumpf (2008b) conduct an extensive study into the manipulation of
real-world political prediction markets and find that while manipulators can initially
change prices in their desired direction, these changes are quickly undone. Firstly,

16 Pari-mutuel markets have been explained previously in Section 1.5.2
17 http://www.wahlstreet.de
18 http://www.wahlboerse-berlin.de/
they did a field experiment on the IEM for the 2000 U.S. presidential elections. They simulated the actions of manipulators by placing investments on a political party (equal to about 2% of the total trade volume) unrelated to news or changes in fundamentals on either the vote-share market, the winner-take-all market or both at the same time. Out of the eleven attempted manipulations, nine had a significant initial impact on the market price but these changes were not long lasting. Traders saw these changes as profitable trading opportunities so the effects of the manipulation were quickly undone and prices returned to their initial levels after a few hours. The manipulations had no effect on the long-term price level and were thus unsuccessful.

Secondly, they studied manipulation on historical political betting markets from 1880 to 1944. These were betting markets in New York run on U.S. presidential elections but also state and local elections. At the time, these were very popular and the results were widely published in newspapers. Rhode and Strumpf write that traders with close connections to the political parties, and the election, traded in the markets and that manipulation could have large effects in these markets if successful, since due to their popularity they could sway public opinion. Both parties were often accused of manipulating the markets to make their candidate appear stronger with the goal of influencing people who would not normally vote, to vote and to sway undecided voters. The authors collected odds from newspapers and also public allegations of manipulation. By analyzing the data, they found that despite the active participation of insiders, the market was popular and made accurate predictions. They also found that when public allegations of manipulation were made, these were not associated with any large or permanent change in the prices. Again, it was possible to manipulate the price but only for a short while, prices returned to normal levels within days.

Lastly, they looked at the TradeSports markets for the 2004 U.S. presidential elections and found two particular attacks from speculators. On September 13th 2004 and on October 15th 2004, the Bush winner-take-all contract dropped in price without any new information being made public that could explain it. Again their conclusions were the same, the attacks moved prices initially but they quickly returned to their initial levels and would not have been profitable for the trader. However, these two trades did receive considerable media attention which may have been the desired outcome of the manipulator.

For more information on these markets, see the next part of this paper or Rhode and Strumpf (2004).
Rhode and Strumpf conclude that while it is possible to manipulate political stock markets for a short time, doing so for a longer time period is both difficult and expensive.

Theory

While there have been many studies on prediction market manipulation in practice, there has been little theoretical work on the subject. Hanson and Oprea (2009) start with a simple single-period version of the Kyle model, adapt it to be more fitting for a thin prediction market and add a manipulator. Other traders know the strength of the manipulator’s preferences but do not have perfect information about their target price. They conclude that the manipulator’s efforts have no effect on the market price and no direct effect on the accuracy of the prices. They find that the manipulator can be seen as a noise trader, since his trades are not based on information, and actually increases the accuracy of the market. By giving other traders an incentive to collect information from which they can profit, manipulators cover the cost of information discovery by non-manipulators. By providing liquidity, manipulators thus improve the predictive accuracy of the market. However, this model makes a lot of assumptions which may not hold in practice and more research is needed before we can confirm their conclusion.

1.6.4 How can we interpret the predictions of prediction markets?

The standard view, as has been mentioned previously, is that the current price of a contract traded on a prediction market reveals the mean belief of traders and so the price of a winner-take-all contract can be directly interpreted as the probability of the event occurring (this has been shown in Section 1.4.2). For example, if a political winner-take-all contract which is worth 100 if a candidate wins and 0 if he loses, currently trades at 60 we say that his chances of winning are 60%.

However, Manski (2004) points out that this interpretation of prediction market prices has always been assumed without an adequate theoretical base. He derives the equilibrium price of a prediction market (similar to the IEM, where the number

\[20\] The Kyle model involves uniform traders (“noise traders”), informed traders and a market maker who sets prices efficiently dependent on the combined quantities that the traders wish to trade. For more details see Kyle (1985, 1989).
of contracts on both sides of the market is always the same) when traders are risk-neutral price-takers with heterogeneous beliefs and have a fixed trading budget. He finds the equilibrium price for a contract \( m \), that is valued between 0 and 1, to be (Mansi 2004, p.2):
\[
\pi_m = P(q_m > \pi_m)
\]
(1.4)

Where \( \pi_m \) is the price of the contract \( m \) and \( q_m \) is the subjective probability of the individual trader that event \( m \) will occur. \( P(q_m, y) \) is the cross-sectional distribution of beliefs \( (q_m) \) and budgets \( (y) \). This shows that when the market price of the contract is above 0.5, most traders have beliefs higher than the price and when the price is below 0.5, most traders have beliefs lower than the price. The equilibrium price thus equals the \((1 - \pi_m)\)-quantile of the distribution of beliefs, not the mean.

Wolfers and Zitzewitz (2006c) give a numerical example of these results; suppose all traders are willing to invest exactly $100 in the prediction market on a contract that pays $1 if the event occurs and $0 otherwise. If the contract is currently selling for $0.667, then those who have a higher valuation than the price will be wanting to buy contracts and can each buy 150 contracts while those having a lower valuation than the price will want to sell contracts and can sell 300 contracts (at a price of $0.333). In equilibrium there must therefore be twice as many buyers as sellers which means the market price must fall at the 33rd percentile\(^{21}\) of the belief distribution, not at the mean. This also implies that at a market price of \( \pi_m \), \( 1 - \pi_m \)% of the traders believe the event has a lower probability than \( \pi_m \)% of occurring. In the example (with a price of $0.667) this would mean that 33% of the population believes that the event has less than a 66% chance of occurring. Manski also shows that the mean belief must lie in the interval \((\pi_m^2, 2\pi_m - \pi_m^2)\), the market’s prices do thus not allow us to know the mean belief of the traders, only a wide interval in which this mean belief lies. However, Manski’s model relies on the assumption that traders are risk neutral and always willing to invest a fixed sum in the market.

Wolfers and Zitzewitz (2006b) expand on this and describe the conditions under which the prices of a prediction market can correspond with the mean beliefs of the traders. They study the mapping from the distribution of beliefs to the equilibrium price in a prediction market. Firstly they derive the market price where individual traders \( j \) have log utility \( U_j \) and an initial wealth \( y \). They each choose how many contracts \( x \) they want to buy on the market at the current price \( \pi \) given that they

\(^{21}(1 - \pi_m) = (1 - 0.667) = 0.333\)
have a subjective probability of winning of $q$ (Wolfers and Zitzewitz, 2006b, p.3):

$$\max_{\{x\}} EU_j = q_j Log[y + x_j(1 - \pi)] + (1 - q_j) Log[y - x_j\pi]$$  \hspace{1cm} (1.5)$$

yielding:

$$x_j^* = y \frac{q_j - \pi}{\pi(1 - \pi)}$$  \hspace{1cm} (1.6)$$

The market is in equilibrium where supply is equal to demand:

$$\int_{-\infty}^{\pi} y \frac{q - \pi}{\pi(1 - \pi)} f(q) dq = \int_{\pi}^{\infty} y \frac{\pi - q}{\pi(1 - \pi)} f(q) dq$$  \hspace{1cm} (1.7)$$

If beliefs ($q$) and wealth ($y$) are uncorrelated, this implies that:

$$\pi = \int_{-\infty}^{\pi} q f(q) dq = \bar{q}$$  \hspace{1cm} (1.8)$$

Here we see that with the assumption that the traders have log utility, the market prices will be equal to the mean belief of the traders.

However, if beliefs and wealth are correlated, traders with certain beliefs will have a greater effect on the market price since they are capable of making larger trades. The result is that the market price is a wealth-weighted average of the beliefs of the traders, not the mean belief of the traders.\footnote{This could be interesting for the study of play-money markets where money can only be accumulated by previous correct predictions. If the market price is weighted by wealth and the wealthy have a history of making accurate predictions, the market price may be more accurate as a result.} They also look at alternative utility functions\footnote{They look at constant relative risk aversion (CRRA), constant absolute risk aversion (CARA), quadratic utility, and hyperbolic absolute risk aversion (HARA); for details see Wolfers and Zitzewitz (2006b).} and find that prices can diverge from mean beliefs but that this divergence tends to be quite small. They conclude that the case studied by Manski is the worst-case scenario and is not realistic in practice. Gjerstad (2004) independently came to the same conclusions.

Since the Manski model is unrealistic in practice, we will in the rest of this paper refer to the equilibrium price on a prediction market as the mean belief of the traders and thus interpretable as the probability of the event occurring.
1.7 Conclusion: how to design a prediction market

In this part of the paper we have thoroughly discussed the issues that should be taken into account when designing a prediction market. The types of contracts offered (winner-take-all, index, spread) all allow for the prediction of different aspects of an event. Four different trading mechanisms; the continuous double auction, call auctions, dynamic pari-mutuel markets and market scoring rules, have also been looked into. By comparing each on the basis of their ability to continuously incorporate information, whether they guarantee liquidity and whether there is risk for the market operator it was found that each system has its own advantages and disadvantages.

Another consideration that we identified is how contracts will be introduced to the market, will a unit portfolio system be used? We then looked at one of the most important considerations; whether the market will use real-money or play-money. While we find some evidence that a real-money market can provide better predictions on certain types of contracts, the open question is whether this (potential) increase in predictive accuracy is worth the extra organizational and regulatory difficulties of using real-money.

Another important consideration is how to attract a population of traders that has information that you are looking to aggregate. Other considerations we identify include; the transaction costs, trading limits, the information available to traders and the anonymity of traders. We also mention the competition between trading venues and the importance of the trading interface of the market.

We then looked at some often asked questions that question the performance of prediction markets. Our findings tend to be favorable for the markets. Arbitrage opportunities tend to be few and quickly removed, traders make mistakes and are biased by their preferences but this has little effect on the market’s price and long-term manipulation has been shown to be difficult and costly. Prices do also not diverge much from the mean belief of the traders in most practical situations.

We continue our study of prediction markets in the next part where we apply what we have learned in this part to data from prediction markets in the run-up to the 2008 U.S. presidential election.
Part 2

Prediction Markets in Practice

The theoretical workings of prediction markets have already been examined in the previous part of this paper and it has been shown that, under the right conditions, a prediction market can be used to accurately assess the probability of future events. This part will examine whether this is true in practice by looking at the performance of prediction markets in the run-up to the 2008 United States presidential election.

As has been seen, prediction markets have been used to predict anything from the box-office returns of new movies to forecasting the weather, but one particular subject that prediction markets have always been used for is to predict the outcome of political elections. Recently, the U.S. presidential election of 2008 attracted a lot of attention from the media and also the prediction markets; nearly every online market offered contracts to predict the outcome (and various other aspects) of the election. Due to the amount of interest in this subject there was also a lot of data available from different contracts and different markets. The presidential election should thus prove to be an interesting case which we can use to assess the predictive value of prediction markets.

This part of the paper proceeds as follows: firstly, we will briefly discuss the various prediction markets from which data was collected. Then we use various techniques to assess how accurate the predictions were in absolute terms and compared to other more traditional methods of predicting the outcome of elections.
2.1 Prediction markets and U.S. presidential elections

As already mentioned in Section 1.2.2, one of the oldest and most popular applications of prediction markets is to forecast political elections. This is of great importance since the outcome of elections can have far-reaching effects on individuals, corporations and other governments. Being able to predict the outcome of an election allows them to account for this in their current decision making. We often see that the share price of certain companies is very dependent on the outcome of an election, since different candidates can have very different policies with respect to certain industries.

Betting on U.S. presidential elections is far from new, Rhode and Strumpf (2004) study large-scale betting markets on presidential elections between 1868 and 1940. These betting markets were focused in New York and had a much larger scale than prediction markets do today. For example, during the 1916 election some $165 million (in 2002 dollars) was wagered in the New York markets, this is more than twice the sum spent on the election campaigns that year. The average betting volume of these markets was also over two hundred times the largest volume in any election on the Iowa Electronic Markets (IEM). They quote the New York Times as writing: the “old axiom in the financial district [is] that Wall Street betting odds are ‘never wrong’.” (Rhode and Strumpf, 2004, p.3) and find that in only one election did the candidate that was clearly favored a month before the election, lose. From the 1940s onwards these markets declined, mainly due to questions of ethics.

Most of the often quoted evidence for the accuracy of prediction markets comes from the evidence of the performance of the Iowa Electronic Markets (IEM) on forecasting U.S. presidential election outcomes. Since 1988, these markets have been used to forecast the outcome of presidential elections both in terms of who will win the election (using winner-take-all contracts) and what percentage of the vote each party will receive (using vote-share contracts). Research using data from these markets have shown that the market makes accurate predictions, both in absolute terms and relative to polls. Berg, Forsythe, Nelson and Rietz (2008) look at the short-run predictions of various prediction markets including the IEM. They compare the prices of 237 contracts from 49 vote-share election markets run in 13 countries (including data from the IEM presidential markets of 1988, 1992, 1996 and 2000). For each contract they have two measures; the market price at midnight on the eve of the election and the volume weighted average price of all transactions.
over the week before the election. They find that the average error of the polls is 1.93% while the average error of the markets is 1.49% by the first measure and 1.58% by the second. They conclude that the market clearly outperforms polls in terms of short-term predictions. \cite{Berg2008} extend this analysis with data from the IEM presidential markets from 1988 to 2004 to study how well the markets perform on longer horizons compared to polls. They find that the market is closer to the outcome of the election in 74% of the observations and performs even better when looking only at observations more than 100 days before the election. We will be testing their results, for the 2008 election so their study will be discussed more in later sections. \cite{Erikson2008} criticize the conclusions of \cite{Berg2008} and these criticisms will also be addressed later.

