PRELIMINARY – Comments welcome

Prediction Markets

Justin Wolfers
Stanford GSB and NBER
jwolfers@stanford.edu
www.stanford.edu/people/jwolfers

Eric Zitzewitz
Stanford GSB
ericz@stanford.edu
http://faculty-gsb.stanford.edu/zitzewitz

This Draft: November 19, 2003

Abstract

We analyze the extent to which simple markets can be used to aggregate disperse information into efficient forecasts of unknown future events. Drawing together data from a range of prediction contexts, we show that market-generated forecasts are typically fairly accurate, and they outperform most moderately sophisticated benchmarks. Carefully designed contracts can yield insight into the market's expectations about not only probabilities, means and medians, and also uncertainty about these parameters. Moreover, conditional markets can effectively reveal the market's beliefs about regression coefficients, although we still have the usual problem of disentangling correlation from causation. We discuss a number of market design issues and highlight domains in which prediction markets are most likely to be useful.

The authors would like to thank David Pennock, Emile Servan-Schreiber of NewsFutures, David Dempsey and John Delaney of Tradesports, David Siegel and Scott Hereld of Trackmaster, and George Neumann of IEM for help with data. Thanks to Kay-Yut Chen, Andrew Leigh, Rohan Wolfers, Betsey Stevenson, Tim Taylor, Hal Varian and Craig Yee for stimulating discussions. Doug Geyser, Chris Lion, Paul Reist, Eric Snowberg, and Ravi Pillai provided outstanding research assistance.

1. Introduction

This paper reviews a new—and emerging—form of financial market, often known as prediction markets, but also going by the name "virtual stock market", or "event futures". Analytically these are simply markets where participants trade in contingent commodities—contracts whose payoff depends on unknown future events. Of course, contingent commodities span a wide space, from buying stock in your company to betting on the Super Bowl. Roughly speaking economists have articulated three reasons that trade in contingent commodities is socially useful. First, trade in financial assets can yield efficient risk sharing and pooling by matching risky assets with risk-acceptant investors. Most trade in financial markets falls into this category. Second, holding or trading risk may be intrinsically enjoyable, and this provides the rationale for sports betting and other gambling markets. And third, market prices aggregate dispersed information that may help predict future events. Markets designed specifically around this information aggregation and revelation motive are our focus in this article.

The article by Rhode and Strumpf in this volume shows that while this taxonomy of markets organized around risk, fun, and information aggregation is useful, the distinctions are easily blurred. In their case, markets organized around gambling provide useful information. Equally, if prediction markets develop sufficient liquidity, they may also prove useful for those wishing to hedge against specific risks, as envisioned by Athanasoulis, Shiller, and van Wincoop (1999) and Shiller (2003).

The Iowa Electronic Market is probably the best known prediction market. The original Iowa experiment, run in 1988, allowed trade in a contract that would pay 2½ cents for each percentage point of the popular vote won by candidates Bush, Dukakis, or others. It should be immediately clear that these prices are likely to be useful predictors of the eventual vote share of each candidate. Indeed, a key intellectual force behind the enthusiasm for prediction markets derives from the efficient markets hypothesis. If these markets are efficient, then not only will market prices provide useful indicators of the unknown election outcomes, they provide the single best indicator of likely outcomes. Put more strongly, in an efficient prediction market, no combination of available polls or other information can be used to improve on the market-generated forecasts. (An aside: aggregate rationality is easily reconciled with individual irrationality, as long as the marginal trade is motivated by rational traders.) Naturally the efficiency of any particular market is an empirical matter, but early successes have generated substantial interest among both political economists and financial economists, and stimulated an intriguing mix of theoretical, experimental and field research.

We begin by describing the types of contracts that might be traded in the next section, before proceeding to survey several applications. We then draw together a rough and fairly optimistic description of what we have learned from early experiments, raise some market design issues, and conclude with some evidence on the limitations of prediction markets.

2. Types of Prediction Markets

Prediction markets are simply markets in which payoffs are tied to unknown future events. Naturally there exist many ways to tie future events to financial payoffs, and careful design can be used to elicit the market's expectations of a range of different parameters.

Table 1: Contract Types: Estimating Uncertain Quantities or Probabilities

Contract	Details	Example	Reveals market expectation of…	More general application
Winner-takes-all	Contract costs $\$p$ Pays \$1 iff event y occurs Bid according to value of $\$p$	Event y: Al Gore wins the popular vote	Probability that event y occurs, $p(y)$	Defining many events, $y_1, y_2, .., y_n$ reveals $F(y)$
Index	Contract pays $\$y$.	Contract pays \$1 for every percentage point of the popular vote won by Al Gore	Mean value of outcome y: $E[y]$	Contract pays some function of y: $\$g(y)$. Reveals $E[g(y)]$
Spread	Contract costs \$1 Pays \$2 if $y>y^*$ Pays \$0 otherwise. Bid according to the value of y^*.	Contract pays even money if Gore wins more than $y^*\%$ of the popular vote.	Median value of y.	\$1 contract pays $\$(1/q)$ if $y>y^*$. Reveals $F_{1-q}(y)$.

