Decision Market Predictions: Separating Efficiency and Bias*

June 1, 2005

Glenn Bean
Stanford University
Department of Economics

ABSTRACT

Information markets provide a revolutionary mechanism for revealing aggregate knowledge. They continue to gain popularity in academic circles and continue to prove themselves through successful prediction for practical applications. Policy makers have shown to be wary of depending upon their results exclusively for decision-making, but there is evidence that information markets could be used for relevant advising. One particular type of information market uses contingent contracts that speculate about unrealized states of the world. These estimates hold an inherent bias. This study develops a model for a more efficient estimator and how to acquire it through conventional contracts.

* I would like to thank Eric Zitzewitz for all of the suggestions, guidance and resources he provided throughout the project. I would also like to thank Sinead Kelliher www.tradesports.com for generous help with the data. Glenn Bean: Undergraduate, 2005, Stanford University; gbean@stanford.edu
Introduction

People, businesses and institutions have always been interested in gathering information to create a glimpse of the unknown because accurately forecasting the future improves the efficiency of their respective activities. Sometimes this takes the form of experimentation, development of analytical theories or simply first-hand experience. With the advent of the Internet and the wealth of information it makes available, it follows naturally that more clever and advanced methods are being pursued to gather, analyze and understand information. Free markets of all types are very good at aggregating information which is disclosed through prices. A recently developed market-based instrument has been shown to have remarkable precision when asking directed, meaningful questions. These “information markets” (also known as “prediction markets” or “event futures”) originated when one was established by a group of economists at the University of Iowa to predict the outcome of the 1988 American presidential election. A prediction market is a forum for exchanging contracts whose payoffs are tied to unknown future events. After becoming overwhelmingly popular in academic circles, the markets continue to operate and have permeated many sectors of society with practical applications.

Examples of Information Markets in the Past

Academics and industry were quick to find applications for information markets after their splash on the political scene in 1988. The markets have been used for product research in the business world by firms such as Microsoft, Eli Lilly, Goldman Sachs, Deutsche Bank and Hewlett-Packard. In 1996, HP and Charles Plott, an economist from the California Institute of Technology, set out on a joint research project whose aim was to set up a software trading platform for a prediction market at HP. The participants selected for the research project
consisted of 30 product and finance managers from HP. They were each given $50 in a trading account and allowed to sell and purchase contracts on the levels they estimated quarterly sales would reach. For instance, if a manager thought sales would be in the range of $201 million and $210 million, he could buy a contract that would pay at the end of the quarter if his prediction was correct. If he revised his personal estimate throughout the quarter, he would likely try to sell the first contract and buy another based on the new estimate. These markets were open during lunch breaks and after business hours, and lasted for a week. The managers were allowed to keep profits earned from owning paying contracts (correct estimates) at the end of the market. When trading stopped, the contract with the highest price (i.e., under the highest demand from the managers) was deemed the “estimate” of the market. The HP marketing manager estimated simultaneously and independently estimated quarterly sales, as usual. The marketing manager’s projections missed by 13% during the experiment while the estimate of the information market had an error of only 6%. In further trials, the market estimate outperformed official forecasts 75% of the time.

Notorious Flash in the Pan

The United States government attempted to put information markets to use. In 2001, the Defense Advanced Research Projects Agency (DARPA), a research think tank within the Department of Defense, began funding a project called FutureMAP (later under the name Policy Analysis Market, or PAM). Professors who had completed seminal research in the field were contracted to establish information markets to allow for trading in various forms of geopolitical risk. The contracts to be traded ranged from indices of economic health, civil stability, military disposition, conflict indicators and even specific events. The final plan was to cover eight nations with five parameters for each, such as political instability and economic growth. Traders
would also predict US GDP, world trade and several other miscellaneous topics. (Hanson, 2004b). Examples of PAM\(^1\) contracts would be “How fast will the non-oil output of Egypt grow next year?” or “Will the Iraqi regime persist for more than one month after hostilities?” In addition to these contracts written for a single index or event, the exchange would have offered contract combinations. One logical combination would be of an economic event and a political event, such as the two examples above. The goal was to see how well trading by experts in this market could predict future events, and how the correlation of some of these events was perceived. Unfortunately, in July 2003, a few months before the first trades would be made in the market, press reports began to surface after two senators claimed that the market would allow people to bet on terrorist attacks, and critics quickly ripped DARPA for proposing “terrorism futures.” Senators Ron Wyden (D, OR) and Byron Dorgan (D, ND) complained that the U.S. Defense Department was planning to let people bet on terrorist attacks, or even worse, easily profit from terrorist activities. DARPA’s public relations director was out of town at the time, the PAM team was not contacted for comment, and so DARPA dropped the small program the very next day.

**Predicting Success**

While efforts to predict international futures have faltered, political prediction markets have gained momentous popularity, especially for Presidential elections. The Iowa Electronics Market (IEM) is a real-money futures market conducted by the University of Iowa Henry B. Tippie College of Business that trades election future contracts. Since the inaugural run with the 1988 Presidential election, the IEM has run markets for every presidential election, many congressional and state elections, and elections abroad. A study was conducted (Berg, Forsythe, ___

---

\(^1\) Robin Hanson has kept a detailed log of the development of the Policy Analysis Markets, the news stories that surrounded it, and its termination. This is maintained at [http://hanson.gmu.edu/policyanalysismarket.html](http://hanson.gmu.edu/policyanalysismarket.html).
Nelson and Rietz, 2003) comparing the accuracy of the IEM markets to that of the major polling sources (Gallup, CNN, and network news, among others). It was found that the markets outperformed the polls on average, and in 9 of 15 presidential elections they outperformed the polls significantly. The precision of these informal “electronic betting parlors” could be surprising to some. To address this, I ask the question, “What is the underlying mechanism, and why does it seem to perform so well?”

The Basics of a Prediction Market

In a world without time or resource constraints, anyone with a question for which a group of people could provide the answer would personally interview everyone that had a relevant insight. They would compile the results and then make a decision. Unfortunately, this luxury is not an option when the group size grows to be larger than just a handful of people. When gathering data from a large group, the demand for an efficient, cost-effective Information Aggregation Mechanism (IAM) arises. Prediction markets are one such mechanism.

A prediction market is a forum for exchanging contracts whose payoffs are tied to unknown future events. These contracts are a superset of the commonplace financial derivative called a “future.” A financial future would typically be a contract to purchase a commodity or security at a certain price in the future. These derivative markets, established primarily for hedging and locking in future prices and quantities, are not without predictive power. One study found that the future prices of orange juice concentrate were better indicators of the weather than forecasts by the National Weather Service of the Department of Commerce (Roll, 1984). In prediction markets, the contracts depend upon future events, and so can be written for a broader range of subjects. Depending on the rules linking the payoff to a future event, the market’s expectations of a range of different parameters can be elicited. While this type of analysis treats
the market as though it were itself a person with a set of expectations, it should be kept in mind that there are subtle but important distinctions between the market and the participants that make it up. The median expectation of the market is not the expectation of the median participant.

The typical mechanism of prediction markets is a continuous double-auction, with buyers submitting bids and sellers submitting asking prices, where a trade is executed any time the two sides of the market reach a mutually agreeable price. There are a few common types of contracts. Two with which this paper deal are “winner-takes-all” and an “index.” Another that is used less often is a “spread” contract.

First, in the “winner-take-all” model, a contract is sold for price $p$ that pays $1$ if and only if the specific event occurs, such as a particular candidate winning an election. The price varies given market demand, and can be interpreted as the market’s expectation that the event will occur. Second, an “index” contract is based on a number that rises or falls. It maps this outcome to a continuous payment function. For example, the percentage of the vote a candidate receives is a common index contract: after the election, it will pay the candidates vote share. The price for such a contract represents the mean expectation of the market. A third type of contract is the “spread.” The contract pays even money if an index is at or above a target. A contract for an election that costs $1$ would pay $2$ if a candidate’s vote share exceeded some value $x$.

Bidders bid on contracts with different $x$’s, according to their belief of the outcome of the event relative to $x$. This contract type reveals the median outcome of the event.

The information to be gleaned from a prediction market does not end with the novel sale of futures contracts on events. Each individual contract can reveal the market’s expectation of a specific parameter: probability, mean, etc., and combinations and families of contracts can be

---

2 This is assuming risk-neutrality. Concerns about interpreting prices as probabilities are raised by Manski (2004) and will be addressed later in the Literature Review.
useful to evaluate the uncertainty about these expectations. Consider a series of “winner-take-all” contracts that pay if and only if a candidate earns 50 percent of the vote, 51 percent and so on. This family of contracts would represent the entire probability distribution over all outcomes. Similarly, the payout of a “spread” contract could be varied to get a bigger picture of the median distribution. A contract that cost $2 and paid $3 if the realized outcome were greater than target would identify the 66th percentile of the distribution.