2.2 Description of data

In practice, we see that there are many different prediction markets currently running on the internet, offering a wide range of contracts.\footnote{As far as we are aware there are currently no large-scale offline prediction markets.} Although they have many differences, one similarity is that nearly all of them offered contracts regarding the U.S. presidential election of 2008.

For our analysis, trading data was collected from the relevant contracts from the following eight online prediction markets: the Iowa Electronic Markets, Intrade, Betfair, Bet2Give, NewsFutures, Hubdub, the Foresight Exchange and Inkling Markets. These eight were chosen mainly for their popularity and the availability of data with regards to the presidential elections but, conveniently for our purposes, these eight also differ significantly in how they function and incentivize trading. The differences in how these markets function and how these differences affect their prediction making ability will be described later in detail. Table I early in this paper showed key facts about these markets side-by-side. The working of each market and the collected data will be explained more in detail as it is used, the amount of detail given for each market is dependent on how much the data is used.

As can be seen from Table I all are online markets founded in the last 20 years. Four of the markets use real-money, while the other four use play-money. They differ in which incentives they give to traders, how the trades are carried out and how the traded contracts are created.


2.3 Comparing markets using winner-take-all contract data

As described in Section 1.5.1, a winner-take-all contract is a contract with two possible outcomes where the correct contract is worth ‘everything’ (usually 100) and the incorrect contract is worth nothing. During the 2008 presidential election this type of contract was very popular due to its simplicity and the fact that the price can be displayed as the probability of the underlying event occurring.

For the 2008 presidential election there were two main types of winner-take-all contracts traded; those for individual candidates to win (for example, Obama to win the election) and those for parties to win (for example, for the Democratic Party to win the election). Data on individual candidates to win was collected from Intrade, Betfair, Hubdub and Inkling Markets. We will look only at those contracts for Barack Obama and John McCain since these were the two candidates that received their party’s nomination. We ignore contracts for third-party candidates that also ran in the election since their chances of winning were very small and the volume in these markets low. Data for parties to win (we will look only at those contracts for the Democratic Party and the Republican Party) was collected from Intrade, the IEM, Bet2Give, NewsFutures and the Foresight Exchange.² A summary of the main characteristics of these markets has already been presented in the previous part of this paper in Table I.

This section will exhibit this data from these markets and compare the data between the different contracts and across the markets to see what we can learn from it.

2.3.1 Intrade winner-take-all data

Figure 1 shows the Intrade winner-take-all contract prices for Obama and McCain for the whole time range of the contract (from 24/10/2006 to 06/11/2008, 743 observations for each contract), these prices are the closing prices for each day. For the Obama contract, we see a total volume of 1,244,882 contracts with an expiry volume of 526,866 when the contract closed at 100. The McCain contract had a total volume of 1,489,313 contracts and an expiry volume of 599,611 when the contract closed at 0.

²The first two offered contracts for both the Democratic Party and the Republican Party, while for the other three markets, only data for the Democratic Party contract was available.

³Other winner-take-all contracts were available on Intrade for Allen, Biden, Bloomberg, Clinton (Hillary), Dodd, Edwards, Gingrich, Giuliani, Gore, Huckabee, Paul, Richardson, Romney,
We see large changes in the contract prices (and thus in the market’s expected probability of either candidate winning the election) over the life of the contract as more information becomes available. Due to the thickness of the market we would expect very fast incorporation of new information into the prices.

Until the end of 2007 there were still many candidates that could receive the nomination from their party and so the prices of the Obama and McCain contracts were still very low. The dates of nomination are shown as vertical solid lines, the first line on March 4\(^{th}\) 2008 shows when McCain was nominated to be the Republican candidate and the second line on June 3\(^{rd}\) 2008 shows when Obama was nominated to be the Democratic candidate. To allow us to better analyze this data we can limit the time range from the start of 2008 until the election (on November 4\(^{th}\) 2008), this can be seen in Figure 2. Both candidates’ prices start low at the beginning of the year (since neither had received their party’s nomination yet), and once both are nominated (shown again by the vertical solid lines) both contract prices increase and sum to nearly 100 (since either one of them is expected to win the election).

Thompson, Warner and a contract for the ‘rest of the field’ (all other candidates).

**Figure 1:** Intrade winner-take-all contract daily closing prices for Obama and McCain for the whole time range of the contracts (24/10/2006 to 06/11/2008)
Figure 2: Intrade winner-take-all contract daily closing prices for Obama and McCain from the beginning of 2008 until the election

Again, the vertical solid lines show the dates of nomination but also the vertical dashed lines indicate various other events in the election campaign and show that prices quickly adapt when new information becomes available. The first line shows a spike in the Obama contract, this was when he won the Iowa Democratic caucuses on January 3\textsuperscript{rd} greatly increasing his chance of receiving the nomination. The second line, shows a sharp decrease in the Obama contract price when Clinton won the New Hampshire primary on January 8\textsuperscript{th}. The third line represents Super Tuesday, on February 5\textsuperscript{th}, when 21 states held their primaries. McCain performed particularly well and so his contract price increased. The fourth dashed line shows a large increase in the Obama contract price, this was around the time of the North Carolina Democratic primary on the 6\textsuperscript{th} of May when it became clear that Obama would almost certainly become the Democratic candidate. The fifth dashed line, on August 29\textsuperscript{th}, shows when McCain announced that he had chosen Palin to be his running mate leading to the “Palin Bounce” and an increase in his contract price. The last dashed line shows the only point late in the election where the McCain contract traded higher than the Obama contract, but from this peak (on the 15\textsuperscript{th} of September when Lehman Brothers went bankrupt), the McCain contract continued to decease
in price to its closing price at 0 and the Obama contract continued to increase to its closing price at 100.\(^4\)

Since these are winner-take-all contracts, we can compare the contract prices at any point in time to the outcome of 100 for Obama and 0 for McCain. A rudimentary way of doing this is using the **average absolute prediction error** (AAPE); we compare the prediction of the market on each day to the final outcome of the election and take the average of these absolute errors. We find, that over the whole time range, the AAPE is 69.2 for Obama and 20.7 for McCain. For just 2008, the values change to 45.6 for Obama and 33.4 for McCain and since the date of each candidate’s nomination, the value is 34.9 for Obama and 35.1 for McCain. It is obviously difficult to draw any conclusions from this, Obama’s error is initially high since he wasn’t seen as a viable candidate early in the election. The predictions proved to be wrong but that does not mean that at the time they were bad predictions since they were constructed using only the data available at the time.

Another simple way of judging the accuracy of the prices is to look at how often the contract price of the Obama contract was higher than the contract price of the McCain contract (we interpret this as the market giving Obama a higher probability of winning the election than McCain). We see that this is the case on 79% of the days since the beginning of the data set and on 92% of the days since the start of 2008. Again, it is difficult to draw any conclusions from this but it gives an indication of the predictive accuracy of the market.

Our conclusion is that while the winner-take-all data is useful for showing the trend in prices and how prices are affected by new information, we are not able to judge the prices for accuracy against the final election outcomes. We have no way of deciding what constitutes an accurate prediction. A market that overvalues Obama, for example; a market that always gives Obama a 100% chance of winning the election from the start of 2008 until the day of the election, will have an error of zero even though it is clear that this is not an accurate prediction. For the Obama contract we have no way of distinguishing between an overvaluation and an accurate prediction. The same is true for the McCain contract, we cannot distinguish between what is an accurate prediction and what undervalues his chances of winning.

This does not mean to say that the data is useless. In the next section, we use winner-take-all data to compare contracts and to compare markets. The judging of accuracy however will be left for a later section where we use vote-share data which

\(^4\)This list of dates was compiled by Betfair. For a similar graph using Betfair data see: [http://predicts.betfair.com/2008/11/how-barack-obama-won-the-presidency/](http://predicts.betfair.com/2008/11/how-barack-obama-won-the-presidency/)
Figure 3: Intrade, Betfair, Hubdub and Inkling Markets contract prices for Obama and McCain from the beginning of 2008 until the election can be compared to the outcome to judge its accuracy.

2.3.2 Comparing winner-take-all contracts for individual candidates across markets

Figure 3 is an expansion of Figure 2 which allows us to compare the contracts across different markets. The figure shows Intrade winner-take-all contract prices for Obama and McCain from the start of 2008 until election day but also includes the data from comparable contracts on Betfair, Hubdub and Inkling Markets.\(^5\) Here it is important to note that Intrade and Betfair use real-money while Hubdub and Inkling Markets use play-money. A second important difference is that both Intrade and Betfair use a continuous double auction as trading mechanism while Hubdub and Inkling Markets both use versions of Hanson’s market scoring rule. The reader is referred to Section 1.5.4 for a discussion of the difference between real- and play-

\(^5\)For Inkling Markets only data on the Obama contract was available. The other markets had contracts for both Obama and McCain. A full comparison of these markets can be seen in Table [1](#).
Table III: Correlations between Obama winner-take-all contracts on various markets (calculated from begin 2008 or all available data when the market started later)

<table>
<thead>
<tr>
<th></th>
<th>Intrade</th>
<th>Betfair</th>
<th>Hubdub</th>
<th>Inkling Markets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrade</td>
<td>–</td>
<td>0.977</td>
<td>0.840</td>
<td>0.860</td>
</tr>
<tr>
<td>Betfair</td>
<td>0.977</td>
<td>–</td>
<td>0.842</td>
<td>0.889</td>
</tr>
<tr>
<td>Hubdub</td>
<td>0.840</td>
<td>0.842</td>
<td>–</td>
<td>0.810</td>
</tr>
<tr>
<td>Inkling Markets</td>
<td>0.860</td>
<td>0.889</td>
<td>0.810</td>
<td>–</td>
</tr>
</tbody>
</table>

As can be seen in Figure 3, the daily data from Intrade and Betfair were available from the beginning of 2008 while the Inkling Markets data starts on 22/06/2008 and only data on the winner-take-all contract for Obama was available. The data from Hubdub starts on 29/08/2008.6

The first thing that can be noticed is how well the contract prices for Intrade and Betfair move together. The Obama contracts have a correlation coefficient of 0.977 while the McCain contracts have a correlation coefficient of 0.967 from the start of 2008 until election day. These clearly move very closely together and react to new information equally quickly and at the same time. The absolute average difference in the predictions of Intrade and Betfair is 3.0 in the case of the Obama contracts and 3.2 in the case of the McCain contracts.

The correlation coefficients for all the Obama contracts are shown in Table III.7 The correlation coefficient between the Intrade Obama contract and the Hubdub Obama contract is 0.840 and between the Intrade Obama contract and the Inkling Markets Obama contract we find a coefficient of 0.860 (based on the time ranges available). These move together less well, most likely due to less informed trades being made by traders who do not use real-money. The highest coefficient is that of Intrade and Betfair, the two real-money markets, and the lowest coefficient was between Hubdub and Inkling Markets, the two play-money markets.

As we have clearly demonstrated in the previous section, it is difficult to judge the absolute accuracy of a winner-take-all contract. Instead we try to compare the predictions of the two markets using a simple binomial test to see which is closest to the outcome. Comparing the predictions of two markets, on every day the one that

---

6All calculations were done using data from the start of 2008 when available or from the start of the contract when not.

7Since similar results were found for the McCain contracts these are not shown here.
Intrade Betfair Hubdub Inkling Markets

<table>
<thead>
<tr>
<th></th>
<th>Intrade</th>
<th>Betfair</th>
<th>Hubdub</th>
<th>Inkling Markets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrade</td>
<td>–</td>
<td>4.35%</td>
<td>14.49%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Betfair</td>
<td>95.65%</td>
<td>–</td>
<td>33.33%</td>
<td>1.45%</td>
</tr>
<tr>
<td>Hubdub</td>
<td>85.51%</td>
<td>66.67%</td>
<td>–</td>
<td>26.09%</td>
</tr>
<tr>
<td>Inkling Markets</td>
<td>100%</td>
<td>98.55%</td>
<td>73.91%</td>
<td>–</td>
</tr>
</tbody>
</table>

**Table IV:** Binomial tests: percentage of days where the row market was closer to the outcome than the column market for Obama contracts from 29/08/2008 until the election is closer to the outcome receives a 1 and the other receives a 0, by summing these results for all the days we can see how often each market was closest to the eventual outcome. Days where both markets were equally accurate are not included in these calculations.

For the whole time range available, we see that the price of the Obama contract on Intrade was closer to 100 than the Obama contract on Betfair on 120 out of the 311 days; so only 38.59% of the time. For the McCain contract we see that the Intrade prediction is closer to 0 than that of Betfair for 14 out of the 311 days so only 4.50% of the time. From this we can clearly see that in this case Betfair outperformed Intrade on this measure. Again we face the problem that we can only compare to the outcome of 0 or 100, we again face the difficulty that we cannot distinguish between an overvaluation and an accurate prediction. A market that overvalues Obama and undervalues McCain will score better on this measure than a market that makes ‘accurate’ predictions.