Table 1 summarizes the three main types of contracts, linking payoffs to whether a specific event occurs (the incumbent wins the election), to a continuous variable (the vote share of the incumbent), or a combination, such as in spread betting where traders differentiate themselves by bidding on the cutoff that determines whether an event occurs (the incumbent garners more than y^* votes). Winner-take-all markets yield prices that represent the market's expectation of the probability that an event will occur.[1] When the

[1] The price of a winner-take-all security is essentially a state price, which will equal an estimate of the event's probability under the assumption of risk neutrality. The sums wagered in prediction markets are

unknown is a quantity, rather than a binary outcome, an index contract whose payoff is linear in the outcome will yield insight into the expected mean. "Spread betting", where the price is fixed, but the two sides of the market bid according to the terms of an even money bet (e.g., a point spread in football), can yield the market's expectation of the median, instead. In each case, the relevant contract will reveal the market's expectation of a specific parameter: a probability, mean, or median, respectively.[2]

Beyond eliciting the market's expectations about a specific outcome, such as the incumbent's vote share, prediction markets can also be used to evaluate uncertainty about these expectations. Within financial markets, this role is played by options. With prediction markets, only very small variations on the very simple structure of event contracts described in Table 1 are required to yield insight into the level of uncertainty. For instance, we might replace the single index market tied to the incumbent's vote share with a family of contracts that pay off if and only if the candidate earns 48% of the vote, 49% and so on. This family of winner-take-all contracts will then reveal the entire probability distribution of the market's expectations. A family of spread betting contracts can yield similar insights: just as a $1 under/over contract will elicit expectations of the vote share that is as likely to be an underestimate as an overestimate, a contract that costs $4 and pays $5 if $y>y^*$ will elicit a value of y^* that is a four-fifths probability to be an over-estimate, thus identifying the 80th percentile of the distribution. As a final alternative, non-linear index contracts can yield insight into higher-order moments of the distribution of the index. For instance, if we trade both an index contract that pays according to the square of the index, y^2, and a more standard linear contract, then market prices will reveal the market's expectation of $E[y^2]$ and $E[y]$, which can be used to make an inference about the market's beliefs regarding the standard deviation of $E[y]$, more commonly known as the standard error. (Recall that the standard deviation can be expressed as $\sqrt{(E[y^2]-E[y]^2)}$, or the square root of the mean of the squares less the square

---

typically small enough that assuming that investors are not averse to the idiosyncratic risk involved seems reasonable. But if the event in question is correlated with investors' marginal utility of wealth, then probabilities and state prices can differ. In what follows, we leave this issue aside and use the term probability to refer to risk-neutral probability.

2 There is a subtle, an almost metaphysical question here: What is the "market's" expectation anyway? Throughout we will speak as though the market is itself a representative person, and that "person" has a set of expectations. Consequently there are important but subtle differences between parameters such as the market's median expectation, and the median expectation of market participants.

of the means.) By the same logic, even more complicated index contracts can yield insight into higher order moments of the distribution.

3. Applications and Evidence

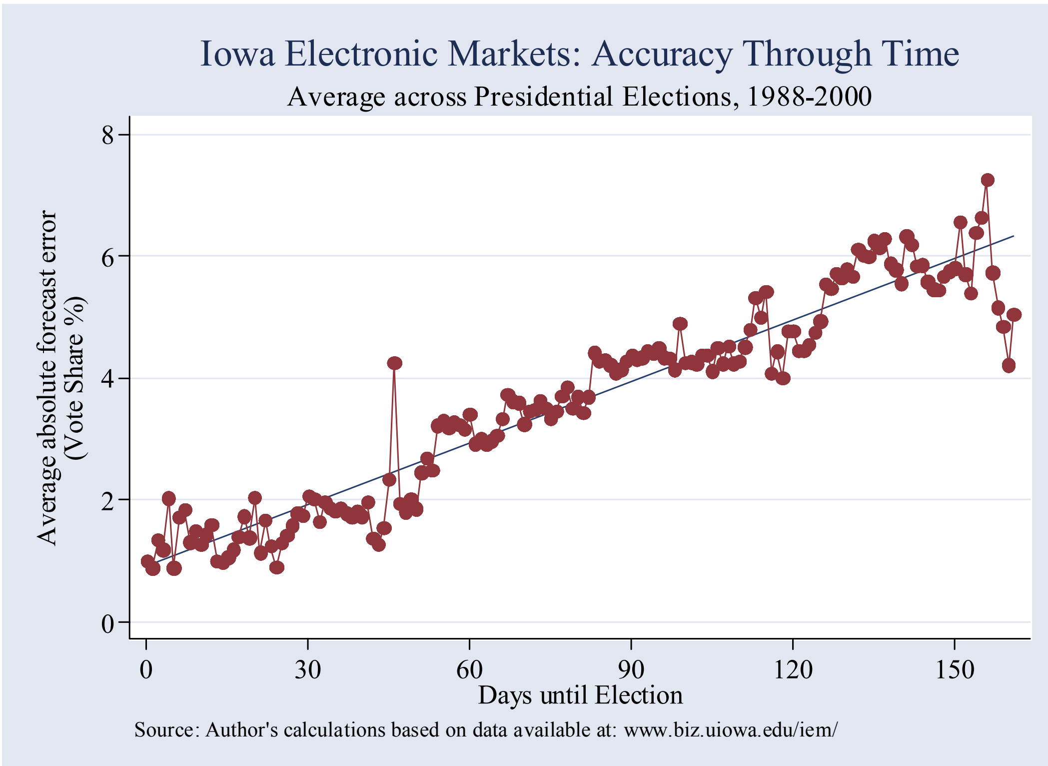
While we are still accumulating evidence on the behavior of prediction markets in different contexts, there are already a few generalizations that can be drawn.