In addition to varying the index or payout, useful information can be attained by tying contract payouts to multiple events. A simple “winner-take-all” combination contract traded before the 2004 Presidential election. It paid $1 if President Bush won the election and Osama bin Laden was captured before the election (Wolfers and Zitzewitz, 2004a). When considered along with the basic “winner-take-all” contract on President Bush winning the election, it is possible to infer the affect of capturing bin Laden on President Bush’s chance at re-election. Other combination contracts were mentioned previously with regards to the PAM project. An example of these would be, “How will Syria’s GDP change with the pullout of x American troops?”

Interpreting prices to reveal information is not a new notion: Friedrich Hayek suggested in an essay published in 1948 that prices in naturally occurring, free markets make important contributions to information in economies (Hayek, 1948). Why haven’t markets like those described above been established earlier with the particular intent of gathering information? The answer is based on the computer age in which we currently live. The IEM began (as indicated by its name) as an electronic exchange. Similar in-house networks as well as the internet have made trading feasible on the somewhat obscure topics by participants whose primary job is not trading. The speed, distance covered and efficiency required mean that computers are a
necessity for feasible market action. This leads to more important questions: Who is participating in these markets, and how does this population affect the accuracy of the market? Wolfers and Zitzewitz (2005a) tackle questions like these in their exploration of the fundamentals of prediction markets.

Information markets require some uninformed traders in order to function. These traders will bring necessary uninformed order flow to the market – with only rational traders in the market basing decisions on expected returns, the No Trade Theorem binds and the market is frozen (Milgrom and Stokey, 1982). Uninformed order flow can come from a variety of motivations – entertainment, overconfidence, hedging – but only the latter fits very well into economic models, and so it can be difficult to quantify the flow and determine which markets will be successful. Most prediction markets to this point do not have enough volume for meaningful hedging, so entertainment and overconfidence serve as the primary two motivators for voluntary participation. Wall Street is over-represented in information markets, but it is reasonable to assume that experienced stock traders will also have diverse views.

Ultimately, these markets work for a variety of reasons. First, in many of the online markets, almost anyone can participate, which leads to the necessary volume and broad variety of beliefs. Second, and perhaps most fundamentally, the markets demand participants to “put their money where their mouth is.” Experts and politicians make forecasts all the time, but talk is cheap. The monetary (or otherwise rewarding) incentive motivates participants to honestly reveal their information in transactions. Finally, due to the profit-potential, traders are driven to find more accurate information to use, which inherently makes the markets even more accurate.

---

3 “Uninformed” is used here mainly to contrast the idea of “well-informed.” It does not mean ignorant, but rather a set of traders whose estimate of the true probability is less precise than the market-maker’s, or those who are willing to trade with that type of behavior for hedging purposes.
This aggregate information manifestation is what columnist James Surowiecki calls “The Wisdom of Crowds” in his new book.

Prediction Markets Today

Despite the DARPA incident of 2003, prediction markets have been gaining popularity. Legal restrictions in the United States established by the Commodity Futures Trading Commission (CFTC) have strictly limited futures trading to financial derivatives. Because of this, a number of off-shore sites have arisen to provide markets on more traditionally-based gambling and political topics. Websites like www.tradesports.com and www.betfair.com are two popular examples. There are also forums that remain onshore, though have made compromises with the CFTC. The IEM has promised to limit positions to $500 (which also hinders meaningful hedging, even if adequate volume is present). Platforms like NewsFutures (www.newsfutures.com), the Hollywood Stock Exchange (www.hsx.com) and The Foresight Exchange (www.ideosphere.com) all operate using play money, though sometimes redeemable for prizes. At least two markets have been sanctioned by the CFTC to operate as futures exchanges, Economic Derivatives, hosted by Goldman Sachs and Deutsche Bank, and HedgeStreet (www.hedgestreet.com). Neither has engaged in trading political contracts, but rather have remained in the economic realm. This avoids the association with pleasure-gambling and moves more in the direction of legitimate hedging and derivative trading.

Now that a general base for this novel information aggregation mechanism has been outlined, I will move to the research that has been done to explain and modify the original model of the IEM, express concerns to using markets for policy-making, and the support for the accuracy and efficiency of the markets.
This Study

Information markets have had numerous examples of accurate forecasts of events. More complicated contracts can reveal more meaningful results, but sometimes at the cost of lower volume (or more noise) for the final contract price. The second layer of complexity generally comes from estimating a variable conditional upon an outcome. These contingent markets were dubbed “decision markets” by Robin Hanson since they are theoretically better at revealing an outcome based upon the decision made. Well-established examples are the contingent markets run by the IEM for presidential elections. The contingent market prices can be used in harmony with traditional information market prices for some novel insights.

Wolfers and Zitzewitz (2004b) analyzed the 2004 Democratic Presidential nomination process from the perspective of information markets.

Table 1: 2004 Presidential Election Contingent Markets
Contracts pay according to vote-share, Conditional on Democratic Nominee

<table>
<thead>
<tr>
<th>Candidate</th>
<th>Democratic Vote Share</th>
<th>Republican Vote Share Against this Candidate</th>
<th>Implied Probability of Nomination</th>
<th>Expected Share of Popular Vote if Nominated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wesley Clark</td>
<td>$0.017</td>
<td>$0.026</td>
<td>4.30%</td>
<td>39.5%</td>
</tr>
<tr>
<td>Howard Dean</td>
<td>$0.018</td>
<td>$0.027</td>
<td>4.50%</td>
<td>40.0%</td>
</tr>
<tr>
<td>John Edwards</td>
<td>$0.084</td>
<td>$0.055</td>
<td>13.9%</td>
<td>60.4%</td>
</tr>
<tr>
<td>John Kerry</td>
<td>$0.390</td>
<td>$0.391</td>
<td>78.1%</td>
<td>49.9%</td>
</tr>
<tr>
<td>Other Dems</td>
<td>$0.014</td>
<td>$0.013</td>
<td>2.70%</td>
<td>51.9%</td>
</tr>
</tbody>
</table>

Notes: Columns A and B show the price of contracts that pay one penny for every percentage of Democratic or Republican two-party popular vote the candidate wins, given the candidate is nominated.
The IEM ran a contingent market for the primary candidates, with some of the contracts are shown in Table 1. These contracts pay a penny for each percentage of the two-party popular vote won by each party, but pays only if the democratic nominee is also successfully predicted. Because of the two-party structure of the contracts, a portfolio containing a Democratic vote share and a Republican vote share contract tied to Kerry will pay $1 if Kerry wins and $0 if he loses. Essentially, a “winner-take-all” market is embedded in the contingent market, and so the sum of the prices of the two contracts in (Column A and B in the table) in this portfolio can be seen as the probability that Kerry will win the nomination (Column C).

Column D is the focus of the table. Dividing the democratic vote share tied to a certain candidate by that candidate’s probability of being nominated yields the market’s probability of winning given that the candidate is nominated.4 In this column, John Kerry is predicted to have a 50% chance against President Bush, but John Edwards is predicted to have a significantly better 60% chance. Wolfers and Zitzewitz interpret this higher chance of success with some wariness. The structure of a contingent contract reads, “Given that John Edwards in the Democratic nominee, what would the Democratic two-party vote-share be?” Considering that the prediction markets considered John Edwards to have a 14% chance versus John Kerry’s 78% chance of being the Democratic nominee, the current question of John Edwards’ vote-share is a somewhat moot point, and so some bias is built into the contingent analysis. What could change the state of the world for John Edwards so that this estimate would be more meaningful? There could be a change in support for his platform, or he could turn out to be a more scintillating orator than expected, but the campaigns had run long enough that it is unlikely the results from the above table would change due to that. There could also be a shift due to an exogenous event: John Kerry falling ill, or being implicated in a sex scandal. If the state of the world changed in

\[ \frac{E[DV|\text{candidate } X]}{Pr\{\text{candidate } X\}} = E[DV \cap \text{candidate } X] = \text{Col D} \]
this way, John Edwards would have likely become the favorite for the Democratic nomination, and the question is raised, “How accurate is the 60% vote-share estimate?” This study will try to emulate a parallel scenario of a favorite falling from its top-ranked position due to an event unrelated to his nearest competitor. The bias of the preceding logic can be estimated by observing an actual exogenous event that affects the probability of the underdog.

**Literature Review**

The paper thus far has explained the fundamental concepts of the mechanism of an information market. As the popularity of the markets has burgeoned from their origins at the Iowa Electronic Markets, theorists, experimental researchers and businesses have come up with inventive ways to apply the instrument and gain meaningful results. Some skepticism has accompanied the new applications, inspiring more research and models to defend prediction markets. This literature review will more thoroughly analyze some of the previous empirical work done with prediction markets, the necessary conditions for functioning prediction markets, and outline the primary concerns about the markets and their refutations. Finally, it will examine and explore the exciting reason for studying information markets at all: applications that are currently in placed or have been suggested for future implementation.