Despite this obvious shortcoming, we apply this methodology to the other markets only for the Obama contracts (since these are available for all markets) and only on the data from the time when all the contracts were available (29/08/2008 until 05/11/2008; 69 observations for each contract). The results are shown in Table IV where the percentage represents the percentage of the days where the row market was closer to the outcome than the column market. The most interesting thing we can see from this table is the surprising ‘accuracy’ of the play-money markets relative to the real-money markets. For example, Inkling Markets beats Intrade every day in the sample while Intrade is beaten by all three other markets. On paper, we would expect Intrade to be more efficient than Inkling Markets due to its use of real-money and its larger volume. Intrade is designed as an investment market while Inkling Markets is designed mainly for entertainment.
What we have shown, however, is not that the predictions of Inkling Markets are more accurate than those of Intrade (winner-take-all data does not allow us to do this). All we have shown is that the Obama contract is valued higher on Inkling Markets, this does not mean that the price was more accurate. While we cannot show this, it would be relatively safe to assume that the prices on Intrade were more accurate than those on Inkling Markets. One possible explanation for the overvaluation could be that the traders are biased (see Section 1.6.2 for more information on this) and will therefore tend to overvalue the Obama contract. The wishful thinking effect, false consensus effect and assimilation-contrast effect could hold true on these markets. The theory states that if a trader prefers Obama to McCain they will overvalue Obama’s chances of winning the election and drive up the contract price higher than objectively it should be. We would expect these biases to be stronger on play-money markets than on real-money markets since having to risk real-money may encourage individual traders to compensate for their biases and an overvaluation leads to profit opportunities for more rational traders. An interesting way of testing our hypothesis would be if we had data on the political preferences of the traders; an out of proportion preference for Obama over McCain could offer an explanation for our results, but these data are not available.

2.3.3 Comparing winner-take-all contracts for individuals to parties

Another interesting use of the winner-take-all data is that we can compare the prices of the contract for the individual candidates (Obama and McCain) to the prices for their political parties (the Democratic Party and the Republican Party respectively) within one prediction market. We would expect the prices of the candidate to converge to the price of their party as it becomes more and more certain that they will be nominated. Once they officially receive their party’s nomination we would expect the price for the candidate and the price for the party to be very close together and to move identically. Any difference in the prices can only be explained by uncertainty over whether the nominated candidate will actually be the candidate on the day of the election.

Figure 4 shows contract prices from Intrade for Obama and McCain, and for the Democrats and the Republicans for the whole time range where we have data available for all contracts (from 24/10/2006 to 06/11/2008). Also shown are two vertical lines, the first (red) shows the date on which the nomination of McCain as the Republican candidate was confirmed (March 4th 2008) and the second vertical line
Figure 4: Intrade contract prices for Obama and McCain and for the Democrats and Republicans for the whole time range of the contracts

(blue) shows the date on which Obama was confirmed as the Democratic candidate (June 3\textsuperscript{rd} 2008). While these are the actual dates when it was confirmed who would be the candidate, it was sometimes already nearly certain before these dates, see for example the McCain-Republican lines.

We confirm our hypothesis; once the candidate has certainty that they will be nominated by their party, the contract price for the candidate and for their party are very close and move together. After the nomination, differences between the contract prices show that there is uncertainty over whether the nominated candidate will still be the candidate on the day of the election. After Obama was certain to be the Democratic candidate, we find a correlation coefficient of 0.993 between the Obama and Democratic contract, for McCain and the Republican contract this is 0.992 after he is certain to receive their nomination. The difference between the Obama and Democratic contracts is often greater than that of the McCain and Republican contracts, this can be explained by there being more uncertainty over the Obama nomination than over the McCain nomination.
Comparing winner-take-all contracts for parties across markets

Similar to the comparison of Obama and McCain winner-take-all contracts, we can also compare winner-take-all contracts for the Democratic Party and the Republican Party across markets.

Figure 5 shows contract price data only for Intrade and the Iowa Electronic Markets (IEM) for the whole available time range (02/06/2006 to 06/11/2008) for the Democratic Party and the Republican Party. Both markets are similar in design; both use real-money and use a continuous double auction. They differ in how they introduce contracts to the market but the significant difference is that the IEM has a limit on investments of ($500) while Intrade has no significant limits (see Table I for more details). Again, we can conclude that the comparable contracts on both markets tend to move together; the Democratic contracts have a correlation coefficient of 0.824 while the Republican contracts have a correlation coefficient of 0.887, these are lower than in the Obama-McCain case seen previously but still high.
Figure 6: Intrade, the IEM, Bet2Give, NewsFutures and the Foresight Exchange contract prices for the Democratic Party from begin 2008 until the election.

Using again the simple binomial test used previously, we find that the Intrade Democratic contract makes better predictions than the IEM Democratic contract on 41.08% of the days while the Intrade Republican contract makes better predictions than the IEM Republican contract on 66.65% of the days. The same criticisms of this method apply and with only these winner-take-all data it is not possible to decide which market makes more accurate predictions.

Finally we look at Figure 6 which shows Intrade, the IEM, Bet2Give, NewsFutures and the Foresight Exchange contract prices for the Democratic Party from the beginning of 2008 until the election. We only show the Democrats since all the markets had a contract for the Democrats but not all the markets had a contract for the Republicans. The most interesting thing about this comparison is that it involves data from two markets using real-money (Intrade and the IEM), two markets using play-money (NewsFutures and the Foresight Exchange) and Bet2Give (which uses real-money with all profits donated to charity). Again the reader is directed to Table I for more details.

Again, we see that the prices of these contracts all move together. Table V shows the
Table V: Correlations between Democratic winner-take-all contracts on various markets from the beginning of 2008 until the election

<table>
<thead>
<tr>
<th></th>
<th>Intrade</th>
<th>IEM</th>
<th>Bet2Give</th>
<th>NewsFutures</th>
<th>Foresight Exchange</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrade</td>
<td>–</td>
<td>0.885</td>
<td>0.719</td>
<td>0.921</td>
<td>0.953</td>
</tr>
<tr>
<td>IEM</td>
<td>0.885</td>
<td>–</td>
<td>0.685</td>
<td>0.867</td>
<td>0.847</td>
</tr>
<tr>
<td>Bet2Give</td>
<td>0.719</td>
<td>0.685</td>
<td>–</td>
<td>0.783</td>
<td>0.755</td>
</tr>
<tr>
<td>NewsFutures</td>
<td>0.921</td>
<td>0.867</td>
<td>0.783</td>
<td>–</td>
<td>0.895</td>
</tr>
<tr>
<td>Foresight Exchange</td>
<td>0.953</td>
<td>0.847</td>
<td>0.755</td>
<td>0.895</td>
<td>–</td>
</tr>
</tbody>
</table>

Table VI: Binomial tests: percentage of days where the row market was closer to the outcome than the column market for Democratic contracts

<table>
<thead>
<tr>
<th></th>
<th>Intrade</th>
<th>IEM</th>
<th>Bet2Give</th>
<th>NewsFutures</th>
<th>Foresight Exchange</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrade</td>
<td>–</td>
<td>71.47%</td>
<td>49.50%</td>
<td>75.16%</td>
<td>19.67%</td>
</tr>
<tr>
<td>IEM</td>
<td>28.53%</td>
<td>–</td>
<td>43.93%</td>
<td>55.56%</td>
<td>20.06%</td>
</tr>
<tr>
<td>Bet2Give</td>
<td>50.50%</td>
<td>56.07%</td>
<td>–</td>
<td>69.83%</td>
<td>42.95%</td>
</tr>
<tr>
<td>NewsFutures</td>
<td>24.84%</td>
<td>44.44%</td>
<td>30.17%</td>
<td>–</td>
<td>13.73%</td>
</tr>
<tr>
<td>Foresight Exchange</td>
<td>80.33%</td>
<td>79.94%</td>
<td>57.05%</td>
<td>86.27%</td>
<td>–</td>
</tr>
</tbody>
</table>

correlations between the prices of one contract and the prices of another contract. As we can see, the markets where the Democratic contract is the most correlated are Intrade and the Foresight Exchange with a coefficient of 0.953, the ones that are least correlated are the IEM and Bet2Give with a coefficient of only 0.685. Looking at Figure 6, another interesting observation is the large difference between the highest and lowest prediction, often this is greater than 10 percentage points at the beginning of 2008 but decreases closer to the day of the election.

As we did previously, we can compare the performance of the markets by using a binomial test to see which market is closest to the outcome on each of the days. This can be seen in Table VI, showing the percentage of the days where the row market was closer to the outcome than the column market. Here we see that the Foresight Exchange ‘outperforms’ all the other markets, this means that its predictions are the closest to the outcome of 100. As mentioned previously, this could be explained, not by the accuracy of the market but by over-predictions for the Democrats caused by biased traders.
2.4 Other types of contract

The contracts relating to the U.S. presidential election of 2008 were not just winner-take-all or vote-share contracts for Obama and McCain or the Democratic Party or the Republican Party. Intrade in particular ran many different markets; on voter turnout, various other candidates to win, vice-presidential candidates (Biden or Palin) to be withdrawn, who would control the House and Senate after the election etc.. In this section we look at two other markets run by Intrade to give a brief example of what else can be done with prediction markets. Firstly, we look at the market on the number of electoral votes that each party would receive in the election and then at the markets run for each state to predict which party would win the electoral votes of each state.

Intrade ran one market offering contracts on “2008 Democratic Electoral College Votes”, these were winner-take-all contracts that would pay 100 if they were correct and 0 otherwise. They ranged from $\geq 160$ electoral college votes to $\geq 440$ electoral college votes, increasing in intervals of 10. For example, the contract $\geq 300$ would expire at 100 if the Democrats received greater than or equal to 300 electoral votes,
if they received less than 300, the contract would expire at 0. This is interesting to
us because, instead of just knowing what the likelihood is of a party winning the
most electoral college votes, we have the whole probability distribution so we can
calculate what the probability is of the party getting each number of votes.\footnote{There was also a market for “2008 Republican Electoral College Votes” also offering contracts from $\geq 160$ to $\geq 440$. Since all the electoral votes are usually split between the two parties, in theory these should be the exact opposite of the Democrats, although in practice this might not hold. These data were not available.}

Figure 7 shows the contract closing price for 18 different contracts traded on this
market on the day exactly one month before the election (on 04/10/2008). It shows
the values of contracts from $\geq 210$ to $\geq 380$. The other contracts did not have prices
on this date due to lack of trading; volume in this market was much lower than in
the general markets. The vertical line at 365 shows the actual number of electoral
votes that the Democratic Party received. At this point in time, the contract for
the highest number of votes that had a value over 50 was at $\geq 310$, telling us that
the market expected the Democratic Party to receive between 310 and 319 electoral
votes, this is 55 to 46 lower than the actual outcome and shows that quite a lot
changed in the last month of the election. The price of the highest contract that
ended up being correct (the “$\geq 360$” contract) was only 22.5 one month before the
election.

Another observation is that the contract price decreases as the number of electoral
votes that the contract needs to win, increases. This is what we would expect; we
expect each bar to be higher than the bar to its right for prices to be consistent,
otherwise arbitrage would be possible. We do see that in general this holds in Figure
7 but there is an exception at $\geq 340$, this can most likely be explained by the bid-ask
bounce due to the low liquidity and thus large spread in these markets.

Figure 8 is also based on data collected from this same market and shows the value of
the contract for the highest number of electoral college votes for the Democrats that
has a price greater than or equal to 50 on each day from 08/06/2008 to 07/11/2008
(this is the total time range that the contracts were traded). The horizontal line at
365 shows the outcome of the election, which means that the highest contract that
‘won’ was that for $\geq 360$. We see that the market was surprisingly far off at the start
of this time period; predicting $\geq 280$ votes, 80 off from the winning contract. On
15/09/2008 the prediction even dropped to 260, 100 off from the winning contract.
The first time that the highest winning contract ($\geq 360$) had a value greater than
50 was on 18/10/2008 but then there was still a lot of uncertainty as to exactly how
many votes the Democrats would receive in the election.
This brief look at the contracts on the number of electoral votes leads us to believe that the market did not do a particularly good job in predicting the outcome. This can be explained by the characteristics of the election (the Democrats proved to be very successful in the end, more so than was predicted by most sources) and by failings in the market; the electoral college market had a low liquidity due to the low number of traders. The $\geq 360$ contract for the Democrats only had a total volume of 4,633 contracts compared to 156,117 for the Democratic winner-take-all contract. By looking at all the possible numbers of electoral votes we gain information in the form of the complete probability distribution but we lose information by fragmenting the market and having less traders interested in the contract, therefore there is a loss of liquidity.