First, market prices tend to respond rapidly to new information. The following anecdote provides an interesting example. In mid 2003 DARPA, the research thinktank within the Department of Defense, proposed a Policy Analysis Market, so as to allow trading in various forms of geopolitical risk. Proposed contracts were based on indices of economic health, civil stability, military disposition, conflict indicators and potentially even specific events. A political uproar followed, as some of the proposed contracts were perceived to be in poor taste. The uproar ultimately led to calls for the head of DARPA, Admiral John Poindexter, to resign. Immediately an offshore exchange, Tradesports.com, listed a new security that would pay $100 if Poindexter was ousted by the end of August 2003. Early trading suggested a likelihood of resignation by the end August of 40 percent, and the price rose as it became clear that pressure was building. Early rumors of his resignation surfaced on NBC on July 30, although it appears that these weren't credible. Around lunchtime on the 31st, reports started citing credible Pentagon insiders who claimed knowledge of an impending resignation. Within minutes of this news first surfacing (and hours before it became widely known), the price spiked to around 80. These reports left the date of Poindexter's proposed departure uncertain, which explains the remaining risk. As August dragged on without a resignation eventuating, the price slowly fell back toward 50. Poindexter then issued a letter of resignation, dated August 12th, suggesting that he would resign on August 29. On the 12th, the market rose sharply, to a price of 96 by the end of the trading session.

This anecdote is simply representative of a more general feature: the revelation of information is rapidly incorporated into prices, yielding more accurate predictions. Figure 1 illustrates this point more formally, aggregating data from the vote-share (or index) contracts that the Iowa Electronic Markets have run on each of the U.S.

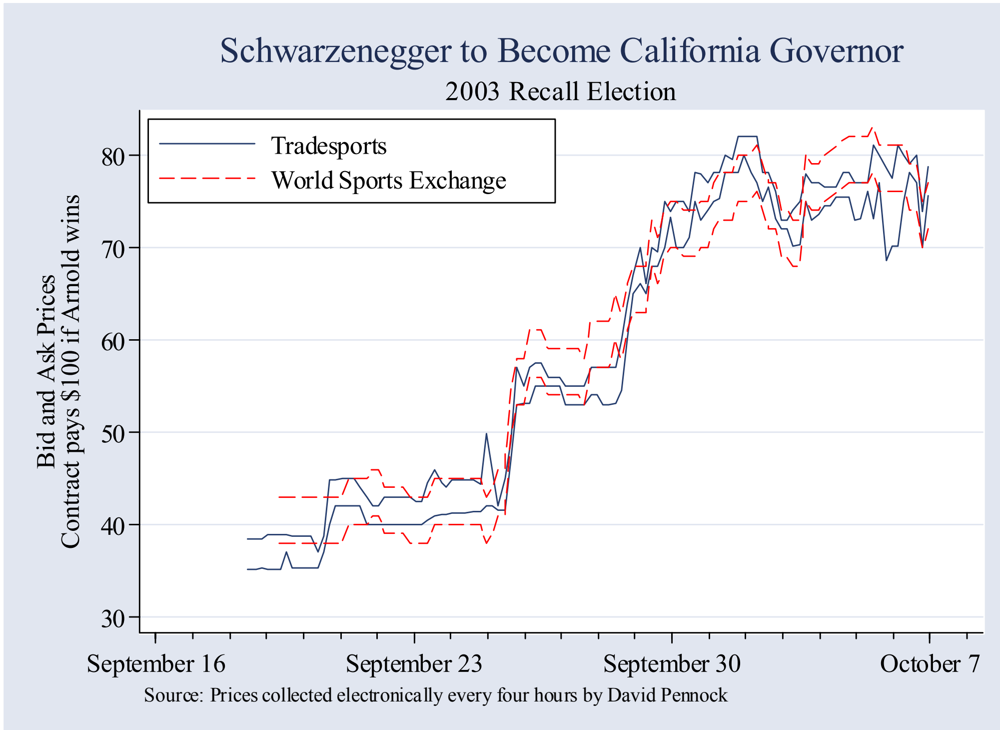
Presidential elections since 1988, showing increasingly accurate forecasts as election day approaches.