**The Need for Information Markets – Traditional Methods and Their Failures**

The ability of information markets to incorporate information could be useful above and beyond the benign examples of presidential elections and GDP estimates. Some of the following examples illustrate cases in which information markets could have been to produce a more efficient outcome, and even in some cases, avoid tragedy.

Traditionally, information from groups is provided in one of two ways: the statistical mean (for any size of group) and group deliberations (if the group is small enough). Sunstein
(2004) outlines the benefits and challenges of gathering information from a group in each of these ways. The statistical mean can be obtained by any simple method of averaging beliefs (the guesses of the number of jelly beans in a jar) or discrete variables (taking a vote or a poll). This method has merit in situations where the question posed has a definite answer such as “How many home runs did Babe Ruth hit?” or “When was Herbert Hoover President of the United States?” An accurate estimate is achieved when a group is polled and their responses are averaged. While consultation with an expert would also provide a correct answer, studies have shown that the average response of a group for an estimate of jelly beans will outperform most of the individual responses, consulting a group of experts about inconclusive topics will yield more meaningful results than most individual experts. In a series of thirty comparisons, groups of experts had 12.5% fewer errors on forecasting tasks involving such diverse issues as company earnings, cattle and chicken prices, real and nominal GNP, survival of patients, and housing starts (Armstrong, 2001).

Errors in this method arise when a group has systematic bias. Suppose that a jar had 800 jelly beans. If an experimenter said, “Some jars, though not necessarily this one, hold 500 jelly beans,” before a subject submitted their guess, guesses would be biased down because of the artificial “anchor point.” Another source of error can arise when a group is unfamiliar with the material. A study at the University of Chicago Law School asked faculty members to estimate the weight, in pounds, of fuel that powers a space shuttle. The correct answer is about 4 million pounds, but the median was 200,000, and the mean was 55,790,555 (due to an obvious outlier) – both estimates wildly off. Similar results have been reported when people are asked to estimate the number of atoms in a jelly bean jar, as opposed to the jelly beans themselves.
Similarly, group deliberations (e.g. a jury or focus group) are often employed to aggregate information or reach a consensus, but these can encounter problems that lead to inaccuracy. Deliberations can be useful because of interpersonal contact: they allow the group to inquire about vague statements, possibly point out errors in logic of other group members, expose misleading information or display confidence in knowledge. Unfortunately, this same interactive feature is one of the largest drawbacks of these types of groups. Statements and acts of some group members will convey relevant information, and that information will cause other group members not to disclose what they know. The social setting of deliberations can cause some group members to not reveal their information due to fear of disapproval and associated harms. Confidence in personal information can lead to group polarization, which can be detrimental if that information is inaccurate. The term “groupthink”, coined by Irvin Janis, is often used for this type of behavior. The aspects of groupthink may simply seem like the setting for a psychology experiment or the premise of the movie *Twelve Angry Men*, but its perils have recently been implicated in issues of national security.

Faulty information from the Central Intelligence Agency (CIA) upon which the Bush administration justified the invasion of Iraq caused a wave of insecurity to sweep the country. The Senate Select Committee on Intelligence accused the CIA of groupthink in its report on the situation, claiming that the agency’s predisposition to find a serious threat from Iraq led it to fail to explore alternative possibilities or to obtain and use the information that it actually held.\(^5\) In the Committee’s view, the CIA, “demonstrated several aspects of group think: examining few alternatives, selective gathering of information, pressure to conform within the group or withhold

\(^5\) Available at http://intelligence.senate.gov
criticism, and collective rationalization.”6 In sum with other conclusions, the Committee
stressed the CIA’s failure to elicit and aggregate information.

A second example of government agency failure that did not utilize information in their
possession to prevent disaster comes from just a year earlier: the 2003 Space shuttle Columbia
accident. The Columbia Accident Investigation Board squarely placed the blame for the accident
on NASA’s internal culture. In the Board’s view, NASA “lacks checks and balances.” At
NASA, “it is difficult for minority and dissenting opinions to percolate up through the agency’s
hierarchy.” In his book, “The Wisdom of Crowds,” James Surowiecki reports that the Mission
Management Team (MMT) leader for the final Columbia mission, Linda Ham, received the
report of damage done by a large piece of foam that had broken off one part of the shuttle and
smashed into the ship’s left wing and unilaterally dismissed it moments later as inconsequential.
The information from the Debris Assessment Team (DAT) engineers was innocuously presented
to her, and she exacerbated the mistake by ruling out any probability of significant damage. In
retrospect it is easy to point out the relevant data from vast volumes the team had to sift through
in a short amount of time, but with a better system to percolate the relevance of information to
decision makers, perhaps the crisis could have been averted.

A third example that strikes close to home is the case for the abolished Policy Analysis
Markets. Surowiecki points out that the CIA did not anticipate the 1993 bombing of the World
Not until the September 11th incident was significant attention paid to this shortfall in
information aggregation. Surowiecki posits that the CIA may have had relevant information in
these scenarios, but the vast amounts of information the CIA processes and its inefficient internal
organization in passing that information precluded meaningful steps from being taken to prevent

6 Id., conclusion 4
these incidents. The PAM could have been an instrument within the CIA that could have helped
to sift information that simple human observation could not tackle.

**Questions Often Raised About Information Markets**

The concept of prediction markets has significant appeal – they allow quick, efficient,
small-scale information gathering that could help address major problems. How do these
features function, and what could cause the markets to under perform, not function, or plainly be
inaccurate? Wolfers and Zitzewitz (2005a) establish some necessary conditions for functioning
markets. As mentioned above, it is necessary to have some “uninformed” traders in the market
to provide motivation for trade. This can come from overconfidence, the thrill of a gamble or
hedging. It is essential that there is enough diversity in the information of the participants so that
both sides of a trade will be taken since all trades happen at the margin in the continuous double-
auction format. Given that, the price will move toward market beliefs irregardless of market
volume. The authors also point out the value of the wording of the contracts themselves to
attract traders. Contracts must be written on topics that are interesting and meaningful to traders.
Here again arises the example of guessing the number of atoms in a jelly bean jar rather than the
jelly beans. Thin markets can function well, but adding traders typically increases accuracy.
Furthermore, contracts require clear wording to avoid ambiguities in the definition of outcomes.
Wolfers and Zitzewitz point to the contract, “When will Yasser Arafat leave the Palestinian
state?” and the discussion that arose as to whether the wording included visiting Paris for
medical treatment, or just death. While some financial and economic indices can be hard to
write for, policy outcomes can be even more difficult to summarize. Suppose an information
market were run on the crime rate in Baghdad. The market designer cares about actual crime
rate, but the contracts will probably be written on reported crime rates. One can imagine where
these rates would move in opposite directions: if crime were so high it caused people to report fewer crimes, or if individuals with influence over the reported crime rate chose to exercise it to their benefit.

Honesty and Foul Play

As long as information markets have existed, there have been concerns about the intentions of participants. What if a subset of the population wants to manipulate the policy decision that will result from the outcome of the market and so participates dishonestly? The 1996 US Presidential election prediction, one of the most inaccurate in the history of the IEM, was explained away by a rush of trading the night before the Election Day with the aim of putting Bob Dole in better standing (Berg, Forsythe, Nelson and Rietz, 2003). The prediction error that arose in the week before the election was mitigated by the fact that the history of the market had been very accurate, even though the final prediction deviated significantly from the realized outcome. What if the outcome of a prediction market could affect something more important than election predictions, such as troop movements or school funding? How would policy makers know which estimate to consider most accurate? Is it easy to manipulate the market? A broad body of research has been conducted to address these concerns.

A cornerstone for explaining the accuracy of information markets has been to point to the “put your money where you mouth is” system. In the U.S., this literal interpretation keeps many markets from getting off the ground because they are viewed by the government as betting parlors. As a result, some market creators have introduced play-money exchanges where participants are endowed with some credit and allowed to turn it in at the end of the day for prizes. An immediate concern arises that these markets will be less accurate because participants will be more cavalier with the wealth bequeathed to them. Servan-Schreiber, Wolfers, Pennock
and Galebach (2005) suggested just the opposite, and produced results to support their claim. Wealth in this type of market can only be acquired through previous accurate trades, increasing the incentive to be correct and limiting the possibility of manipulation. The study used market predictions from TradeSports.com (real-money) and NewsFutures.com (play-money) of NFL games to enter a competition called ProbabilityFootball (www.probabilityfootball.com). In this game, participants give personal estimates of the probability a team will win a football game. The quadratic payoff function $100 - 400(w - q)^2$ (where $w$ is an indicator variable for whether the team wins and $q$ is the stated probability assessment) rewards accuracy, but severely punishes extreme incorrect estimates, which encourages commonly-accepted risk-averse behavior. The play-money predictions actually outperformed the real-money in both guessing the winner of games and in the ProbabilityFootball competition.