Another market that Intrade ran during the 2008 election was the “US Election by State”. For each of the 50 states and Washington D.C. there was a contract for the Democrats to win the majority of the state’s electoral votes, the Republicans to win the state’s electoral votes and for the Field (any other party) to win. These thus give a prediction for each state of what the probability is of each party winning the electoral college votes of that state at any point in time. Figure 8 shows these

Figure 8: Predicted Democratic electoral votes from 08/06/2008 to 07/11/2008
Figure 9: Maps showing Intrade predictions for winner of the majority of the electoral college votes per state.
predictions at three different points in time; one year before the election (panel a, on 04/11/2007), one month before the election (panel b, on 04/10/2008) and on the eve of the election (panel c, 03/11/2008). Also shown is the actual outcome of the election (panel d). The colors of the states depend on the value of the contract according to the key in the figure.\(^9\)

Looking at the election eve map and the market’s predictions, we see that the market only predicted one state incorrectly, this was Indiana which the Democrats won and had not won since 1964. It was expected that they would not win this time either. The values of the Indiana contracts were 47 for the Democrats and 53 for the Republicans so the market was not far off. One month before the election the market predicted 49 out of the 51 states correctly and for one year before the election the market predicted 43 states correctly, 3 incorrectly and 4 were tied (50-50, most likely due to lack of trading).

The market here proved to be relatively accurate at predicting who the states would vote for. The trading volume was very much dependent on the state, this is as we would expect. The higher the uncertainty as to who the state will vote for, the more it is traded. There is not much interest in trading in a state where the outcome is certain. For example, none of the contracts in the Washington D.C. market were traded while the Indiana market had a total trading volume of 63,030 contracts.

We collected similar data from Inkling Markets. On the eve of the election, the market predicted 49 out of the 51 states correctly (Indiana was also incorrect and so was Missouri), one month before the election they also predicted 49 out of 51 states correctly and one year before the election 45 were correct. These predictions were slightly worse than Intrade but not by much.

### 2.5 Comparing prediction markets to polls using vote-share data from the IEM

The fundamental question that still needs to be addressed is how accurate prediction markets were in predicting the outcome of the presidential election. As we have discussed in the previous sections, data from winner-take-all contracts for individual candidates or parties cannot be judged for accuracy against the outcome of the

\(^9\)Figure 9 was inspired by the maps on: [http://electoralmap.net/](http://electoralmap.net/) where the Intrade predictions for the states can be seen for various days from 24/05/2008 until the election. However, Figure 9 was made using our own calculations from the Intrade data.
election. To solve this problem, we now look at vote-share contracts which, instead of attempting to predict who will win the election, try to predict which percentage of the vote each party will receive. The data from these types of contracts can be judged for their absolute accuracy and, more importantly, can also be compared to other ways of predicting the election outcome such as polls, which have traditionally been used to forecast election results.

For this comparison we make use of data from the vote-share market of the Iowa Electronic Markets (IEM). None of the other major prediction markets ran a vote-share market, so this is the only data that can be used for this purpose. Since these data are used extensively, first the market and its functioning are explained in detail.\footnote{For more details, see the IEM Trader’s Manual: \url{http://www.biz.uiowa.edu/iem/trmanual/}}

\subsection{2.5.1 About the Iowa Electronic Markets (IEM)}

The Iowa Electronic Markets (IEM)\footnote{http://www.biz.uiowa.edu/iem/} are a group of prediction markets operated by the University of Iowa Henry B. Tippie College of Business. The markets are not-for-profit and are run for educational and research purposes. Participation in the market is aimed particularly at students, faculty and staff at educational institutions around the world but is in practice open to anyone willing to participate via its website.

Since 1988, these markets have been used to predict election outcomes, these are the longest running online prediction markets known to us. The IEM is best know for its political markets on U.S. elections but has also run markets on foreign elections, box-office receipts of movies, economic indicators and others.

\subsubsection*{How the IEM function}

Anyone willing to trade in the market can create a trading account by filling in the on-line sign-up form. An account activation charge of 5.00 U.S. dollars ($) is charged when opening a new account but after this (with the exception of an ‘inactivity fee’ if the account is not used) no commissions or transaction fees are charged on trades. When the trading account is opened the trader can deposit real-money into the trading account with a minimum of $5 and a maximum of $500. Traders can add more money to their accounts at anytime, provided they do not exceed the maximum limit. The use of real-money has been discussed in Section \ref{sec:real-money} and the existence of trading limits has been discussed in Section \ref{sec:trading-limits}.
Trading

The IEM uses the WebExchange Trading System for operation of the market and all trades are done electronically via the website of the market using a continuous double auction (as described in Section 1.5.2). As in most financial markets, traders can trade using limit orders (place a bid, place an ask, or withdraw an outstanding bid or ask) or market orders (buy or sell at the current market price).

The IEM does not allow a trader to sell short or buy on margin, this means that a trader cannot sell more contracts than they currently own or spend more money than they have in their account. Any orders violating this are considered as ‘infeasible orders’ and are simply cancelled (traders can do something similar to short selling by purchasing a unit portfolio and selling the contract which they would like to short).

Information available to traders

The traders on the IEM are subject to information constraints, this is important when discussing issues of market microstructure. The only public information available to the traders is the current high bid price, the current low ask price and the last trade prices. They also know the number of contracts in circulation in each market. The historical data of the market are also available, this includes summaries of prices and trading volumes over previous time periods. The traders can therefore not observe the whole ask queue or bid queue in the order book and only see the current best prices. All trading and all bids to buy and offers to sell are also anonymous, so no trader can know who they are trading with or who is buying or who is selling.

Population of traders

The markets were originally designed for academic purposes so the traders were mainly from the academic community, but since then many outsiders have started trading in the market. Berg and Rietz (2006) give various statistics about the traders in the market. They write that in the 2004 presidential election markets, 44% of the traders in the vote-share market were non-academic. They also point out that “traders are typically more educated, richer, and “more male” than both the average U.S. citizen and the average U.S. voter” (Berg and Rietz, 2006, pp.145-146). They, clearly, do not represent a representative sample of the U.S. population. For more discussion on the population of traders see Section 1.5.5.
2.5.2 The U.S. presidential election of 2008

Contracts and payoffs

The IEM ran two different markets in relation to the U.S. presidential election of 2008; a winner-take-all market for the Democrats and Republicans (these data have been used previously and will no longer be looked at here) and a vote-share market.

The vote-share market (PRES08_VS) involved trade in two contracts, one for the Democratic Party and one for the Republican Party: UDEM08_VS and UREP08_VS. The first contract had a payoff of “$1.00 times two-party vote share of unnamed Democratic nominee in 2008 election”, it is thus linearly related to the vote-share received by the party. It is important to mention that the contracts refer to the two-party vote-share, this means that votes for the other parties (other than the Democrats or Republicans) are not looked at. The percentage used for the Democratic contract is the number of votes received by the Democratic Party divided by the sum of the number of votes received by the Democratic Party and the Republican Party. All contract price data were listed as decimals since the contracts are priced between 0 and 1 but, to remain consistent with the prices on other markets, we multiplied all prices by 100 so that they vary between 0 and 100 instead and can be read as percentages.

All contracts on the IEM are issued in the form of unit portfolios which can be purchased from the market at any time by traders. These unit portfolios consist of one of each of the contracts that is traded in the market and has a price equal to the sum of the liquidation values of all the contracts that it contains; in this case 1. For the presidential vote-share market each unit portfolio contained one contract for the Democrats and one contract for the Republicans. The cost of the unit portfolio will be equal to the value of these contracts once the election has been decided. Since both contracts offer a payoff of $1 if the party wins and a payoff of $0 if they do not win, the unit portfolio will be worth exactly $1 regardless of which party wins and so the unit portfolio will be valued at $1 at all times. Once a trader has purchased a unit portfolio from the IEM, they can trade the individual contracts in the market. Contracts can also be removed from the market by selling a unit portfolio back to the market. In this way, the number of each contract in the market will always be the same (there will always be as many contracts for the Democrats as for the Republicans).
Data

Since the IEM was designed for research purposes, all the data that was required for our analysis was available directly from their website.\(^\text{12}\)

For each contract, information was available on; the number of contracts traded each day, the total volume for the day in U.S. dollars, the low, high, average and last prices of the day. For the purpose of our analysis, the last price of the day was used (this can be compared to the closing price of a standard financial market).

For the vote-share contract data set, the first day where both contracts were non-zero was 08/06/2006 and the last observation was on 07/11/2008.

The data set consisted of one observation (the last price of the contract of the day) for each day in the time range. However, the data set contained missing data, these are on the following days: 11-12-13/05/2007, 13-14-15/07/2007 and 29-30/10/2007. The reason for these missing data is unknown but since it only accounts for 8 observations a long time before the election, these missing data were replaced with ‘na’ in the data set and the observations were simply ignored. Another error was that the observation on 04/12/2007 for the Republican vote-share last price was zero, this observation was also removed. Other anomalies (such as on 01/05/2008 when the last price of the Republican vote-share contract spiked to 0.740, this can be clearly seen in all the following figures) were not altered since these were most likely driven by traders in the market and not by errors in the system. Therefore, for each of the contracts there were 891 observations.

Figure 10 shows the daily price graph for this vote-share market (using the last price of the day). The horizontal axis shows the date while the vertical axis shows the price of the contracts. The market starts off giving the Democratic Party a higher predicted percentage of the two-party vote than the Republican Party. The horizontal dashed lines show the actual two-party vote-shares that each party received in the election, we notice that the opening prices for both contracts were surprisingly close to the actual outcome. The figure also shows the total volume traded in the market for each day (measured in total units traded in both contracts), we see that the volume increases drastically closer to the election.

\(^{12}\)http://www.biz.uiowa.edu/iem/markets/data_Pres08.html
2.5.3 Absolute accuracy of the IEM vote-share predictions

The question that we must try to answer is how accurate the IEM was in predicting the outcome of the presidential election. Since we are now using vote-share data, we can judge the market’s accuracy compared to the election outcome. For this we calculate the average absolute prediction error (AAPE); we compare the prediction of the market on each day to the final outcome of the election and take the average of these absolute errors.

The final outcome of the election showed that Barack Obama and the Democrats received 69,498,952 votes with a total of 52.87% of the popular vote while John McCain and the Republicans received 59,949,402 votes giving him a total of 45.60%. This means that 2,008,577 votes were given to other candidates or 1.53%. Since the IEM vote-share contracts look at the two-party vote-share and not the share of the total votes, we have to convert these percentages and say that the Democrats received 53.68% of the two-party vote-share and the Republicans received 46.32%.

We find that over the whole time range of the market, the Democratic contract had

\[\text{http://www.uselectionatlas.org/RESULTS/}\]
an average absolute prediction error of 1.876 and the Republican vote-share contract had an AAPE of 1.781. Looking only at the observations from the beginning of 2008 until the day of the election, we find an AAPE of 1.651 for the Democratic contract and 2.037 for the Republican contract. The absolute accuracy of the market is an interesting statistic but does not tell us much without another value to compare it to.

2.5.4 Relative accuracy: comparing IEM predictions to polls

How accurate the market is in absolute terms is important, but what is more important is how accurate its predictions were, compared to other methods of predicting the election outcome. The most common way of forecasting an election is through the use of polling. By asking a large representative sample of the voting population how they intend to vote in the election and extrapolating the results to the population at large we can make an accurate forecast for the outcome of the election.

Polls in theory

Kou and Sobel (2004, p.287) discuss the conditions necessary for a poll to make accurate forecasts at times $t$ before the election (at time $T$). The forecasted vote-share of the poll $V_c(t)$ for candidate $c$ at time $t$ can be seen not as a forecast of the vote-share on the day of the election ($\theta_c(T)$) but as an estimate (with an error) of the vote-share ($\theta_c(t)$) that would be observed if the election were held at the time of the poll ($t$):

$$V_c(t) = \theta_c(t) + B(\theta_c(t)) + \epsilon_c(t)$$

(2.1)

where $B(\theta_c(t)) = E(V_c(t) - \theta_c(t)|\theta_c(t))$ is the systematic measurement error (bias of the poll) and $\epsilon_c(t)$ is the random error component.

They give five conditions under which both the random and systematic measurement error of using the poll to estimate the vote-share are zero (Kou and Sobel 2004, p.288):

1. A probabilistic sample of the eligible voting population has been taken.
2. Estimates adjusted for non-response are free of systematic measurement error.
3. Estimates adjusted for likely voting status are free of systematic measurement.
4. The procedure used to apportion the votes of the undecided voters is free of systematic measurement error.

5. There are no other sources of systematic measurement error.

If these conditions are met, the error is zero and $E(V_c(t)|\theta_c(t)) = \theta_c(t)$ so the poll makes accurate predictions.

However, even if these conditions are met, there still exists the possibility for polls to contain projection error. This comes about because polls estimate the vote-share as if the election were being held on the day of the poll instead of in the future, on the day of the election. This implies that if the date of the poll is far away from the date of the election, the voters preferences are not yet certain and the poll will not be able to form a good forecast of what will happen on the day of the election. Polls used for short-term forecasting suffer less from this error.

**Methodology**

In their paper, Berg, Nelson and Rietz (2008) use data from the IEM vote-share market for U.S. presidential elections from 1988 to 2004 and compare these to polls. The authors find that the market significantly outperforms the polls and that the market improves in relative accuracy, the longer the time until the election. The methodology that they used will be applied to the 2008 elections and compared to their results.

Data on the contract prices in the IEM vote-share market were collected from the IEM website. Since the market never closes (so there are no closing prices) the last price of the day is used; this is the last trade before midnight of a given day. Only the vote-share market data is used since this is easiest to compare to polls, which try to predict the percentage of the vote each party will receive and not the percentage chance of the party winning the election.