Figure 1: Information Revelation Through Time



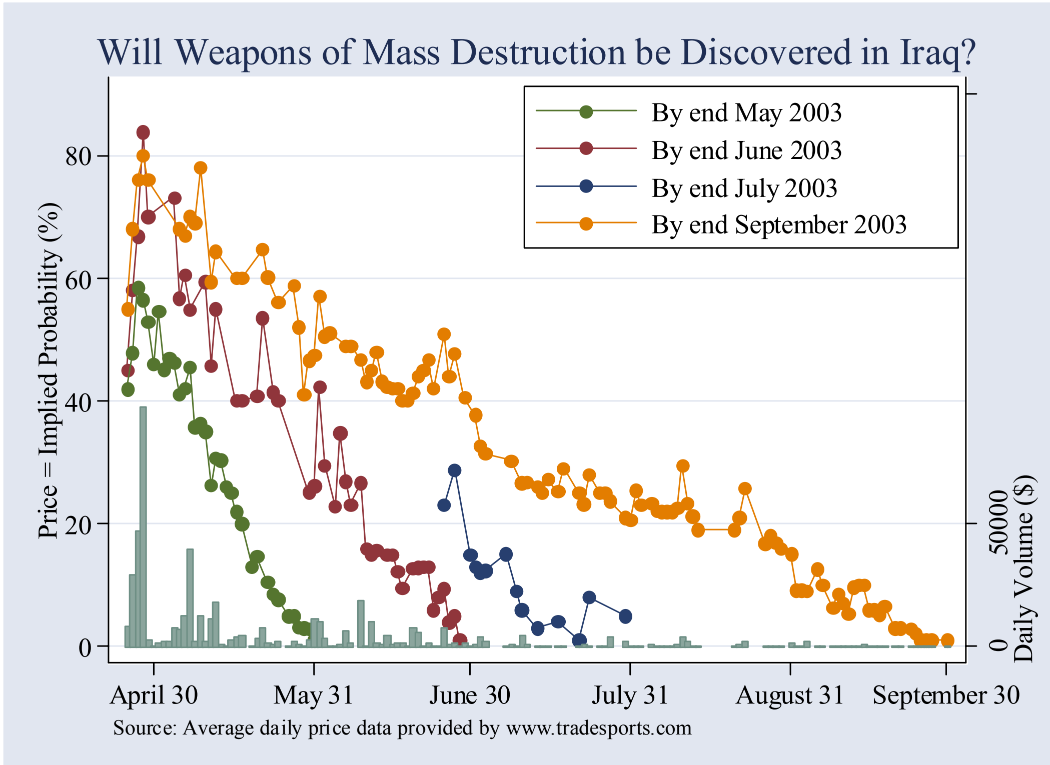
Second, the law of one price appears to (roughly) hold. There appear to be few arbitrage opportunities in these markets, and those that show up are fleeting, and involve only small potential profits. Figure 2 shows the bid and ask prices on a contract that paid \$100 if Schwarzenegger was elected California's Governor in 2003, sampling data on bid and ask prices from two online exchanges every four hours. While both sets of prices show substantial variation, they co-move very closely, and opportunities for arbitrage (when the bid price on one exchange is higher than the ask on another), are virtually absent.

Figure 2: 2003 California Gubernatorial Election



Further, in those instances in which there exist families of related securities, there tends to be an internal consistency in their pricing. Figure 3 provides a simple example, showing the prices of two securities launched by Tradesports, asking whether weapons of mass destruction would be found in Iraq by May, June, July or September 2003. The clear comovement of the four lines indicates that the prices of each contract digested similar information at close to the same time.

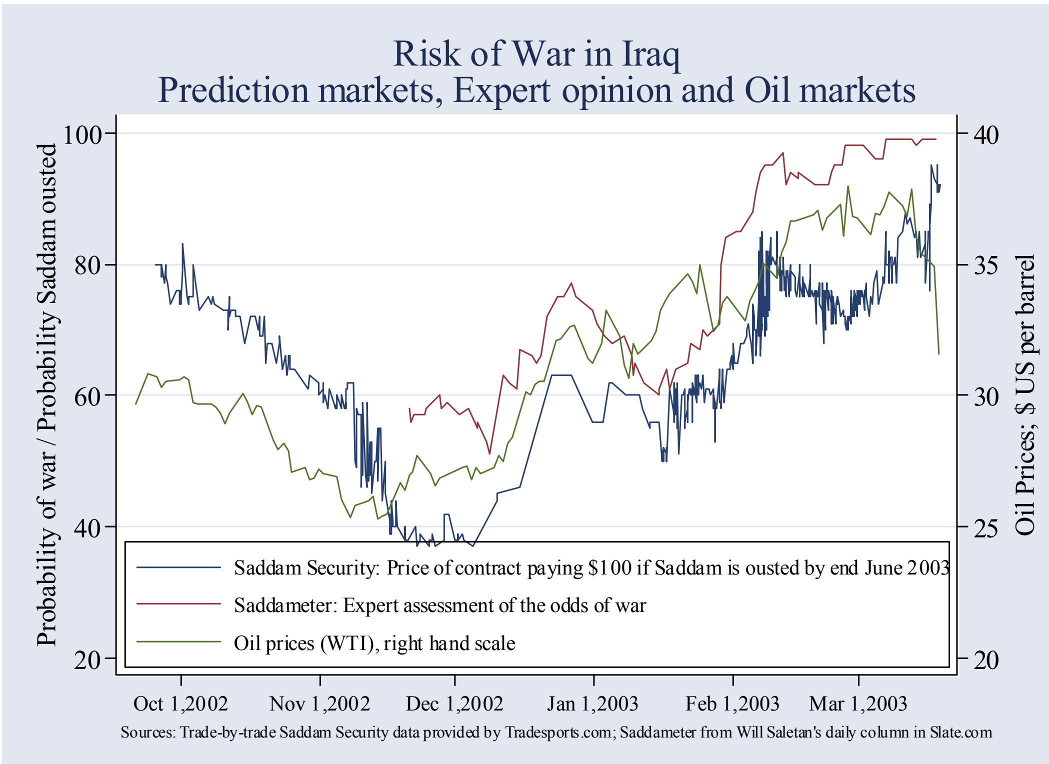
Figure 3: Will Weapons of Mass Destruction be discovered in Iraq?