There have been several known attempts at manipulation in prediction markets, none of which succeeded in a very meaningful way, whether in the actual IEM, the lab, or in historical studies (Wolfers and Zitzewitz, 2004b). This is largely attributed to arbitrageurs monitoring the markets and taking the other side of an offer that takes the price in the wrong direction. In a recent study, Robin Hanson, Ryan Oprea and David Porter (2005) set up their own experimental prediction market and introduced manipulators whose payoffs were tied to something other than accuracy – namely the median price of the market, hence an incentive to drive up the price irregardless of the probability of the outcome. These planted manipulators in a market of merely twelve traders were unable to distort the accuracy of the prices. One hypothesis to explain the effect is that non-manipulators would have an upper threshold at which they would no longer trade – that is, given the common knowledge that some manipulators existed in the market, non-manipulators were less likely to accept prices that seemed inaccurate.
Lab experiments always suffer some scrutiny from critics due to the fact that real markets will always differ inherently – not all population variables can be accurately replicated to get precise results. Robin Hanson has published a handful of papers that develop qualitative and quantitative theoretic models to defend the lab results to complement his lab work. Hanson (2005) explores five various forms of foul play, including lying, manipulation, sabotage, embezzlement and retribution. Ultimately, some of these forms of foul play can be avoided by disallowing particular parties from trading in the market. For instance, if certain advisors to traders were not allowed to trade themselves, they would not have a profitable incentive to lie to their advisees. Increased volume can also help alleviate some of these problems. In the case of manipulation, the market would likely lose some degree of predictive power, though not be rendered useless. An interesting side effect of manipulation comes by virtue of the fact that it increases the incentive for other traders to obtain relevant information and contribute it to the market. Sabotage, a concern that could arise around something like the September 11th attack, or the terrorism futures, can largely be ruled out due to the thinness of information markets. While the markets need some participants and incentives to produce meaningful results, there is relatively little money changing hands, as opposed to similar financial markets. In sum, there are obvious pitfalls with implementing prediction markets, as with any market or mechanism. Thoughtful planning, however, can negate many of the obvious opportunities for foul play so that the markets are not dependent on the good behavior of its participants. In fact, in a previous paper, Hanson (2004) actually proposes that manipulators can actually increase the accuracy of information markets.

Hanson uses a standard Kyle-style microstructure model (a manipulator trades as a noise trader with an informed trader – Kyle, 1989, 1985) to prove that a manipulator added to a market
has no effect on market price, and the increased variance causes informed traders to gather even more relevant information. The increased variance in target trade prices increases the average price, largely due to increasing returns to informed trading in this case. An essential detail lies in the assumption that the markets will continue operating long enough for the correction to be made.

*Interpreting Prices in Decision Markets*

Once a market is established, contracts are well-written and participation is high enough to have adequate trading, the market designer is left with a data set or point to use as a piece of evidence in a decision. Typically closing prices on “winner-take-all” contracts are interpreted as probabilities. A rational, risk-neutral trader would only buy a contract for $.60 that paid a dollar if he felt the event it was written for had a 60 percent probability of occurring. Charles Manski (2004) disputes this common notion by arguing that prices are not an accurate indicator of the market’s beliefs but rather within a range of its beliefs. The basic premise for this charge comes from his model in which he shows that traders’ true beliefs are higher than price when the price would indicate a probability greater than 50 percent, and conversely traders’ true beliefs lie below price when price indicates a probability less than 50 percent. Manski concludes that prediction markets do not follow the assumptions of the traditional efficient market models, “hence it is no surprise that efficient markets do not emerge.” Manski’s solution: “Whereas prediction markets only provide data on persons who choose to trade, surveys enable measurement of the expectations of random samples of populations interest.” Not surprisingly, Manski’s accusatory paper drew attention of other researchers in the field.

Manski essentially challenged the field to produce a credible model of prediction markets and demonstrate that they represent the views of participants. Wolfers and Zitzewitz (2005b)
produced a general model for prediction markets in which prices can deviate from mean beliefs, but show that the deviation is generally small. They emphasize that the model put forth by Manski is a specific case of the general model with extreme assumptions that lead to the most extreme conclusions: that prediction market prices *can* deviate from mean beliefs. These assumptions included a risk-neutral trader who invests his entire wealth on one side of the market or the other depending upon where price falls relative to their personal beliefs. Wolfers and Zitzewitz demonstrate with a set of realistic utility curves and varying range of beliefs that prices generally diverge very little from market mean beliefs. They review a supremely successful prediction market application. Recall the ProbabilityFootball competition in which Wolfers et al. compared the accuracy of play-money estimates vis-à-vis real-money estimates. This is a competition in which typical risk-averse behavior (due to the quadratic payoff function) eliminates polar activity as in the Manski model. Wolfers and Zitzewitz placed mock entrants using these estimates in a season of ProbabilityFootball and placed 7th and 9th out of nearly 2000 entrants – a fairly convincing performance to support interpreting prediction market prices as probabilities.

**Current and Future Applications**

To this point, I have explored the history of prediction markets, some of their theoretical bases and empirical justifications. They have been held up to scrutiny from a number of directions and emerged either unscathed or modified. While all of this background is somewhat necessary and exciting, the real value of information markets is in the applications, current and future.
Old Data, New Ideas

The Policy Analysis Market (PAM) experiment clearly revolved around a sensitive topic after 9/11, and was strongly discouraged because of that. Leigh, Wolfers and Zitzewitz (2003) tackle a slightly less incendiary issue when they analyze the cost of war in Iraq based on predictions of financial markets. The authors acknowledge that war presents profound ethical questions, and financial cost is not always among them. Part of the reason for not heavily weighing the financial burden on America in the case of the 2003 war with Iraq could stem from the broad range of estimates on duration and (correlated) overall cost of the war. The authors propose a method for estimating costs ex-ante, which, if accurate, could be used as a factor in policy-making. Wolfers and Zitzewitz (2004c) revisit the topic (ex-post the capture of Saddam Hussein) to evaluate the precision of their original analysis.

The ex-ante analysis, completed just days before the invasion of Iraq began, relies on prediction markets for “Saddam Securities,” or series of futures markets that trade contracts differentiated by the date by which Saddam Hussein would be captured. The Saddam Security is used as a proxy for the probability of war. It is compared with William Saletan’s “Saddameter,” a weekly journalist for Slate.com who ended each article with a probability of war. The markets for Saddam Securities closely followed the Saddameter. Though the contracts were written for the date of Saddam’s ousting rather than the outbreak of war, most experts believed that Saddam would only be ousted by war; if war broke out, expert consensus was that combat would end very quickly. This indicator is used to project the changes in both oil prices and assets if war broke out: a 10 percent increase in the probability of war correlated with a $1 increase per barrel of oil. When this outcome is scaled, it implies that the cost of war in the oil sector is $10 per
The authors point out at least five valid reasons that assets would likely lose value due to war. Given this premise to build upon, they regressed the S&P 500 on the Saddam Security. Their results were not dissimilar from negative effects in the oil sector: a 10 percent increase in the probability of war was associated with between a 1.1 to a 2.6 percent decline in the S&P 500. An increase in oil prices will slow GDP growth more than defense spending would increase it, a secondary detrimental effect of higher oil prices, the impact in different sectors, and on foreign GDPs. In conclusion, projected decline of the S&P 500 alone would account for the destruction of approximately $1.1 trillion of wealth, an order of magnitude greater than the most extreme expert ex-ante estimate. Using S&P option prices, the authors were able to create a slightly more telling distribution of the probabilities of S&P percentage decline: 70 percent probability that the war would have an effect in the range of 0 to -15 percentage points, 20 percent probability of a -15 to -30 percentage point drop-off, and a 10 percent chance of even larger declines. Fifteen months later, Wolfers and Zitzewitz returned to their work to examine their methodology and the validity of the predictions they developed using the Saddam Securities.