The polls used to compare to the market were all the nation-wide poll reports that

---

14 Apart from last prices we also had data of the low price, high price and average price for each day. Instead of using the last price, we could have used the average price or the midpoint between the high and the low price. The correlation between the last price and the average price is 0.959 and the correlation between the last price and the midpoint is 0.942. Repeating our analysis using any of these other measures had very little effect on our results and did not change our conclusions so we opted to use closing price to use data more comparable to financial markets.
Figure 11: All the polling data: predicted Democratic winning margin over time could be found.\textsuperscript{15} Where possible, we used data based on likely voters and two-way match-ups. Data from polls using data from “Registered Voters” and “All Adults” or polls offering a longer list of parties were also used, but preference was given to the first type. Following the methodology of \cite{Berg, Nelson and Rietz (2008)} we use raw polls (the polling numbers are used as published and are not adjusted to compensate for projection error), later we will look at other options in more detail.

The first observed poll had a last day-in-field of 07/11/2006 (this is the last date on which data for that poll report were collected) and the last observed poll, 07/11/2008. In total, we ended up with 357 unique observations of polls which were linked to the closing price of the IEM markets from the last day that the poll was in the field (following \cite{Berg, Nelson and Rietz (2008) we do not lag the polls), giving a total of 357 pairs of observations which could be compared. Figure 11 shows these polls measured by their predicted Democratic winning margin, the horizontal line shows the actual outcome of the election of 7.38\%. We see that the predictions of different polls differ greatly even when they were held on the same date, the consequences of this will be discussed in a later section.

\textsuperscript{15}The sources of these data were two websites: PollingReport.com and RealClearPolitics.com. \cite{Berg, Nelson and Rietz (2008) only use the first source, but the second source was used to increase the size of the data set. These websites only report the results of the polls, the polls were administered mainly by polling organizations (such as Gallup) or news agencies (such as CNN or ABC). All duplicates (polls reported by both websites) were removed from the data set.
Since the market and the polls give the predicted percentage of the popular vote received by either party, the first step is to convert this into the predicted winning margin. This is done from the point of view of the winning party (in this case the Democratic Party) and gives the predicted percentage of the votes (margin) that they are expected to win by (this is done simply by subtracting the percentage for the Republicans from the percentage for the Democrats). The problem with this is that there are other parties running in the election (other than the Republicans and Democrats) so we need to adjust the polls to represent only the two-party vote-share. This is done by splitting the percentage going to ‘other candidates’ or ‘undecided’ proportionally among the two parties. This is also done with the IEM vote-share data in the same way to make sure that the two contracts sum to 1.

The normalized poll margins are calculated as:

\[
s_{\text{Poll}}^{\text{Democrat-Republican},t} := \frac{r_{\text{Poll}}^{\text{Democrat},t} - r_{\text{Poll}}^{\text{Republican},t}}{r_{\text{Poll}}^{\text{Democrat},t} + r_{\text{Poll}}^{\text{Republican},t}}
\]

(2.2)

and the normalized IEM market price margins are calculated as:

\[
s_{\text{IEM}}^{\text{Democrat-Republican},t} := \frac{p_{\text{IEM}}^{\text{Democrat},t} - p_{\text{IEM}}^{\text{Republican},t}}{p_{\text{IEM}}^{\text{Democrat},t} + p_{\text{IEM}}^{\text{Republican},t}}
\]

(2.3)

Here, \(s\) represents what we will refer to as the normalized spread since it refers to the percentage of the two-party vote-share by which the Democratic candidate is predicted to win the election. \(s_{\text{Poll}}^{\text{Democrat-Republican},t}\) is the normalized spread of the polls at time \(t\) and \(s_{\text{IEM}}^{\text{Democrat-Republican},t}\) is the normalized spread of the IEM vote-share market at time \(t\). \(r_{\text{Poll}}^{\text{Democrat},t}\) is the poll prediction for the Democratic vote-share at time \(t\) and \(r_{\text{Poll}}^{\text{Republican},t}\) is the poll prediction for the Republican vote-share at time \(t\). Likewise, \(p_{\text{IEM}}^{\text{Democrat},t}\) is the IEM market price for the Democratic vote-share contract at time \(t\) and \(p_{\text{IEM}}^{\text{Republican},t}\) is the market price for the Republican vote-share contract at time \(t\).

We then compute the average absolute prediction error (AAPE) of each poll, using the formula:

\[
AAPE_{t}^{\text{Poll}} := \frac{|s_{\text{Poll}}^{\text{Democrat-Republican},t} - v_{\text{Actual}}^{\text{Democrat-Republican}}|}{2}
\]

(2.4)

and also calculate the average absolute prediction error for each observation from the IEM:

\[
AAPE_{t}^{\text{IEM}} := \frac{|s_{\text{IEM}}^{\text{Democrat-Republican},t} - v_{\text{Actual}}^{\text{Democrat-Republican}}|}{2}
\]

(2.5)
Here, we use the normalized spreads calculated with the previous formula and $v_{\text{Actual Democrat}}-\text{Republican}$ is the actual normalized spread that is the outcome of the election. The average absolute prediction error (AAPE) compares the prediction of either the poll or the market to the actual outcome of the election and gives us one number for each observation by which we can judge how accurate the prediction was. Then, for each of the observations (days in which we have a poll result and a market price) we ask which was closer to the actual outcome of the election; the poll or the market? This is the same as asking which has the smallest average absolute prediction error. Using a simple binomial test, for each observation we assign a 1 if the market has the smallest AAPE and a 0 if the poll has the smallest AAPE. Ties are broken in favor of the polls.\textsuperscript{16} By summing the results of the binomial tests, we can see how many times the IEM had a lower error than the polls and we can then see which makes better predictions on average. The number of observations is the number of polls and if more than one poll was released on the same day we compare all to the same market prediction for that day.

The significance of the results is shown by calculating the $p$-value of the binomial test. This is the exact binomial probability of a number of 1’s being observed that is as large or larger than the observed number, given the number of observations and a hypothesized probability of 0.50. They are based on the assumption that polls and markets preform equally well and so each has a 50% chance of being closest to the outcome.

Results

Figure 12 shows the implied vote-share margins for the IEM and selected polls.\textsuperscript{17} On the vertical axis we have the predicted Democratic winning margin (the normalized spread) and on the horizontal axis, the date. The continuous line shows the data from the IEM contracts and the dots represent polling data as the letters of various selected polling companies. The horizontal line shows the two-party vote-share margin by which the Democrats won in the election which was 7.38%.

The vertical lines on the figure draw attention to important dates similar to those discussed previously. The first is ‘Super Tuesday’ on the 5\textsuperscript{th} of February 2008,

\textsuperscript{16}It was decided to do this since we are studying the prediction market. In all the analysis that follows in this paper, we only observed 3 observations where there was a tie, so this decision had little effect on our results.

\textsuperscript{17}Not all the polls are shown to make the figure clearer, we opted for those by CNN, CBS, NBC, Hotline and Gallup.
Figure 12: Predicted Democratic vote-share winning margins for the market and selected polls from begin 2008 until the election

This was when the greatest number of states held primary elections to choose each party’s candidate. The next pairs of vertical lines represent the conventions, the first pair is the Democratic National Convention (from 25/08/2008 to 28/08/2008) and the second pair is the Republican National Convention (from 01/09/2008 to 04/09/2008). The three vertical lines after that represent the three presidential debates, these can have a large effect on the election and were held 26/09/2008, 07/10/2008 and 15/10/2008. For more details on the effect of the news on the price in the market, see Section 2.3.1. The data shown in the figure run from the beginning of 2008 and end on election day the 4th of November 2008, our data set starts earlier but due to the scarcity of polls this is not shown in the figure.

From the figure we conclude that the polls and market prices follow the same trend but again we see a large variance in the predictions of the polls; polls held on the same day can have very different predictions. This will be discussed more in a later section.

As described previously, for every observation the absolute average prediction error is calculated and it is seen whether the market or the poll was the closest to the final outcome. The results of this process are shown in Table VII. The first two columns
of data show the results from Berg, Nelson and Rietz (2008) for ‘All Years’ (this includes the U.S. presidential elections of 1988, 1992, 1996, 2000, and 2004) and the election of 2004. The third column (labeled A) shows the results when this same methodology described above was applied to our data for the 2008 election.

Looking at our results (panel A) we see that, for all the observations, the market is closer to the outcome of the election than the polls in 235 out of the 357 observations, or 66% of the time (this is significant at a 99% significance level) showing that the market appears to outperform the polls in the 2008 election. We are also interested in how the market performs relative to polls in other time ranges. Looking at only the observations that were made more than 100 days before the election we find the market wins 105 out of 160 observations, or 66% of the time (significant at 99%), this still favors the market but we would expect this percentage to be higher than that for all the observations since polls are expected to become less and less accurate as we move further away from the election. Looking at polls that were taken from 66 to 100 days before the election, we find the market only wins in 14 out of the 28 observations, or 50% of the time. We can attempt to explain this due to uncertainty in this election but possibly also by the small number of observations. Looking at the results from 32 to 65 days, we see that the market wins in 47 out of the 64 observations, or 73% of the time (significant at 99%) this is our highest percentage and shows that the market performed particularly well in this time period. Moving closer to the election, we look at the observations from 6 to 31 days before the election and find that the market wins on 51 out of 79 observations, or 65% of the time (significant at 99%) again the market performs well. Looking at just the last five days before the election we see that the market wins on 18 out of 26 observations or 69% of the time (significant at 95%). In general our results are favorable for the prediction market, we find the market significantly outperforms the polls in every period apart from the 66 to 100 day period.

Since we used the same methodology, we can also compare our findings to those of Berg, Nelson and Rietz (2008). The first two columns of results show their results from ‘All Years’ in their sample (1988-2004) and from 2004 alone. For all the days in the sample they find that the market wins in 74% of the cases for all the years and 70% of the cases for 2004 (both significant at 99%). They find the market performs better overall than our results show despite the similar number of observations between our 2008 and their 2004 data sets. Interesting in their results is that the period where they find that the market performs the best is the period

<table>
<thead>
<tr>
<th>Days included in sample</th>
<th>Item</th>
<th>Berg, Nelson and Rietz (A)</th>
<th>2008 Adjusted (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>77</td>
<td>2008</td>
</tr>
<tr>
<td></td>
<td>Number of polls</td>
<td>964</td>
<td>357</td>
</tr>
<tr>
<td></td>
<td>Poll “wins”</td>
<td>255</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td>Market “wins”</td>
<td>709</td>
<td>235</td>
</tr>
<tr>
<td></td>
<td>% market wins</td>
<td>74***</td>
<td>66***</td>
</tr>
<tr>
<td></td>
<td>p-value (1 sided)</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>More than 100 days</td>
<td>360</td>
<td>160</td>
</tr>
<tr>
<td></td>
<td>Number of observations</td>
<td>92</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Poll “wins”</td>
<td>268</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>Market “wins”</td>
<td>74***</td>
<td>66***</td>
</tr>
<tr>
<td></td>
<td>% market wins</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>66 - 100 days</td>
<td>131</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Number of observations</td>
<td>56</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Poll “wins”</td>
<td>117</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>Market “wins”</td>
<td>56</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>% market wins</td>
<td>84***</td>
<td>73***</td>
</tr>
<tr>
<td></td>
<td>p-value (1 sided)</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>32 - 65 days</td>
<td>173</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>Number of observations</td>
<td>58</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Poll “wins”</td>
<td>111</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>Market “wins”</td>
<td>111</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>% market wins</td>
<td>68***</td>
<td>73***</td>
</tr>
<tr>
<td></td>
<td>p-value (1 sided)</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>6 - 31 days</td>
<td>222</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>Number of observations</td>
<td>61</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Poll “wins”</td>
<td>161</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>Market “wins”</td>
<td>161</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>% market wins</td>
<td>73***</td>
<td>65***</td>
</tr>
<tr>
<td></td>
<td>p-value (1 sided)</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Last 5 days</td>
<td>78</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>Number of observations</td>
<td>25</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Poll “wins”</td>
<td>53</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Market “wins”</td>
<td>53</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>% market wins</td>
<td>68***</td>
<td>69**</td>
</tr>
<tr>
<td></td>
<td>p-value (1 sided)</td>
<td>0.001</td>
<td>0.038</td>
</tr>
</tbody>
</table>

*** p < 0.01, ** p < 0.05, * p < 0.1

Table VII: Comparing the IEM vote-share predictions to polls

66 to 100 days which is where we found our 50% result. Also interesting is that our result for the last 5 days is higher than either of theirs. Possible explanations for the differences in our results could lie in the characteristics of the particular elections, 2008 in particular had many uncertainties as to candidate nominations and vice-presidential nominations. However, it is impossible to prove what caused these differences.
2.5.5 Conclusions and criticisms

We conclude that our analysis of the 2008 election shows that the market provides better predictions of the election outcome than polls, with the market beating the polls on 66% of the days. Our results, however, are not as favorable for the prediction markets as those of [Berg, Nelson and Rietz]. In some periods, the market is the clear winner while in the time period 66 to 100 days we find that the polls and the market perform equally well at predicting the election outcome. In this section we offer a few criticisms of the results and particularly of the methodology used above (and by [Berg, Nelson and Rietz]) and suggest ways to remedy these.