The third generalization is that the profit motive has usually proven sufficient to ensure that attempts at manipulating these markets were unsuccessful. Several studies are salient here. Rhode and Strumpf document attempts at manipulation in early twentieth century political markets was typically unsuccessful. Wolfers and Leigh (2002) report candidates betting on themselves at long odds in order to create a "buzz", while Strumpf (2003) placed random \$500 bets on the Iowa Electronic Markets to trace their effect. Camerer (1996) attempted to manipulate pari-mutuel betting on horse races by canceling large wagers at the last moment. None of these attempts at manipulation had much of a discernible effect on prices, except during a short transition phase.

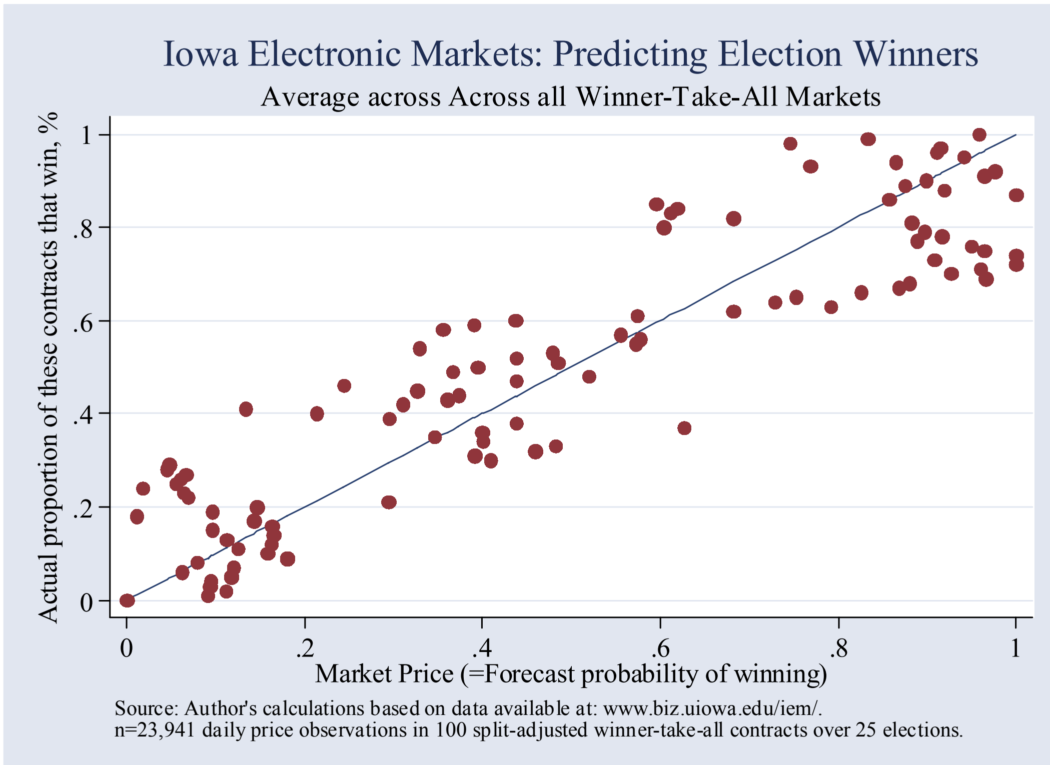
Fourth, in most cases, the time series of prices in these markets does not appear to follow a predictable path and simple betting strategies based on past prices appear to yield no profit opportunities. Figure 4 shows a specific example: the "Saddam Security", a contract listed by Tradesports that paid \$100 if Saddam were ousted by the end of June 2003. Leigh, Wolfers and Zitzewitz (2003) showed this security's prices met the standard definition of weak-form efficiency. (Rhode and Strumpf provide evidence on this point for early 20th century political markets.) Equally, there is some evidence that this small-scale market responded to news about Iraq with a slight lag relative to deeper financial markets.