The invasion began 48 hours after the ultimatum issued by President Bush in March, which did not come as a surprise to most experts. This concept is the first to be addressed: the Saddam Security was largely driven by war-related news, and was in turn highly correlated with asset and oil prices (30 and 75 percent explaining power, respectively). Previous studies had shown that the markets reacted very little to news, but the logic follows here that indeed the markets react quickly to incorporate this type of news. Once the war began, the price of the June Saddam Security shot to 95 percent, an indication that the markets felt it would be a short engagement. The S&P actually rallied during this period, and oil prices fell. This reversal in

---

7 It’s worthwhile to note that most experts expected oil production in a post-Saddam era to increase oil production, perhaps by a factor of two. The increased supply would have a long-term positive effect, but that is not the forte of the prediction markets in this case.
correlation with the Saddam Security could be explained by the states of the world: before the war, an increase in the probability of war was a bad prospect for the markets. Once the war began, the Saddam Security became a proxy for the probability of duration. From that point on, the security should be interpreted on the basis that given the present war, combat would be short. Increasing confidence in a short war correlated with increasing asset value and falling oil prices.

The mixed success of this study set some valuable precedents. This was the first time a study in this vein had been conducted, which leaves room for modifications and larger observations sets in the future. More complex conditional contracts could extract more meaningful results, like \( E(S&P|\text{War}) \) and \( E(S&P|\text{NoWar}) \). In addition to estimating the effects of a policy, the study was able to create a probability distribution over all cost outcomes. If the government were to consider these estimates in their policy-making, an endogeneity issue would lead to different biases. Given the White House’s explicit disdain for considering polls and mechanisms similar to these in policy decisions, it will probably not be a problem in the near future.\(^8\)

A study that the President may have found more useful (though intuitively straightforward) also came from Wolfers and Zitzewitz (2004a). Tradesports.com has offered contracts written for the date of Osama bin Laden’s capture (not dissimilar from the Saddam Security above, or some of the PAM proposed contracts), GDP, unemployment, and many other indices. Wolfers and Zitzewitz asked tradesports.com to run a contingent prediction market selling contracts that linked the outcome of the 2004 presidential election with these specific events that could influence the election. Specifically, the capture of bin Laden before the

---

\(^8\) In a national news conference held April 28\(^{th}\), 2005, President Bush made the comment, “I don't think you can make good, sound decisions based upon polls. And I don't think the American people want a president who relies upon polls and focus groups to make decisions for the American people.” Full transcript at [http://www.usatoday.com/news/washington/2005-04-28-bush-transcript_x.htm](http://www.usatoday.com/news/washington/2005-04-28-bush-transcript_x.htm)
election, unemployment rate falling to 5 percent or below, and the terror alert level reaching red. By interpreting the prices as probabilities, Wolfers and Zitzewitz concluded that capturing bin Laden before the election would increase Bush’s chances of winning from a healthy 64 percent to an overwhelming 91 percent. Their results indicated that a healthier economy would give Bush an 80 percent chance of winning, and a high terror alert level would also rally the country around the incumbent. These are amazing swings in public opinion compared to more traditional methods of wooing the public; the probability of a Bush victory moved from 66 percent to 64 percent after the first presidential debate, which was considered to have been a clear victory for Kerry.

The Coming Revolution

Economists are not afraid to speculate about the future applications of information markets and have set no limit on their practicality or scope. Robin Hanson outlines a modified government that would serve to oversee the execution of economic policies which prediction markets determine to be the most favorable for national welfare in an article entitled, “Shall We Vote on Values, But Bet on Beliefs?” (2003). While this radical idea will not likely be implemented in its entirety (he addresses no fewer than thirty design objections that could be immediately raised), it is a bold step forward in suggesting that information markets could permeate deeper into policy-making and have a meaningful impact.

Robert Hahn and Paul Tetlock (2004) have also endorsed ideas of similar magnitude. They use a relatively tame but paramount example in domestic public education. Suppose the school board of New York City is interested in improving the standardized tests scores if its high school students, and is willing to pay $1 million for each point that test scores go up in a year. The government could hold an auction for education contractors and the highest bidder (i.e. the
bider who can increase scores the most for the lowest cost) and it would be a classic form of
performance based policy (PBP). If instead an information market is designed and trades two
contracts, one that pays $x for test scores with no change in policy and one that pays $y for the
new policy. If these contracts were trading for $75 and $80 respectively just before a policy’s
implementation, the market-maker could infer the policy would increase scores about 5 points.
Another information market could be run to estimate the likely revenues of the ensuing auction.
While these results are still estimates and not precise, they have been shown to perform better
than expert opinions, and so the market-maker has valuable information for the very low price of
running an information market. The bidders in the auctions could use these results to modify
proposals and bidding strategies. A third derivative benefit could come in the form of hedging to
manage the risk of a proposed policy. Parents wary of the change could participate on the short
side of contracts and then have enough money to send their child to private school if the new
management did not fulfill their promises.

Hahn and Tetlock suggest a similar example that fits this model in the non-profit sector.
What if the Gates Foundation could put a value on the reduction of individual HIV infections in
sub-Saharan Africa before a certain time. They could run similar prediction markets to produce
estimate of the number of infections with and without a certain policy, and hence also the
benefit. Another market could be run to estimate the cost of the policy, as with the school
example. Wolfers and Zitzewitz (2004c) suggest a similar contingent contract written on oil
prices. In 2002, a contract could pay $P if Saddam were ousted in 2002 (where P is the future oil
price) and another that would pay $P if Saddam were not ousted. The purchase price would be
refunded in the case not realized. The difference in the price of these two securities is the
market’s expected change in oil prices if Saddam were to be ousted. The benefit of this accurate
estimate is principally time: researchers do not have to wait for political variation to run a regression that estimates this effect. Each of these examples demonstrate the potential utilization of prediction markets in a slightly different way, though each through a remarkably efficient method.

**Study Design**

**Background**

In the Literature Review I covered the primary research foci and applications of prediction markets – accuracy and foul play, business, political and leisure. My study will explore a very specific though useful derivative: contingent prediction markets. The payoffs of the contracts in these markets are tied to multiple events rather than just one. The complexity brings a tradeoff: it is harder to entice participants, but the enhanced forecasts can be used for more novel applications. Hanson (1999) suggested these contingent markets could be used as the basis for “decision markets.” Decision markets are those designed specifically for the purpose of basing one single decision on the outcome of the market. In contrast, the HP sales forecasts may be useful for a series of decisions, and markets run on sporting events are run for no decision making purpose at all, so their loose categorization comes from predictive powers.

Rietz and Berg (2003) performed a study ex-post on the role prediction markets could have played in the Republican nomination process of the 1996 presidential elections using market data from the IEM. Their logic was simple: through the primary nomination process, the nominating party is more likely to select a candidate on the grounds of popularity within the party rather than with the population as a whole. But if the party wants to win the election, they need a candidate who appeals to the population at large. The IEM ran conditional index markets with contracts like, “$1 times the two-party Republican vote-share if Bob Dole is nominated,”
“$1 times the two-party Democratic vote-share if Bob Dole is nominated,” “$1 times the two-party Republican vote-share if Lamar Alexander is nominated,” etc. By using these contingent prices, the party could have estimated the strength of the candidates in the national election prior to a nomination. Their study found Colin Powell to be a very strong candidate before he declared a year before the election that he would certainly not be running. Their regression of President Clinton’s probability of re-election on the probability of Powell’s nomination showed that the increase in probability of Powell receiving the Republican nomination (based on a winner-take-all contract of “Powell as Republican Nominee”) had a significant negative coefficient; had he run, Powell would have been likely to win against the incumbent president. The Republican Party could have used this type of result to support stronger recruiting of Powell, especially given the predictions of the rest of the nominees. As can be seen in figure 2, when Powell announced that he would not pursue the Republican nomination, Bill Clinton’s probability of victory spiked.

**Figure 1:** Forecast probabilities for Powell’s name being placed in nomination at the Republican convention from the Powell nomination market and forecast probabilities for Clinton winning the election (defined as receiving the most popular votes) from the Presidential winner-takes-all market. Source: Rietz and Berg, 2003.
Bob Dole, the eventual Republican nominee, was shown prior to his nomination to be a weak candidate against President Clinton. An increase in the probability of Dole’s nomination implied an increase in the probability of a Clinton victory (Figure 2). Further, there was always another candidate for the nomination that the market predicted to be stronger than Dole through Super Tuesday and beyond, when Dole’s nomination was essentially assured. Irregardless, Dole was eventually nominated and Clinton was re-elected by a comfortable margin.

**Figure 2** - Forecast probabilities for Dole becoming the Republican nominee from conditional contracts in the Presidential vote-share market, the percentage of delegates committed to Dole from primary and caucus results and forecast probabilities for Clinton winning the election (defined as receiving the most popular votes) from the Presidential winner-takes-all market. Source: Berg and Rietz, 2003.