Variable polls

One criticism is that the individual polls make very different predictions. Looking at Figure [11] it is clear to see that there is a very large difference in the predictions made by the polls. Even polls that were held on the same day can vary enormously in their predictions due to different polling techniques and different samples. Looking at Figure [12] we can also see that while the data from the IEM are quite volatile, the polling data are even more volatile. One explanation of this could be that the polling data come from many different polling agencies. Even though we use data only from respectable large-scale organizations, they do use different methods to collect their data, leading to different results. Looking at Figure [12] we see that even polls from the same polling agency can be very variable.

This leaves the methodology open to the criticism that the favorable results for the market were found by taking all the available polls together; the good and the bad. By looking at only the polls from a certain polling agency or an index of polls we might be able to solve this problem. After all, we do not look at many prediction markets but take data from only one (because only one offered a vote-share contract in this case). We will suggest a solution to this problem in Section 2.5.6.

Undecided voters

A second criticism involves the existence of undecided voters when the polls are administered. As discussed earlier, the poll predictions for the Democrats and Republicans did not always sum to 100, this is because there were other parties in the election but also because some voters were undecided at the time of the poll. To solve this problem we calculated the normalized two-party vote-split; the difference
of the sum of the two parties from 100 was divided between the two parties depend- 
ent on the percentage they each received in the poll. However, there are many 
different ways one could have solved the ‘undecided’ problem.

Looking at the poll data that we used, we find that the average absolute difference 
between 100 and the sum of the Democratic and Republican predictions is more than 
9%, meaning that on average 9% of the respondents in the polls said they would 
vote for a third-party or are undecided (for polls in the last 5 days this decreases 
to 4.3%). Since, in the election, only 1.53% voted for a third-party, we can assume 
that there were a large number of undecided voters in the polls. So, we must solve 
this problem if we want to compare the polls to the election outcome (where there 
are no undecided voters) or to the IEM where vote-share prices reflect the two-party 
split on the day of the election.

In the 1948 presidential election, polls failed to predict the re-election of Truman, 
it was suggested that this was due to the large number of undecided voters. This 
led to the Mosteller Commission (Mosteller et al., 1949) which published a book 
in which different ways of measuring the accuracy of polls that include undecided 
voters were proposed. One of these alternative methods could be used to account 
for the undecided voters but this will not be looked at further in this paper.

**Projection error**

A last criticism is that polls suffer from projection error (as discussed previously 
in Section 2.5.4). The prediction market, by definition, tries to predict what the 
outcome of the election will be on the day of the election. However, polls ask 
participants how they would vote if the election were held on the day that the 
poll is administered. This means that a poll that is held a long time before the 
election is unlikely to be accurate since the opinions of the individual voters will 
still significantly change by the time of the election. Closer to the election we 
expect the projection error to be less and polls to be more accurate.

We can solve this problem and compare the polls and the market as equals by 
built a model for the polls, this will be discussed in detail in Section 2.5.7.

### 2.5.6 Expansions

To address the criticism that our results are in favor of the markets only because we 
are combining data from both ‘good’ and ‘bad’ polls, we replace our polling data
Figure 13: Expansions: comparing the IEM vote-share predictions to other forecasts from individual spot-polls with an index of polls.\footnote{All the data for this section were collected from the PollyVote data set at: \url{http://spreadsheets.google.com/pub?key=pr1ZdfEZ874nj4KRoxUTUXA}} For this we collected data from the ‘RealClearPolitics Poll Average’ (further abbreviated as RCP) that combines polling data from various organizations into an average poll which we can then compare to our market data using the same methodology as used before.\footnote{For more information about the RCP poll average see: \url{http://www.realclearpolitics.com/}} Since this is an average, we have one observation for each day. The results are shown in panel A of Table VIII and can also be seen graphically in panel a of Figure 13. We see that overall the market wins in 55\% of the cases against the RCP (significant at 95\%) compared to 66\% of the cases in the previous section. It would be fair to say that one of the advantages for the market was that it was being compared to very variable polls, using the RCP instead has fixed this and the market’s advantage has
significantly decreased although it still performs slightly better.

Since using the RCP average instead of our individual polls had such a large effect on the result (the percentage of market wins decreased from 66% to 55%), we decided to also evaluate the market’s performance against other potential ways of predicting the outcome of the election. Firstly we test the market data against quantitative models. These are regression models designed by political scientists and economists to use data from past elections to forecast the result of the current election using current observable factors such as current economic conditions, public opinion and presidential approval ratings. There are many different quantitative models which each use different variables but for our purpose we use various models combined into one measure by PollyVote. The results are shown in panel B of Table VIII and can also be seen graphically in panel b of Figure 13. Here the models clearly beat the market, looking at all the observations the market is only closer to the outcome of the election 40% of the time. The market does seem to improve relative to the models the closer we get to the election but in the long-run the quantitative models seem to outperform the market.

We then decided to test the market against a survey of experts, this is also a continuous data set which is a combination of various surveys compiled by PollyVote.21 The results are shown in panel C of Table VIII and can also be seen graphically in panel c of Figure 13. We find that looking at all the observations, the market wins only 50% of the time; both perform equally well. Looking at the short-run before the election, the market performs much better than the experts (for example winning 100% of the time in the last 5 days), this is probably because the survey of experts is unable to incorporate new information quickly.

Lastly, we compare the market against the PollyVote, this is a combinational forecast that includes data from the Iowa Electronic Markets, polls, quantitative models and expert surveys. The authors argue that by combining these methods, the forecast error is reduced, leaving a more accurate forecast.22 The results are shown in panel D of Table VIII and can also be seen graphically in panel d of Figure 13. Here we see that for all the observations, the market wins 66% of the time (the same as against the polls). The market performs particularly badly in the 6 to 31 days range and without this the percentage would be higher. It is interesting to note that the market wins more often against the PollyVote than against any of its individual components.

21For more information about expert surveys, see Jones et al. (2007).
22For more details on the PollyVote, see Cuzán et al. (2005).
**Table VIII**: Comparing the IEM predictions to the RCP poll average, quantitative models, expert surveys and the PollyVote

<table>
<thead>
<tr>
<th>Days included in sample</th>
<th>Item</th>
<th>All Years</th>
<th>2008</th>
<th>(A)</th>
<th>(B)</th>
<th>(C)</th>
<th>(D)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of polls</td>
<td>964</td>
<td>357</td>
<td>427</td>
<td>429</td>
<td>429</td>
<td>429</td>
</tr>
<tr>
<td></td>
<td>Poll “wins”</td>
<td>255</td>
<td>122</td>
<td>192</td>
<td>258</td>
<td>213</td>
<td>148</td>
</tr>
<tr>
<td></td>
<td>Market “wins”</td>
<td>709</td>
<td>235</td>
<td>235</td>
<td>171</td>
<td>215</td>
<td>281</td>
</tr>
<tr>
<td></td>
<td>% market wins</td>
<td>74***</td>
<td>66***</td>
<td>55**</td>
<td>40</td>
<td>50</td>
<td>66***</td>
</tr>
<tr>
<td></td>
<td>p-value (1 sided)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.021</td>
<td>1.000</td>
<td>0.462</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Market “wins”</td>
<td>268</td>
<td>105</td>
<td>174</td>
<td>123</td>
<td>137</td>
<td>203</td>
</tr>
<tr>
<td></td>
<td>% market wins</td>
<td>74***</td>
<td>66***</td>
<td>53</td>
<td>37</td>
<td>42</td>
<td>62***</td>
</tr>
<tr>
<td></td>
<td>p-value (1 sided)</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.999</td>
<td>0.000</td>
</tr>
<tr>
<td>More than 100 days</td>
<td>Number of observations</td>
<td>360</td>
<td>160</td>
<td>327</td>
<td>329</td>
<td>329</td>
<td>329</td>
</tr>
<tr>
<td></td>
<td>Poll “wins”</td>
<td>92</td>
<td>55</td>
<td>153</td>
<td>206</td>
<td>192</td>
<td>126</td>
</tr>
<tr>
<td></td>
<td>Market “wins”</td>
<td>268</td>
<td>105</td>
<td>174</td>
<td>123</td>
<td>137</td>
<td>203</td>
</tr>
<tr>
<td></td>
<td>Market “wins”</td>
<td>100</td>
<td>47</td>
<td>29</td>
<td>20</td>
<td>29</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>% market wins</td>
<td>84***</td>
<td>60**</td>
<td>29</td>
<td>60</td>
<td>83***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>p-value (1 sided)</td>
<td>0.000</td>
<td>0.075</td>
<td>0.020</td>
<td>0.000</td>
<td>0.997</td>
<td>0.155</td>
</tr>
<tr>
<td>66 - 100 days</td>
<td>Number of observations</td>
<td>131</td>
<td>28</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Poll “wins”</td>
<td>21</td>
<td>14</td>
<td>11</td>
<td>25</td>
<td>14</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Market “wins”</td>
<td>110</td>
<td>14</td>
<td>24</td>
<td>10</td>
<td>21</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>% market wins</td>
<td>84***</td>
<td>60**</td>
<td>29</td>
<td>60</td>
<td>83***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>p-value (1 sided)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.196</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>32 - 65 days</td>
<td>Number of observations</td>
<td>173</td>
<td>64</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Poll “wins”</td>
<td>56</td>
<td>17</td>
<td>5</td>
<td>14</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Market “wins”</td>
<td>117</td>
<td>47</td>
<td>29</td>
<td>20</td>
<td>29</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>% market wins</td>
<td>68***</td>
<td>73**</td>
<td>59</td>
<td>59</td>
<td>85***</td>
<td>94***</td>
</tr>
<tr>
<td></td>
<td>p-value (1 sided)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.423</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>6 - 31 days</td>
<td>Number of observations</td>
<td>222</td>
<td>79</td>
<td>26</td>
<td>26</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>Poll “wins”</td>
<td>61</td>
<td>28</td>
<td>19</td>
<td>12</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Market “wins”</td>
<td>161</td>
<td>51</td>
<td>7</td>
<td>14</td>
<td>24</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>% market wins</td>
<td>73***</td>
<td>65**</td>
<td>27</td>
<td>54</td>
<td>92***</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>p-value (1 sided)</td>
<td>0.000</td>
<td>0.006</td>
<td>0.095</td>
<td>0.423</td>
<td>0.000</td>
<td>0.721</td>
</tr>
<tr>
<td>Last 5 days</td>
<td>Number of observations</td>
<td>78</td>
<td>26</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Poll “wins”</td>
<td>25</td>
<td>8</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Market “wins”</td>
<td>53</td>
<td>18</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>% market wins</td>
<td>68***</td>
<td>69**</td>
<td>20</td>
<td>80</td>
<td>100**</td>
<td>100**</td>
</tr>
<tr>
<td></td>
<td>p-value (1 sided)</td>
<td>0.001</td>
<td>0.038</td>
<td>0.969</td>
<td>0.188</td>
<td>0.031</td>
<td>0.031</td>
</tr>
</tbody>
</table>

*** p < 0.001, ** p < 0.05, * p < 0.1
2.5.7 Adjusting the polls

Our last criticism was that polls suffer from projection error, meaning that polls that are taken long before the election are not good forecasts for the results on the day of the election. To solve this, we have to adjust the polls so that they can be interpreted as forecasts for what will happen on the day of the election. Erikson and Wlezien (2008) argue that it is not fair to compare the predictions of a prediction market directly to polls since the market tries to predict what will happen on the day of the election while the polls collect the opinions of those surveyed as if the election were held on that day. The direct comparison as performed in the previous section and by Berg, Nelson and Rietz (2008), is thus unfair since polls and market prices are not directly comparable. To solve this problem, Erikson and Wlezien transform raw poll data into predictions for election day and compare these to the vote-share prices of the Iowa Electronic Markets (IEM) for U.S. presidential elections from 1988 to 2004. They find that when the daily projected vote from the adjusted polls is used rather than the daily raw poll predictions, the market’s advantage for forecasting elections vanishes. We will test if this is also true for the 2008 data.

Methodology

Since prediction market contract prices are forecasts of what will happen on the day of the election but polls ask the question of how a person would vote if the election were held that day, they are not directly comparable. To take this into account, polling numbers need to be transformed into projections for the election day outcome. To do this we must understand how predictions from a poll, held on a certain day before the election, translate into an expected vote-share on the day of the election. To do this we build an adjustment model using data from how polls in previous elections translated into an election day result.

For this, Erikson and Wlezien (2008) use a data set of polling data from U.S presidential elections from 1952 to 2004. For every day leading up to every presidential election since 1952 (going up to 200 days before) the two-party vote-share for the incumbent party (the party that won the previous election) of each poll is recorded. With this time series of a vote-share for each day, they regress the actual vote margin for the incumbent (the outcome of the election) on the latest polls. The outcome of these 200 equations allows for the creation vote projections out of the polls which can then be compared to the market predictions. For each election they use only the data that would have been available at that time. For example, the regression
to make projections for the 1960 election would use only polling data from 1952 and 1956 to build the adjustment model. So, in order to build a similar adjustment model for the 2008 election, polling data are required for all the presidential elections from 1952 to 2004.