Figure 4: The Saddam Security



Arguably the most important issue with these markets is their performance as predictive tools. The evidence of this is as varied as the imagination of experimenters who have applied these markets in a range of domains. In the political domain, Berg, Forsythe, Nelson, and Reitz (2001) summarize the evidence from the Iowa Electronic Markets, documenting that the market has both yielded very accurate predictions, and also outperformed large-scale polling organizations. Figure 5 shows the aggregate forecast performance of all these experimental markets (or at least those for which data is publicly available). Over the past four Presidential elections, these markets have predicted two-party vote shares with an average absolute error of only 1.4 percentage points. By comparison, the Gallup poll yielded forecasts that typically erred by 2.4 percentage points.

Figure 5: Prediction in the Iowa Electronic Markets



Perhaps more surprising—in terms of information aggregation—is the performance of markets at the level of the individual district. Typically districts are sufficiently small that there is little interest (or funding) for local polling, yet when Australian bookmakers started betting on district-level races, Wolfers and Leigh (2002) report that they were extremely accurate.

That said, taking opinion polls to be forecasts of vote share may not yield a particularly sophisticated counterfactual. A more relevant starting point might be to compare the predictions of markets with those of independent analysts.

Again, the Saddam Security provides an instructive example, and Figure 4 shows that the price of this contract moved in lockstep with both expert opinion (Will Saletan's "Saddameter" – his estimate of the probability of the US going to war with Iraq), and with oil prices, an obvious barometer of political strife in the Middle East.

In a corporate context, Chen and Plott (2002) document that an internal market at Hewlett-Packard produced more accurate forecasts of printer sales than the firm's internal processes. Ortner (1998) turns to project planning, and in his experiment an internal market predicted that the firm would definitely fail to deliver on a software project on time, even when traditional planning tools suggested that the deadline could be met. Pennock, Lawrence, Giles, and Nielsen (2003) show that the Hollywood Stock Exchange predicts opening weekend box office success with useful accuracy, and is about as accurate at forecasting winners as an expert panel.

Another recent innovation, markets in "economic derivatives," provide a useful comparison between expert opinion and market-based predictions. These new markets provide a market mechanism in which traders attempt to predict the likelihood that economic data released later in the week will take on specific values. (The Chicago Mercantile Exchange is planning to open a market in inflation futures to supplement these markets in non-farm payrolls, retail trade, and the ISM purchasing managers' index). The traditional approach to aggregating forecasts is to simply take an average from a survey of fifty or so professional forecasters. This average is often called the consensus estimate. We now have data from the first year of operation of these markets, and Table 2 analyzes these early outcomes, comparing market and consensus forecasts.

Table 2: Predicting Economic Outcomes:

Comparing Market-Aggregated Forecasts with Consensus Surveys

	Non-Farm Payrolls (Montlhy change, '000s)	**Retail Trade (ex Autos) (Montly change, %)**	**ISM Manufacturing Purchasing Managers' Index**
Panel A: Correlations			
Corr(Market, Consensus)	0.91	0.94	0.95
Corr(Consensus, Actual)	0.26	0.70	0.83
Corr(Market, Actual)	0.22	0.73	0.91
Panel B: Mean absolute error			
Consensus	71.1	0.45	1.10
Market (empirical)	72.2	0.46	1.07
Market (implied expectation)	65.7	0.34	1.58
Panel C: Standard deviation of forecast errors (Standard error of forecast)			
Consensus	99.2	0.55	1.12
Market (empirical)	97.3	0.58	1.20
Market (implied expectation)	81.1	0.42	1.96
Sample size	16	12	11

Notes: "Market" = market-implied mean forecast from www.economicderivatives.com
"Consenus" = average of around 50 forecasters from survey run by www.briefing.com
"Actual" = Preliminary estimates from original press releases (BLS, Census, ISM).

The market-based predictions of these economic indicators are always extremely close to the corresponding "consensus" forecast. Indeed, the two forecasts are so similar that there are no statistically (or economically) meaningful differences in forecast performance – measured as either the correlation with actual outcomes, or in terms of average forecast errors. That said, this early sample is sufficiently small that precise conclusions are difficult to draw.

Interestingly, these markets yield not just a point estimate for each economic indicator, but a full probability distribution. Consequently we can back out the market's implied uncertainty, measured either as the expected absolute forecast error, or the standard deviation of the market estimate. These market-based assessments of uncertainty are shown in the last line of panels B and C, respectively. Comparing these implied expectations with outcomes in the first two rows of these panels suggests that the market-based assessments of uncertainty are of about the right magnitudes.