Wolfers and Zitzewitz (2004b) return to contingent markets run during the 2004 Democratic presidential primaries. Prediction market contract prices for specific nominees
reasonably followed probabilities of nomination, but the resulting contingent contracts probably misrepresented the expected vote share if nominated. Expected vote share if nominated can be constructed by the simple Bayesian relation between two contracts on the right side of equation 1 that were run by the IEM:

\[
E[\text{Dem. Two-Party Vote Share} \mid \text{Candidate X}] = \frac{E[\text{Dem. Two-Party Vote Share} \cap \text{Candidate X}]}{\Pr\{\text{Candidate X Wins Dem. Nomination}\}}
\]

(1)

**Table 2: Average Expected Vote Share for Democratic Nominee Candidates**

<table>
<thead>
<tr>
<th>Candidate:</th>
<th>Clinton</th>
<th>Clark</th>
<th>Dean</th>
<th>Gephardt</th>
<th>Kerry</th>
<th>Liebermann</th>
<th>Edwards</th>
<th>ODems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Expected Vote Share:</td>
<td>49.28%</td>
<td>50.14%</td>
<td>47.63%</td>
<td>56.75%</td>
<td>48.29%</td>
<td>45.68%</td>
<td>52.01%</td>
<td>45.39%</td>
</tr>
<tr>
<td></td>
<td>(0.066)</td>
<td>(0.061)</td>
<td>(0.042)</td>
<td>(0.086)</td>
<td>(0.047)</td>
<td>(0.096)</td>
<td>(0.036)</td>
<td>(0.084)</td>
</tr>
</tbody>
</table>

Notes: Average over the life of the Implied Expected Democratic Two-Party Vote Share as Calculated in Column D, Table 1; Standard Deviations in Parentheses; Source: Closing prices as of February 2, 2004, Iowa Electronic Markets (http://www.biz.uiowa.edu/iem)

The data in Tables 1 (above) and 2 indicate that John Edwards would have been a strong candidate in the same manner that Colin Powell would have been a strong candidate, or Bob Dole likely was a weak candidate.

Figure 3 shows the markets were consistently more optimistic about John Kerry’s performance in the general election than Howard Dean’s, though John Edwards regularly surfaced as the “best candidate” after officially entering the race in late January. Since Edwards’ probability of being nominated was so low, it is fair to assume the contingent estimate actually represents a different state of the world, and the estimate overstates the expected vote share in this state of the world. This misrepresentation would yield tainted information not suitable for making policy decisions. To have meaningful information, it will be important to understand the
nature and magnitude of the problem. The following research is aimed at determining how much bias is in this estimate.

Figure 3: Probability of Nomination and Implied Expected Vote Share of Democratic Candidates

Environment and Data Selection

Sports betting markets have long been understood to be remarkably accurate estimates of the outcomes of games. A broad body of literature has been published defending this notion with data from Las Vegas odds-makers. (For some sources on the topic, see: Woodland and Woodland, 1994, Brown and Abraham, 2002, Dowie, 1976, Even and Noble, 1992, Williams, 1999.) All of the literature reviewed above indicates a remarkable accuracy of prediction markets, and so it logically follows that prediction markets about sporting events in particular will also be accurate. Wolfers and Zitzewitz (2004) demonstrate the accuracy of sports
prediction markets by entering prediction market estimates into the ProbabilityFootball.com competition. Sports prediction markets are a useful place to turn for data given their broad participation, high volume and long duration.

Professional sports leagues determine an annual champion through a series of regular season games followed by a post-season tournament (usually called the playoffs) for teams that qualify during the regular season. The five leagues that I focus on (and have the most volume in trading) are Major League Baseball (MLB), the National Basketball Association (NBA), the National Football League (NFL), the National Hockey League (NHL) and the National Collegiate Athletic Association’s (NCAA) Division I men’s basketball. Each of the leagues seed teams for their post-season tournament (as opposed to a round-robin or random tournament), albeit in unique ways. For this study I define the #1 and #2 seed to be the teams that are designated by the league as the team with the greatest chance and second greatest chance, respectively, to represent the conference in the championship matchup.

I look at these tournaments as scenarios that parallel the presidential candidate nomination process. In particular I look at movement in prices of the #2 seeds when a #1 seed is eliminated or approaches elimination by a team other than the #2 seed. I can use this event to do an IV regression similar to the hypothetical John Kerry scandal (#1 seed) and see what underlying bias there is in estimating how the John Edwards (#2 seed) would perform in the different state of the world. Each sports team has two contracts. The first is, “Team X will win the conference,” and the second is, “Team X will win the league championship.” The first is

---

9 I considered using contract prices the night before the playoffs began to make seeding based upon the probability of winning the league championship. Sometimes these seedings would differ from league seedings since the leagues would give priority to divisions and conferences within the league, so a team may be seeded below a weaker team by virtue of the fact of being in the same division or conference as a team stronger than both. This method was abandoned for a combination of two reasons. First, the top seed by probability was always the top seed in the league tournament. Second, rearranging seeds by market probabilities would not rearrange the matchups in the tournaments, which are the events I am analyzing and part of the information the semi-strong Efficient Market Hypothesis says the prices account for anyhow.
analogous to winning the primary Presidential nomination: if Team $X$ wins the conference, they represent their conference in the league championship against the winner of the other conference.\(^{10}\) The second is analogous to the contingent contract of expected vote share in the election given that the candidate is $X$: it is implicitly written that Team $X$ will win the conference given that it is playing in the championship. Similar to the vote-share contract paying 0 if the wrong candidate is selected (i.e. the candidate receives 0 of the popular vote), the championship contract will not pay if Team $X$ does not actually make it to the championship.

Data was provided by tradsports.com, an online forum for cash prediction markets on events that range from politics to the economy to current events to sports. The data for each of the contracts are a time series of midnight closing prices. Tradesports.com charges a nominal $.04 fee on transactions for contracts that could pay $10. Data included all of the tournaments for the five leagues for which tradsports.com ran the appropriate contingent contracts, which encompassed the seasons from 2001 through 2004. I used the following rules to choose which series to compare:

Case #1: Upset
If highest seed loses before the semifinals\(^{11}\), the contract prices of the next highest seed are included in the data set.

Case #2: Elimination Game\(^{12}\)
If highest seed reaches an elimination\(^{13}\) game, the contract prices of the next highest seed is included in the data set.

\(^{10}\) This is nuanced for NCAA basketball where there are seedings for four conferences (or regions) rather than two conferences, but the analysis principle remains the same.

\(^{11}\) If a #1 seed loses beyond this point, it will be to the next highest seed in its league, and so the loss will no longer be an exogenous event to the next highest seed.

\(^{12}\) Note: All the contracts from rule #1 can be used for some data points in #2 since the top seed will reach an elimination game before they are eliminated.

\(^{13}\) In single elimination tournaments (NCAA, NFL), every game is an elimination game. There is not enough trading during games involving the #1 seed on the idle #2 seed, so only the MLB, NBA and NHL playoffs are considered for this case.
Case #3: Change in Probability of Upset
Of the contracts that fall under case #1 and 2, I consider the effects of wins and losses by
the #1 seed on the idle and adjusted prices of the #2 seed.

Methodology

The conference and championship sports contracts are analogous to the Democratic
nomination and general election vote share contracts that the IEM ran. If a team wins the
class, they advance to the championship game. Before the playoffs begin, a championship
contract paying $1 if the team wins it all and $0 if they lose anywhere along the way yields the
probability that the team will win the championship and the conference. Thus a team’s strength
upon reaching the finals can be written as

\[
\frac{Pr\{Team\ X\ Wins\ Championship\ |\ Team\ X\ is\ in\ Championship\}}{Pr\{Team\ X\ Wins\ Conference\}} = \frac{Pr\{Team\ X\ Wins\ Championship\}}{Pr\{Team\ X\ Wins\ Conference\}}
\]  

(2)

This ratio is the probability the team would have conditional upon their presence in the finals –
ignoring how likely that event is to occur. An instrumental variables approach could be used
with the sports contracts to estimate the strength in different states of the world – namely, those
in which an upset is more likely, or has actually occurred. The outcome of a game played by the
#1 seed will change the price of the #2 seed’s contracts, ceteris paribus. A loss or upset of the #1
seed would move the #2 seed contract prices up, and conversely a win would drive them down.

For the following model, I’ll look at Case #1: the upset. It can be also be applied when the #1
seed reaches an elimination game, or any time the #1 seed plays and the #2 seed is idle. In each
case, the outcome of the #1 seed’s game is the instrument for the exogenous variable.