For 2008, we use the polling data collected for the previous sections. Following their methodology we turn this into a time series so we have only one observation for each day, up to 200 days before the election. This can also partly solve the problem of variability mentioned earlier. If there was more than one poll ending on certain day, we pool the available polls for that day into one poll-of-polls. If there was no poll ending on a day we simply take the poll result from the most recent poll for that day. Following their methodology, we also lag the polls two days, effectively treating the poll as if it was released two days after its last day in the field, this also puts polls and market prices on a more equal footing.\textsuperscript{23} In the end, this leaves us with one polling observation for the incumbent party’s vote-share for each day up to 200 days before the election (-1 to -200) we also have the result of the election for the incumbent party. We opted to use only the observations going back 200 days before the election and using data starting from 1952 (following \textit{Erikson and Wlezien}).

We also collected data for the adjustment model. The polling data and election outcomes from 1952 to 2000 come from \textit{Wlezien and Erikson} (2002).\textsuperscript{24} The data was for the Democratic share of the two-party vote so we rewrote this so that we had data for the incumbent party instead. On days where there was no poll ending, we used the observation of the most recent poll. The data for 2004 were collected from PollingReport.com. This gave us a trial-heat poll result for the incumbent party and the outcome of the election for the incumbent party for every day up to 200 days before the election for every election from 1952 to 2008.

To build the adjustment model to transform the raw trial heat polls into vote projections, we use the following regression equation to regress the actual vote-share outcome of the election on the polls for each day before the election (-1 to -200):

\begin{equation}
V_y = \alpha_t + \beta_t P_{yt} + \epsilon_{yt}
\end{equation}

where, $V_y$ is the actual incumbent-party percent of the two-party vote in year $y$ minus 50 and $P_{yt}$ is the corresponding trial heat poll division minus 50 in year $y$.

\textsuperscript{23}To test for robustness, we also repeated our analysis without lagging the polls and found that lagging the polls by two days had no significant effect on our results.

\textsuperscript{24}This is an older article than \textit{Erikson and Wlezien} (2008) but uses similar data. The data can be found at: \url{http://webapp.icpsr.umich.edu/cocoon/ICPSR-STUDY/01304.xml}
on day $t$ of the campaign. The vote-share for the incumbent is thus measured as a deviation from 50%. Separate equations were made for each day up to 200 days before the election using the data for 1952 to 2004. The alphas ($\alpha_t$, the intercepts) and betas ($\beta_t$, the coefficients) of these regressions can be seen in Figure 14; these are very similar to those that can be seen in Figure 2 of [Erikson and Wlezien (2008, p.199)] which only uses data from 1952 to 2000. We also report the root mean squared error of the regression.

From this, we find the same results as [Erikson and Wlezien (2008)]. Our regression coefficient ($\beta_t$) starts just above 0.3 for 200 days before the election and ends just below 0.8 one day before the election. This shows that polls gain weight (make better predictions) closer to the election.

The intercept ($\alpha_t$) is generally positive until just before the election and is highest
around 100 days before the election, during the summer. Erikson and Wlezien claim that this stems from the fact that polls tend to underestimate the support for the incumbent party and so the challenging party receives a higher than justified prediction which will not remain closer to the election.

The root mean squared error shows us the error that we make when predicting the election outcome using the data from the polls, this decreases as we get closer to the election showing that the polls become more accurate the closer we get to the election.

Using these coefficients, we can compute the projected vote on election day that the poll predicts, for each day for the 2008 election ($Y = 2008$):

$$\hat{V}_{2008} = \alpha_T + \beta_T P_{2008,T}$$ (2.7)

Where $\hat{V}_{2008}$ is the forecast for the two party vote-share for the incumbent party (minus 50) based on the polls for year 2008 for the actual outcome of the election $V_{2008}$ in 2008. Using our intercepts for each time $T$ ($\alpha_T$), coefficients for each time $T$ ($\beta_T$) and the raw poll result at time $T$ ($P_{2008,T}$).

If $\alpha$ were zero and $\beta$ one, the raw polls could be directly interpreted as forecasts for the results on the day of the election but, as Figure 14 shows, this is not the case. Once polling data for the year in question have been adjusted using this method, it can be compared directly to prediction market contract prices using the same methods as in the previous sections.

Results

When the polling data are not adjusted, Erikson and Wlezien (2008) reach a similar conclusion as Berg, Nelson and Rietz (2008) and as we find for 2008 in Section 2.5.4 when polls are directly compared to market forecasts the market’s predictions are significantly more accurate. However, when the polling data are adjusted using the above methodology, they find that the market’s advantage in predicting the outcome of the election vanishes. Table IX shows these outcomes, showing for each year the average absolute prediction error of raw polls and market forecasts (for example, row 1988) and then the results for the same year when the adjustment model is used to place polls and the market on equal footing (for example, row 1988(a)). Looking at all the years in their sample (1988-2004), Erikson and Wlezien find that the market has an average absolute prediction error (AAPE) of 2.41 while the raw polls have a, much larger, AAPE of 4.46. They find, when comparing the predictions of the
raw polls to those of the market, the market is closer to the outcome on 75% of the days. However, once the polls are adjusted so that they can be interpreted as predictions for the outcome of the election (as described in the previous section) the average absolute prediction error of the polls decreases to 2.13. This is less than that of the market and the market then only wins on 45% of the days. Similar results were found when looking at other years; adjusting the polls significantly improves their predictive value relative to the market. They conclude that when comparing raw polls directly to a prediction market the market tends to win but if one understands how polls work and how polls do not predict the outcome on the day of the election, using this knowledge to interpret the polls, the market loses its advantage and polls are more accurate.

We follow the same methodology for our data from 2008 and find completely different  

---

<table>
<thead>
<tr>
<th>Election Year</th>
<th>Average Absolute Prediction Error: IEM</th>
<th>Average Absolute Prediction Error: Polls</th>
<th>IEM AAPE as Proportion of Poll AAPE</th>
<th>Percentage of Days Market Beats Polls</th>
</tr>
</thead>
<tbody>
<tr>
<td>1988</td>
<td>3.93</td>
<td>5.55</td>
<td>0.71</td>
<td>60%</td>
</tr>
<tr>
<td>1988(a)</td>
<td>3.93</td>
<td>2.86</td>
<td>1.34</td>
<td>28%</td>
</tr>
<tr>
<td>1992</td>
<td>5.68</td>
<td>7.06</td>
<td>0.80</td>
<td>67%</td>
</tr>
<tr>
<td>1992(a)</td>
<td>5.68</td>
<td>4.21</td>
<td>1.35</td>
<td>25%</td>
</tr>
<tr>
<td>1996</td>
<td>1.01</td>
<td>5.32</td>
<td>0.19</td>
<td>92%</td>
</tr>
<tr>
<td>1996(a)</td>
<td>1.01</td>
<td>1.76</td>
<td>0.57</td>
<td>64%</td>
</tr>
<tr>
<td>2000</td>
<td>0.84</td>
<td>3.08</td>
<td>0.27</td>
<td>88%</td>
</tr>
<tr>
<td>2000(a)</td>
<td>0.84</td>
<td>1.16</td>
<td>0.72</td>
<td>62%</td>
</tr>
<tr>
<td>2004</td>
<td>0.89</td>
<td>1.52</td>
<td>0.59</td>
<td>65%</td>
</tr>
<tr>
<td>2004(a)</td>
<td>0.89</td>
<td>0.79</td>
<td>1.13</td>
<td>41%</td>
</tr>
<tr>
<td>All Years</td>
<td>2.41</td>
<td>4.46</td>
<td>0.47</td>
<td>75%</td>
</tr>
<tr>
<td>All Years(a)</td>
<td>2.41</td>
<td>2.13</td>
<td>1.13</td>
<td>45%</td>
</tr>
<tr>
<td>2008</td>
<td>1.73</td>
<td>1.76</td>
<td>0.98</td>
<td>56%</td>
</tr>
<tr>
<td>2008(a)</td>
<td>1.73</td>
<td>3.89</td>
<td>0.45</td>
<td>89%</td>
</tr>
</tbody>
</table>

Table IX: Our results from 2008 and those of [Erikson and Wlezien (2008)] from 1988 to 2004 and All Years (Table 1, p.198 and Table 2, p.201). For each year the market prices are compared to the raw polls first and then to the polls which have been transformed into projections using the adjustment model (a).
results. Comparing the raw polls to the IEM market prices, we find that the market wins in 111 out of the 198 cases or 56% of the time. The market has an average absolute prediction error of 1.73 while the polls have a slightly higher error of 1.76. However, once we adjust the polls using the above methodology, the market is closer to the outcome for 177 out of 198 observations or 89% of the time. The AAPE of the polls increases from 1.76 to 3.89. Adjusting the polls has clearly decreased their predictive accuracy, interpreting the raw polls as predictions would have given more accurate predictions than adjusting the polls. Possible explanations are suggested in the next section.

To also allow us to compare the results of this new methodology to the results achieved in Section 2.5.4 and by Berg, Nelson and Rietz (2008) we also use our adjustment model to adjust the raw spot polling data instead of making a time series with one observation on each day like Erikson and Wlezien. The results of this can be seen in Column B of Table VII. The same results are found; when the polls are adjusted using the methodology of Erikson and Wlezien (2008) the percentage of observations where the market’s prediction is closer to the outcome than the polls increases from 66% to 85%.

Discussion

We have found the opposite results of Erikson and Wlezien (2008), the adjustment model which improved the predictions of polls in all five elections from 1988 to 2004, worsened them in 2008. Looking at Figure 15 offers a possible explanation for this. We see the predicted incumbent vote-share of both the raw poll and the adjusted poll as well as the outcome of the election (the horizontal line). As can be seen, the raw poll predictions are relatively close to the outcome but the adjustment of the polls causes the line to be transformed upwards (increase the predictions for the incumbent party) and away from result of the election (that the incumbent party lost).

To test for robustness, Figure 16 shows the same using the RCP poll average for 2008 instead of our individual poll data and finds that the adjustment model has the same effect. A possible explanation for this is the characteristics of the election.

---

26 This is more comparable to our comparison to the RCP poll average than our comparison to all the polls due to the problem of variability described earlier.

27 There are less observations used for the calculations in column B than column A of Table VII, this is because we only look at spot polls carried out up to 200 days before the election since our adjustment model only goes this far.
As mentioned earlier, we can see from the intercepts of the regression equations that the polls historically tend to undervalue the support of the incumbent party and overvalue the support for the challenging party. If this is the case, increasing the prediction for the incumbent party (as the adjustment model does) will provide better predictions. However, this is only true if the incumbent is really undervalued. As we see in the figures, in 2008 this was not the case, the incumbent party received only 46.3% of the two-party vote-share which means that increasing the prediction for the incumbent will only move it away from the outcome and thus have negative effects on its predictive value.

To study this further, we use our same regression intercepts and coefficients to adjust the polling data from 1988 to 2008, these are shown in Figure 17. We see that in all the years 1988 to 2004, the adjusting of the polls had the desired effect on the prediction; moving it closer to the outcome (the horizontal line). In 1988, for example, the raw polls predicted the incumbent’s chances as too low so the adjusting model increased the predictions which was correct in the end. In 1996, the raw polls

---

28 It is important to note that these were all calculated using the intercepts and coefficients for 2008, so using polling data from 1952 to 2004. As mentioned before, to adjust their data, Erikson and Wlezien only use the data available at the time of the election. The purpose of these figures is thus only to show what effect the adjusting model has on the raw polls in different election situations and not to draw any other conclusions.
Figure 16: 2008 RCP poll average: raw poll and adjusted poll predicted the chances of the incumbent too high and the adjusting model decreases the predictions of the polls which was also correct in the end. The raw polls from 2000 look rather similar to those of 2008, between 200 and 100 days before the election they are between 45% and 50% and in both cases the adjusting model had the same effect; increasing the prediction for the incumbent party upwards. In 2000 this improved the predictions since the outcome for the incumbent was 50.25% while in 2008 this made the predictions worse since the outcome was 46.3%.

Conclusions

From this we can draw the conclusion that the results of Erikson and Wlezien (2008) are not robust. Their adjusting model improves the forecasts of polls relative to the market for the five observations in their data set but do not for the election of 2008. The characteristics of the elections from 1988 to 2004 allowed their adjusting model to work well while the different characteristics (in particular the incumbent being predicted significantly less than 50% early on and also receiving significantly less than 50% of the vote) do not allow their conclusions to hold for 2008. A different adjusting model may have led to different conclusions but, of course, the purpose of polls is to forecast and you cannot know the characteristics of an election before it is completed.
2.6 Conclusion: policy implications

What we have shown in this part of the paper is that prediction markets can potentially provide more accurate predictions of election vote-shares than polls.

Looking at our analysis for 2008, we find that the market significantly outperforms all polls and slightly outperforms polls when compared to a poll average. Building an adjustment model to adjust the polls also does not improve their forecasts for the 2008 election. A possible explanation is that the polls performed particularly badly in the 2008 election due to the characteristics of the election; large uncertainties about candidates, large numbers of undecided voters and even the ‘Bradley effect’ (where voters tell pollsters that they are undecided or would vote for the black candidate for fear of being criticized as racist while on the day of the election voting for the white candidate) have been given as explanations for badly performing polls in 2008. The adjustment model for 2008 fails because the incumbent performed as badly as predicted early on, unlike in previous elections.