A final interesting comparison is to compare the implied standard errors of the forecasts with the standard errors of the actual estimates that the market is attempting to forecast. For instance, the Census Bureau reports that the first reading of the change in retail trade is estimated with a standard error of around 0.5%, which is in fact larger than the average market-implied standard error of 0.42%. Taken literally, this suggests that the market believes that it is less uncertain about the Census Bureau estimate than the Census Bureau are. A similar comparison can be made for non-farm payrolls, although the inference is less direct. The BLS estimates that their final estimate of the change in non-farm payrolls has a standard error of around 64,000, while the preliminary estimate is more uncertain.[3] Comparing these numbers with the average standard error of the market forecast of 81,100 suggests that the market is about as sure of the advance estimate as the BLS. While it is easy to imagine that traders are more informed about underlying level of economic variables than the statistical agencies, it is harder to believe that the markets are more informed about the agencies' estimates. The most likely reconciliation is either that the statistical agencies are excessively cautious, issuing estimates of their uncertainty that are too high, or that traders are overconfident.

This latter interpretation is consistent with the findings of both psychologists and behavioral economists that traders tend to be overconfident. More generally, it seems likely that prediction markets will suffer many of the same biases observed in other market contexts.

[3] The BLS has yet to estimate a standard error for their preliminary estimates, but the root mean squared error of the preliminary estimate relative to the final estimate, is around 50,000. If the revision to the preliminary estimate and the subsequent error in the revised estimate were uncorrelated, this would imply a standard error for the preliminary estimate of about 81,500.

Figure 6: The Favorite-Longshot Bias

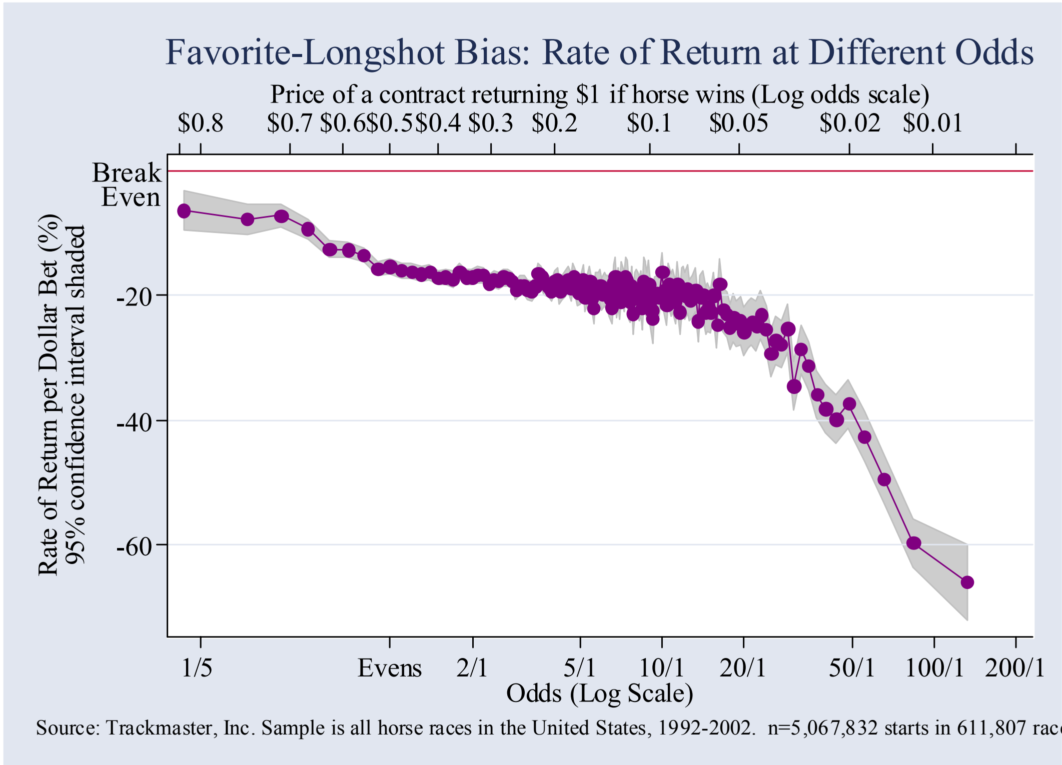


Figure 6 shows field data on precisely this point, exploiting data from a new dataset on horse racing to get a sense of the accuracy of market-based assessments of probability. On average, gamblers lose about 18 cents of every dollar wagered, and this appears to be about true for most horses—those with a 5% to 50% chance of winning. More importantly, there are substantial deviations at the extremes, with much lower returns for wagers on longshots, offset by somewhat higher (albeit still negative) returns for betting on favorites. The overbetting of longshots ties in with a range of experimental evidence suggesting that people tend to over-value small probabilities and under-value near-certainties. Thus even beyond the psychology lab, we see these errors persist in equilibrium even in large and extremely active markets.

The "volatility smile" in options (Bates, 1991 and Rubenstein, 1994) refers to a related pattern in financial markets. This phenomenon refers to over-pricing of strongly out-of-the-money options, and under-pricing of strongly in-the-money options relative to Black-Scholes benchmarks; thus the "smile" refers to the shape of the relationship between implied volatility and strike price. Aït-Sahalia, Wang, and Yared, 2001 argue that the conclusion of mis-calibration is less clear cut in this context, because these prices may be driven by small likelihoods of extreme price changes.