We could define four contracts required for an IV estimate:

- \( P_1 = Pr\{#2\ Seed\ Wins\ Conference\} \)
- \( P_2 = Pr\{#2\ Seed\ Wins\ Conference\ |\ Upset\} \)
- \( P_3 = Pr\{#2\ Seed\ Wins\ Championship\} \)
- \( P_4 = Pr\{#2\ Seed\ Wins\ Championship\ |\ Upset\} \)
Where $P_1$ and $P_3$ are the prices given by the conference and championship contracts and $P_{2,t} = P_{1,t+1}$ and $P_{4,t} = P_{3,t+1}$ for days when the #1 seed is upset. $P_2 - P_1$ is analogous to the first stage regression coefficient, which measures the effect of the upset. The difference in the championship price, $P_4 - P_3$ is analogous to the coefficient of the reduced-form regression. The ratio of the reduced-form and first-stage regression coefficients yields the “Prediction IV” Wald estimator:

$$\beta^{\text{Prediction IV}} = \frac{P_4 - P_3}{P_2 - P_1}$$

This estimate is the ratio of the effect of an upset on the probability the #2 seed wins the championship to the change in probability it wins the conference. Whether the news from the #1 seed’s performance is good or bad from the #2 seed’s perspective, it is expected that the contracts will move in the same direction, and the conference make a larger absolute change. Because of this, the Wald estimator is expected to be positive and bounded, or $\beta^{\text{Prediction IV}} \in (0,1)$.

Given the assumption that the upset of a #1 seed doesn’t affect the chances of a #2 seed in the championship, the Wald estimator provides an estimate for states of the world in which the #2 seed would not have likely advanced to the championship save for the exogenous event of the #1 seed being upset.

**Data**

Tradesports.com was able to provide trade-by-trade data for the contracts in the study. Naturally, each trade moves the price at the margin to the next available pair of buyers and sellers, and so each price movement is small. I normalized the set by using the midnight closing
price as a proxy for the aggregate beliefs of the market for that day.\textsuperscript{14} Some days, particularly early in the regular season, had little movement. Conversely, there was relatively heavy trading during the playoffs. The low volume is to be expected at the beginning of the season for two reasons. First, a win or loss early in the regular season provides very little new information (or at least information that would not be lost in noise of a forecast that far in the future) for the strength of a team in the playoffs. Second, the playoffs carry a “hype” in which even casual sports fans will be more likely to be aware of postseason wins and losses. This introduces a higher population base interested in trading with a higher variance of beliefs, both necessary for higher (and generally more accurate) volumes of trade, as discussed above.

<table>
<thead>
<tr>
<th></th>
<th>% Overall Trades</th>
<th>% Overall Volume</th>
<th>Average Contracts/Trade</th>
<th>Average Change in Price each Trade</th>
<th>Average Change in Midnight Closing Prices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Playoffs</td>
<td>68.8%</td>
<td>43.1%</td>
<td>25.5</td>
<td>1.32</td>
<td>.396</td>
</tr>
<tr>
<td></td>
<td>(60.45)</td>
<td>(3.43)</td>
<td></td>
<td></td>
<td>(1.07)</td>
</tr>
<tr>
<td>Regular Season</td>
<td>31.2%</td>
<td>56.9%</td>
<td>74.36</td>
<td>.685</td>
<td>3.61</td>
</tr>
<tr>
<td></td>
<td>(217.68)</td>
<td>(.969)</td>
<td></td>
<td></td>
<td>(7.65)</td>
</tr>
</tbody>
</table>

Notes: Prices are on a scale of 1-100; standard deviation in parentheses
Source: www.tradesports.com selected professional sports league contracts, 2001-2004

Results and Analysis

\textit{#1: The OLS Estimate is Upward-Biased}

The phenomenon that I want to observe is the inherent bias built into the contingent market prices accounted for by different states of the world. Namely, the underdog is favored to perform better than he actually would, should he reach the finals of his tournament. The natural starting point to investigate this type of movement is in the synthesized probability itself ($P_3/P_1$

\textsuperscript{14} At times I also used volume-weighted price averages for a day – the closing price could be significantly different from the opening price due to a couple large trades, and so badly misrepresent the price of most of the day. When the weighted-price set is used, I make a note of it.
above). In contrast to the previously outlined “Prediction IV,” I will refer to this variable as the OLS estimator.

<table>
<thead>
<tr>
<th>Variable</th>
<th>#1 Seed Wins</th>
<th>#1 Seed Loses</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS_change</td>
<td>0.14%</td>
<td>-4.094%</td>
</tr>
<tr>
<td>OLS_adjusted_change</td>
<td>-0.33%</td>
<td>-3.246%</td>
</tr>
<tr>
<td>OLS_percentage_adjusted_change</td>
<td>0.54%</td>
<td>-4.37%</td>
</tr>
</tbody>
</table>

Notes: OLS\_change is the difference in probability of success of the #2 seed given they make the championship between midnight closing prices before and after the #1 seed is active and the #2 seed is idle. OLS\_adjusted\_change and OLS\_percentage\_adjusted also include observations when the #2 seed is active, and change in contract prices due to the #2 seed’s performance have been subtracted out. For a more thorough description, see the Appendix. Source: www.tradesports.com midnight closing prices for selected MLB, NHL, NBA, NFL and NCAA playoff #1 and #2 seeds, 2001-2004.

Table 4 shows that there is likely an over-evaluation of the #2 seed’s performance in the championship before the #1 seed plays. If the #1 seed wins, there is not statistically significant movement in the OLS estimator. This is not very surprising as the state of the world has not changed much – the #1 seed remains in the tournament, looking perhaps marginally stronger than before, or at least performing to expectation. As a result, the performance of the #2 seed in the championship is not re-evaluated. Conversely, the OLS estimator makes large movements if the #1 seed loses – an indicator that changes in the state of the world to one where it is much more likely the #2 seed will actually make it to the championship. In each case, the estimate of the chance of success falls by 3-4%. In the sports world, this amount may be captured in the error term, but in the political world that margin could easily be the difference between victory and defeat, or a landslide and a narrow outcome.
The “Prediction IV” is difficult to estimate from the OLS Estimate

The OLS estimate is a very intuitive measure that stems from the most basic Bayesian dependent relationship. When the data suggested the conventional estimate was biased, I developed the model outlined in the Study Design that would remove the bias. Result #1 shows that the estimate is indeed upward biased, and the next logical step is to determine by how much, and if the “Prediction IV” can be estimated from the OLS ratio.

Table 5: Estimating the Prediction IV from given OLS estimate

<table>
<thead>
<tr>
<th>Dependent Vars:</th>
<th>IV_Estimate</th>
<th>IV_Adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(I)</td>
<td>(II)</td>
</tr>
<tr>
<td>Independent Vars:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>.599</td>
<td>.661</td>
</tr>
<tr>
<td></td>
<td>(.862)</td>
<td>(.861)</td>
</tr>
<tr>
<td>OLS_estimate</td>
<td>-.192</td>
<td>-.329</td>
</tr>
<tr>
<td></td>
<td>(1.19)</td>
<td>(1.19)</td>
</tr>
<tr>
<td>OLS_change</td>
<td></td>
<td>3.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.12)</td>
</tr>
<tr>
<td>OLS_adjusted</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>.002</td>
<td>.092</td>
</tr>
</tbody>
</table>

Notes: ** indicates significance at the 95% level, *** indicates significance at the 99% level. Standard errors in parentheses. Data points selected for (I) and (II) when #1 seed plays and #2 seed is idle. (III) uses an adjusted series described in the appendix.

Source: tradesports.com midnight closing prices

Equations (I) and (II) indicate very little relation between the two measures. The coefficient for the OLS estimate is always negative, which could indicate that as the market projects weaker candidates to be stronger and stronger, the Prediction IV deflates these beliefs more and more. The coefficient in equation (III) is statistically significant, but meaningful for only a small window of OLS estimates.
The IV estimate is a noisy statistic in this data set. The Wald estimator ranges outside of its theoretical bound of (0,1) with extrema at -5.62 and 4.20. This is probably due to the thinness of the markets which results in the conference contract trading very lightly after the upset, but the more popular championship contract reacts more quickly. Table 6, which was created from the weighted midnight prices, shows the championship contracts have significantly more activity both in both quantity and regularity. This logic probably adequately explains the Wald estimator falling out of the bound suggested by the model. In particular, Wald estimators that fell within the (0,1) range resulted from trade volume upwards of 50% larger than days that fell outside of the range. These days were below the playoff means shown in table 6. The difference wasn’t statistically significant, though, due to the high standard deviation of quantities traded. As a result of this thin trading, I eliminated the irregular “prediction IVs” and ran some of the models again. This time, there was more significance, correlation and prediction power.

#3 Prediction IVs That Fit the Model Can Be Estimated By Their Corresponding OLS Estimates

The numbers in table 7 come after a couple transformations of the data. The primary differences are all concepts discussed above: using midnight prices from trades weighted throughout the day, adjusting some prices to account for #2 seed activity and paring Wald
estimators that do not behave well. By accounting for these, regressions are available that estimate the prediction IV quite well. The following models have parallel intentions, but arrive at their conclusions through nuanced transformations. The similarity in results supports the conclusions while the diversity among them helps shed light on the true nature of the relation.