Looking at the results presented from other years (1988 to 2004) we find that the market tends to greatly outperform raw polls. If we assume for a moment that the findings of Erikson and Wlezien (2008) were correct, they find that the adjusted polls outperform the market predictions, but this does not mean that the markets are useless. Looking at all five elections, 1988 to 2004, they find that the market still beats polls 45% of the time which is still very close and using their limited sample it is too soon to say one outperforms the other. We could even conclude from this that the markets work exceptionally well; instead of building an adjustment model using 200 regression equations and data from all the elections since 1952, we can just look at the market price and see a prediction that is better 45% of the time. The market accurately interprets the information in poll reports for us. Maybe there is a way to adjust prediction market forecasts to compensate for biases, to provide an even more accurate forecast, this has yet to be researched.

Does this mean that we can stop running costly polls and rely on the markets for our predictions? Not necessarily; traders on prediction markets incorporate information from polls, we can assume that the information from polling organizations allows them to set prices more accurately. However, Rhode and Strumpf (2004) find that markets made accurate predictions before the introduction of large-scale polls. Also, Erikson and Wlezien (2009) find that market prices were far better predictors before the introduction of polls than after polls became available, implying that the introduction of polling made the predictions of the markets worse.
Another important question is whether our findings for the U.S. presidential elections also be applied to prediction markets to forecast political elections in other countries. In general, our answer is yes; a prediction market can be used to forecast any election, but the more complicated the political system, the less accurate we would expect the results to be. Jacobsen et al. (2000) report on prediction markets in Europe where there are many equally sized parties and find that predictions are on average less accurate than those in the U.S. where there are only two main parties. They find that price biases tend to be more pronounced in systems of proportional representation.
Figure 17: Raw polls and adjusted polls from presidential elections 1988 to 2008
General Conclusion

This paper has been an extensive study into prediction markets both in theory and in practice. The first part of this paper showed that prediction markets have been successful in predicting a large range of different events and discussed the design of prediction markets in great detail. Looking at the types of contracts on prediction markets (winner-take-all, index, spread) we showed how each can be used to predict a different aspect of an event and the choice thus depends on what the designer wants to predict. We looked at possible trading mechanisms for prediction markets and distinguished four: the continuous double auction, the call auction, dynamic pari-mutuel markets and market scoring rules. By comparing them on whether they allow for the continuous incorporation of information, whether they guarantee liquidity and whether there is risk for the market operator in using this mechanism, we found that each has its own advantages and disadvantages. Most often we see continuous double auctions used for their simplicity but there can be liquidity problems while market scoring rules are the only mechanisms that guarantee liquidity on both sides of the market but there is a (bounded) cost for the operator. One of the most important design decisions is whether a market will use real-money or play-money. The traditional belief is that the use of real-money encourages traders to make more accurate predictions. However, by looking at evidence from experimental economics and evidence from various prediction markets, we found that this is not necessarily the case. Many studies find that real- and play-money markets can perform equally well in many situations. Due to the added legal difficulties and costs of setting up a real-money market, any potential gains in predictive accuracy have to be weighed against the added costs and in many cases it would be recommended to use play-money.

We then discussed the population of traders and found that large numbers of traders are not necessarily needed (as long as the market is liquid) since even a small group can make accurate predictions if they collectively possess relevant information about the subject. Other considerations when designing a prediction market include the
transaction costs, trading limits, the information available to traders and their anonymity. We also mentioned limiting trading hours, the competition between trading venues and the trading interface of markets.

Then, we tried to answer some often asked questions that challenge the performance of prediction markets. Firstly, a requirement for a market to make accurate predictions is that there are no long-term arbitrage opportunities. We found evidence that not many arbitrage opportunities arise in large markets but also point to significant opportunities in experimental markets. Secondly, we asked if traders on prediction markets are rational and found a large literature that claims that average traders are prone to make mistakes and are biased by their preferences in their valuation of contracts. It was also mentioned that both the favorite-longshot bias and speculative bubbles have been observed on prediction markets. Despite these shortcomings, prediction markets seem to be capable of making accurate predictions. The explanation offered for this is that, while the average traders suffer from these shortcomings, they are not the ones who set the prices. Prices are set by the marginal traders who are more rational and less mistake-prone than the average traders.

Another important question is whether manipulation of prediction markets is possible. Looking at evidence on manipulation from both experimental and real-world markets, we conclude that influencing the price for a short time is possible but that this has little effect on the long-run price level and can be very costly for the manipulator. Lastly, we addressed the criticisms that the prices of prediction markets cannot be interpreted as the belief of the mean trader. While this does hold in some situations, more realistic models of predication markets show that the market price is very close to the mean belief of the traders.

The second part of this paper looked at a case of how prediction markets functioned in practice, in the run-up to the 2008 U.S. presidential election. By looking at data collected from eight online prediction markets with very different designs, we confirmed many of our findings from the first part of the paper.

Looking at the data from Obama and McCain winner-take-all contracts on Intrade, we showed how the prices are driven by news and adapt quickly when new information becomes available. However, we discovered that it is difficult (if not impossible) to judge the winner-take-all data for accuracy against the final election outcome. Since it is not possible to define what constitutes an accurate prediction, we used the data instead to test other hypotheses. Firstly we compared winner-take-all contracts for Obama and McCain across different markets and found a very high correlation coefficient between the two real-money markets and a much lower (but
still high) correlation between the two play-money markets, possibly due to less informed trades being made on the play-money markets. We also found that the predictions for the Obama contract on the play-money markets tend to be higher than those on the real-money markets. We suggest an explanation for this; that traders on play-money markets could be more biased by their preferences and thus overvalue Obama’s chances of winning. Real-money markets should be plagued less by biased prices since they provide profit opportunities for other traders.

We then compared winner-take-all contracts for individual candidates to parties using data from Intrade and found that once a candidate is nominated, the contract prices of the candidate and their party are nearly identical and move together. Comparing winner-take-all contracts for parties across markets we found that the Democratic and Republican contracts also move together well on different markets but the correlation coefficients were lower than for the Obama and McCain contracts on different markets. Here, we again found that a play-money market was closest to the outcome on the most days, we also explain this by biases being more predominant on play-money markets.

Other contracts related to the election were then discussed. Intrade contracts were used to forecast the number of electoral votes each party would receive. Looking at the data, we found that they were surprisingly inaccurate, possibly due to the characteristics of the election or of the markets. Looking at Intrade contract data for individual states we found that the market performed well, predicting all but one state correctly on the eve of the election and also predicting 49 out of 51 correctly one month before the election.

Since winner-take-all contracts cannot be judged for accuracy against the outcome of the election, we then looked at vote-share contracts from the Iowa Electronic Markets. We looked at the absolute accuracy of the vote-share prices but found that we needed something to compare them to, so it was decided to compare them to the traditional way of predicting election outcomes: polls.

To compare the market’s predictions to those of polls, we followed the methodology of Berg, Nelson and Rietz (2008). After collecting polling data, we adjusted the polls and market prices so that they represented predictions for the two-party normalized spread. Then we compared, for each observation, the prediction of the poll and the prediction of the market to the outcome of the election to see which was more accurate. Confirming the results of Berg, Nelson and Rietz (2008), we found that the market was closer to the outcome than the poll in 66% of the observations and that the market is closer in all, except one, of the time periods we looked at.
This confirms the evidence for the accuracy of prediction markets, however, we discuss three criticisms of these results. Firstly, the fact that we are taking many polls together may be favorable for the markets. Looking at only the best polling company could improve the predictions of the polls. To test this, we replaced our spot-poll predictions with the RealClearPolitics poll average and found that the market then only wins on 55% of the observations, this criticism was probably justified, the market still performs better but by significantly less. We also compared the market to other potential ways of predicting the outcome of the election and found that the market outperforms the PollyVote, is beaten by quantitative models and is tied with surveys of experts.

The second criticism was that a lot of the inaccuracy of the polls could stem from the fact that they have a large number of undecided voters. The use of other methods to account for undecided voters was suggested but this was not explored further. The last criticism was that polls and market prices could not be directly compared since market prices attempt to forecast what will happen on the day of the election while polls ask how people would vote if the election were held on the day of the poll. Because of this, polls suffer from projection error which increases the further away we are from the day of the election. To account for this, we used the methodology of Erikson and Wlezien (2008) to build an adjustment model for the polls. Using polling data from 1952 to 2004 we ran regressions to see how poll predictions on each day before the election in previous elections translated into election day outcomes and then used this to adjust the 2008 polling data. Erikson and Wlezien (2008) found that when the polls were adjusted, the market’s advantage disappeared. However, we found the opposite; the adjusted polls were significantly less accurate than the raw polls for the 2008 election. We suggest that this was due to the characteristics of the 2008 election and conclude that Erikson and Wlezien's results are not robust since they are only valid if the election meets certain conditions, which occurred from 1988 to 2004 but not in 2008.

To conclude, nearly all our results are in favor of the use of prediction markets to forecast election outcomes. Prediction markets also have many more potential applications and we expect them to perform equally well in other situations. To those looking for an accurate forecast for any future event we would fully recommend them to consider the use of prediction markets.
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Intrade winner-take-all contract daily closing prices for Obama and McCain for the whole time range of the contracts (24/10/2006 to 06/11/2008)</td>
</tr>
<tr>
<td>2</td>
<td>Intrade winner-take-all contract daily closing prices for Obama and McCain from the beginning of 2008 until the election</td>
</tr>
<tr>
<td>3</td>
<td>Intrade, Betfair, Hubdub and Inkling Markets contract prices for Obama and McCain from the beginning of 2008 until the election</td>
</tr>
<tr>
<td>4</td>
<td>Intrade contract prices for Obama and McCain and for the Democrats and Republicans for the whole time range of the contracts</td>
</tr>
<tr>
<td>5</td>
<td>Intrade and IEM contract prices for the Democratic Party and the Republican Party for the whole time range of the contracts</td>
</tr>
<tr>
<td>6</td>
<td>Intrade, the IEM, Bet2Give, NewsFutures and the Foresight Exchange contract prices for the Democratic Party from begin 2008 until the election</td>
</tr>
<tr>
<td>7</td>
<td>Probability distribution of the number of Democratic electoral votes one month before the election (on 04/10/2008) from Intrade data</td>
</tr>
<tr>
<td>8</td>
<td>Predicted Democratic electoral votes from 08/06/2008 to 07/11/2008</td>
</tr>
<tr>
<td>9</td>
<td>Maps showing Intrade predictions for winner of the majority of the electoral college votes per state</td>
</tr>
<tr>
<td>10</td>
<td>IEM vote-share markets daily closing price and traded units</td>
</tr>
<tr>
<td>11</td>
<td>All the polling data: predicted Democratic winning margin over time</td>
</tr>
<tr>
<td>12</td>
<td>Predicted Democratic vote-share winning margins for the market and selected polls from begin 2008 until the election</td>
</tr>
<tr>
<td>13</td>
<td>Expansions: comparing the IEM vote-share predictions to other forecasts</td>
</tr>
</tbody>
</table>
14 Intercept, coefficient and root mean squared error of the 200 regression equations using data from 1952 to 2004 85
15 Our 2008 polling data: raw poll and adjusted poll 89
16 2008 RCP poll average: raw poll and adjusted poll 90
17 Raw polls and adjusted polls from presidential elections 1988 to 2008 93
### List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Comparison of various online prediction markets</td>
<td>10</td>
</tr>
<tr>
<td>II</td>
<td>Comparison of various trading mechanisms (adapted from Luckner (2008, p.239))</td>
<td>16</td>
</tr>
<tr>
<td>III</td>
<td>Correlations between Obama winner-take-all contracts on various markets</td>
<td>53</td>
</tr>
<tr>
<td>IV</td>
<td>Binomial tests: percentage of days where the row market was closer to the outcome than the column market for Obama contracts from 29/08/2008 until the election</td>
<td>54</td>
</tr>
<tr>
<td>V</td>
<td>Correlations between Democratic winner-take-all contracts on various markets from the beginning of 2008 until the election</td>
<td>59</td>
</tr>
<tr>
<td>VI</td>
<td>Binomial tests: percentage of days where the row market was closer to the outcome than the column market for Democratic contracts</td>
<td>59</td>
</tr>
<tr>
<td>VII</td>
<td>Comparing the IEM vote-share predictions to polls</td>
<td>77</td>
</tr>
<tr>
<td>VIII</td>
<td>Comparing the IEM predictions to the RCP poll average, quantitative models, expert surveys and the PollyVote</td>
<td>82</td>
</tr>
<tr>
<td>IX</td>
<td>Our results from 2008 and those of Erikson and Wlezien (2008) from 1988 to 2004 and All Years (Table 1, p.198 and Table 2, p.201). For each year the market prices are compared to the raw polls first and then to the polls which have been transformed into projections using the adjustment model (a).</td>
<td>87</td>
</tr>
</tbody>
</table>
Bibliography


**URL:**  http://hanson.gmu.edu/futarchy.html [20/07/2008]


URL: [http://bpp.wharton.upenn.edu/jwolfers/Papers/Favorite_Longshot_Bias.pdf](http://bpp.wharton.upenn.edu/jwolfers/Papers/Favorite_Longshot_Bias.pdf)