Equation (I) and (II) both suggest that the OLS prediction overstates the strength of the #2 seed in the championship by between a third and a quarter. As would be expected, both of these models suggest the estimates converge around 1. The relatively high $R^2$ and level of significance support the hypothesis that the Bayesian ratio is not applicable to every state of the world.

Equation (III) suggests that weighting wins and losses in different rounds clouds the results – the coefficient is insignificant and the fit of the model is poor. Equation (IV), though, is also adjusted by round and yields an interesting result. It shows that when the data are adjusted by the more-accurate round-weighted data, the OLS estimate has a constant bias built in – it is simply 20% too large at any given time. In general, table 7 supports the findings in table 4 with the direction of the bias, and helps to quantify it.

### Table 7: Estimating the Wald Coefficient from Within the Model

<table>
<thead>
<tr>
<th></th>
<th>$IV_{Adjusted}$</th>
<th>$IV%_{Adjusted}$</th>
<th>$IV_{Adjusted_Round}$</th>
<th>$IV%_{Adjusted_Round}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C$</td>
<td>.296</td>
<td>.115</td>
<td>.677**</td>
<td>-.212</td>
</tr>
<tr>
<td></td>
<td>(.253)</td>
<td>(.160)</td>
<td>(.224)</td>
<td>(.214)</td>
</tr>
<tr>
<td>$OLS_Adjusted$</td>
<td>.625*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.346)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$OLS%_{Adjusted}$</td>
<td></td>
<td>.747**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.267)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$OLS_{Adjusted_Round}$</td>
<td></td>
<td>-.095</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.104)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$OLS%_{Adjusted_Round}$</td>
<td></td>
<td></td>
<td></td>
<td>1.000**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.364)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.424</td>
<td>.611</td>
<td>.173</td>
<td>.519</td>
</tr>
</tbody>
</table>

Notes: * indicates significance at the 85% level, ** at 95%, and *** at 99%; Adjusted Estimates explained in appendix. Adjusting by round simply weights wins and losses differently in different rounds of the playoffs; Adjusting by percentage of contract price makes more accurate adjustments near the extreme prices of 0 and 100.

Source: [www.tradesports.com](http://www.tradesports.com) selected contracts and days when the #1 seed is playing and the #2 seed is idle or adjusted idle.
Conclusions

Decision markets have a promising future as a tool for informed policy making. The 2004 Democratic presidential nomination example shows that the conventional interpretation of prices may overstate the projection for a candidate’s realistic performance. This hypothesized error led me to develop an indicator that would estimate a candidate’s performance in different states of the world, namely one in which the underdog becomes the favorite due to some exogenous event. The analogous sports scenario of earning a berth to the championship game by winning a conference tournament provides a data set with observable exogenous changes to the underdog’s probability of having chance to compete for the championship: the elimination of a higher-seeded team. The contracts from major sports leagues which were traded recently at tradesports.com provided price movements to confirm the hypothesis. When a #1 seed was upset, prices moved to reflect the greater probability that the #2 seed would make it to the championship. The market’s prediction of the #2 seed’s performance in the championship, though, deflated as the event realization became more realistic. My instrumental variable estimate, which used the #1 seed upset, indicated that a more reasonable probability of success for the #2 seed in the championship falls between 65 and 75% of the conventional “OLS” estimate, derived from equation (1) above.

Returning to the Democratic presidential nomination example, John Edwards’ projected two-party general election vote share would have deflated to between 39-45%. Kerry would then be realized as the favorite of the party at 49.9% unless an event unrelated to the John Edwards’ strength of candidacy decreased Kerry’s chance of nomination.
Shortcomings of the Study

Data in this particular vein of research is relatively sparse. That is the unfortunate nature of evaluating a concept before it has had very much time to prove itself. In this particular case, the data set yielded relatively noisy estimates when they were not expected. There were also significant times when the data did not “behave” well. Some unexpected results included only one of the two contracts reacting to relevant information, or the two contracts actually moving in opposite directions. The study did not account for activity in other conferences, injuries to key players, momentum or home court advantage, all of which play significant roles in probability of success. The sample only included one #2 seed that ended up winning the tournament. This is largely due to the fact that trading futures contracts on sports has become popular only relatively recently. It would interesting to consider a larger sample that was able to approximate the accuracy of forecasts given enough #2 seeds that made it to the championship game.

Future Applications

This research is most readily applicable to decision-making in the primary election infrastructure, but can surely have value to other applications. One possible application could also be used with regards to the performance-based policy model that Hahn and Tetlock support, as outlined above. Instead of running an information market to predict the success of a single policy, decision markets could be run for multiple policy change proposals. Contracts could be written that estimate both the strength of a policy vis-à-vis the other proposals and the success of the given policy once implemented. Perhaps the success of the most popular policy would depend on the personnel within the administration, and turnover at these positions would change the likelihood of success.
There could also be logical business applications. A company trying that is trying to choose one product from a set of substitutable prototypes might run an in-house decision market to choose which of the prototypes should be launched, and to estimate their success once out in the market. The probability of a recall for a defective (but initially popular) prototype could serve to help estimate the actual strength of the second-most popular prototype in the market. This is just another parallel scenario to the primary/general election model.

The exciting new field of prediction and decision markets is just beginning to gain momentum. The aim of this study has been to shed light on accurately interpreting their forecasts and supporting the case for their precision. Both research and popularity outside academic circles need to increase before there is widespread use of information markets in the private and public domain. With such a novel and efficient idea, it should not take much more than spreading the word.
Appendix

Calculating OLS\_adjusted and OLS\_percentage\_adjusted

The change in conference and championship contract prices during the playoffs followed a very predictable path depending on whether a team would win or lose in a series. By creating a model to estimate the change based upon winning or losing, and the round in which it happens, I was able to adjust conference and championship contract prices on days when both the #1 and #2 seed played to emulate the #2 seed being idle; that is, adjust for the change in price that came as a result of their performance. Table A1 has the models that were used in the estimates. The coefficients are statistically different from zero in every case but a loss in the first round of the playoffs. Wins generally reward with a greater change than losses penalize. The percentage change was created to accommodate prices near 0 or 100 which cannot move as much with a loss or win, respectively.

Table A1: OLS estimates for Contract Price Adjustments

<table>
<thead>
<tr>
<th>Independent Variables:</th>
<th>Championship Change</th>
<th>Conference Change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>absolute</td>
<td>percentage</td>
</tr>
<tr>
<td></td>
<td>(I)</td>
<td>(II)</td>
</tr>
<tr>
<td>C</td>
<td>.2419</td>
<td>.0444</td>
</tr>
<tr>
<td></td>
<td>(.565)</td>
<td>(.063)</td>
</tr>
<tr>
<td>Win</td>
<td>5.870*</td>
<td>.3087*</td>
</tr>
<tr>
<td></td>
<td>(.776)</td>
<td>(.034)</td>
</tr>
<tr>
<td>Lose</td>
<td>-7.054*</td>
<td>-.2103*</td>
</tr>
<tr>
<td></td>
<td>(.892)</td>
<td>(.039)</td>
</tr>
<tr>
<td>Round1_win</td>
<td>1.679*</td>
<td>.1827*</td>
</tr>
<tr>
<td></td>
<td>(.512)</td>
<td>(.055)</td>
</tr>
</tbody>
</table>

\(^{15}\) Naturally this excludes the NFL and eliminating games since probabilities drop to 0 if the team loses.  
\(^{16}\) Table A1 represents models for the midnight closing prices. A similar but different set of estimates was developed for the midnight weighted prices used in previous adjustments.
| Round1_lose | -1.123 | -.0381 | -1.800 | -.1411 |
|            | (.829) | (.086) | (1.293) | (.115) |
| Round2_win | 5.993* | .3602* | 13.66* | .5955* |
|            | (.500) | (.053) | (.796) | (.073) |
| Round2_lose| -4.378*| -.1586*| -8.991*| -.3041*|
|            | (.599) | (.063) | (.894) | (.081) |

| $R^2$       | .283 | .076 | .263 | .137 | .265 | .182 | .222 | .266 |

Notes: t-stats in parentheses; * indicates significance at the 99% level; Championship_Change and Conference_change are the daily differences in the respective contract prices. Win, Lose and Round are dummies.

Source: www.tradesports.com: midnight closing prices of selected series of professional sports playoff teams.
References


Berg, Joyce, Robert Forsythe, Forrest Nelson and Thomas Rietz, 2003, “Results from a Dozen Years of Election Futures Markets Research”, in Handbook of Experimental Economic Results, (eds: Charles Plott and Vernon Smith)


Hanson, Robin. 2004. “Shall We Vote on Values, But Bet on Beliefs”, mimeo, George Mason University (available at: http://hanson.gmu.edu)