An example where mis-calibration appears clearer comes from the pricing of the state securities on financial variables that trade on Tradesports. Table 3 reports the price of securities that pay off if the S&P finishes 2003 in a certain range. These securities can be approximated using December CME S&P options. Comparing Tradesports prices with the state prices implied by CME option prices suggests that deep-out-of-the-money options are relatively over-priced on Tradesports, reinforcing the volatility smile in CME option prices. In the case of the most bearish securities, the price differences implied a (small) arbitrage opportunity, one which persisted for most of the summer. Similar patterns existed for Tradesports' state securities on other financial variables (e.g., crude oil and gold prices, exchange rates, other indices). This is consistent with the favorite-longshot bias being more pronounced on smaller-scale exchanges.

Table 3: Price of S&P state securities on Tradesports vs. CME
Market close, July 23, 2003

	Price on tradesports		
S&P level at end of 2003	**Bid**	**Ask**	**Estimated state price from CME S&P options**
1200 and over	2	6	2.5
1100 to 1199	11	16	13.2
1000 to 1099	28	33	33.3
900 to 999	25	30	30.5
800 to 899	14	19	13
700 to 799	3	8	5
600 to 699	4	7	2
Under 600	5	8	1
S&P level on July 23, 2003		985	

Notes: Prices given are the price of a security that pays $100 if S&P finishes 2003 in given range State prices are estimated from CME option settlement prices using the method in Leigh, Wolfers, and Zitzewitz (2003), adjusting for the 13 day difference in expiry date.

Another important limitation of financial market pricing arises from the possibility that speculative bubbles drive prices away from likely outcomes. Prediction markets seem like less fertile ground for bubbles than traditional markets, since most prediction markets place no constraint on short selling, and the small-scale nature of the markets makes it unlikely that informed investors will be capital constrained as in Shleifer and Vishny (1997). Unfortunately, it is almost impossible to make any serious attempt at describing the frequency of bubbles in the data we have so far. To take a simple example, through September 2003 we suspected a bubble in the security trading on whether Hillary Clinton would win the Democratic nomination. Our suspicions were based on her public statements that she was not a candidate, and the tenor of discussion among traders. Equally, these high prices may have reflected campaign insiders who knew more about her state of mind than we did. Empirically, the best that we can say is that the performance of past markets at predicting the future has been, on average, pretty good, and that observation holds whether or not specific markets were distorted by bubbles.

Laboratory experiments hold out the possibility of learning more about bubbles, as it is possible for the experimenter to know the "true price", and hence to observe deviations. Plott and Sunder (1982 and 1988) have set up extremely stylized examples in which bubble-like behavior occurs in simple prediction markets. At the same time, bubbles in experimental markets often burst and give way to more rational pricing.

4. Market Design

While we have described the broad contours of prediction markets, there remain several important design issues regarding their implementation. First, there is the issue of the market mechanism that matches buyers to sellers. In most cases a continuous double-auction has been used, with buyers submitting bids and sellers submitting asking prices, and with the mechanism executing a trade whenever the two sides of the market reach a mutually agreeable price. Pari-mutuel systems represent an alternative mechanism and are the dominant form of trading on horse racing. We mention this because the new market in economic derivatives operates under something close to a pari-mutuel system. Beyond these mechanisms, most sports bets are placed with bookmakers who post prices, while many market mechanisms are augmented by market-makers. Finally, while these mechanisms are relatively useful for simple markets, Hanson (2003) has proposed the use of market scoring rules to allow for simultaneous predictions over a range of outcomes.

Beyond market design, careful contract design is crucial. Contracts must be designed so that it is possible for either side of the contract to win, which explains why we don't see contracts like "Weapons of Mass Destruction are not in Iraq", but rather contracts specifying a date by which they must be found. Contracts must be clear, easily understood, and easily adjudicated. This requirement turns out to be non-trivial: the Iowa markets proposed what looked to be a well-specified market on the 1994 Senate election, with contracts paying according to the number of seats won by each party. The day after the election (and while votes were still being counted in some jurisdictions), Senator Richard Shelby (D-AL) switched sides to become a Republican. As another example, Ortner (1998) ran a market on whether a software project would be delivered to the client on schedule, only to have the client change the deadline.

Contracts must be enforceable, and have a clear adjudication mechanism. While we are not altogether sure that there is a meaningful economic difference between gambling and trading, the former is banned in many jurisdictions. These legal restrictions have led groups like NewsFutures.com to adopt play money exchanges, with those who amass the largest play-fortunes eligible for prizes. Prices on play and real-money exchanges are not linked by arbitrage: in August 2003, Bush was a 2-to-1 favorite to win reelection on real-money exchanges, but was even-money on NewsFutures. We do not yet have enough useful comparisons of play-money and real-money exchanges, and it remains an open empirical question as to whether money matters. (Indeed, it is plausible that the play money exchanges may even outperform real-money exchanges, because "wealth" can only be accumulated through a history of accurate prediction).

Practical limits in the types of contracts that can be written and enforced can turn out to be quite important in limiting the scope for prediction markets, although the "play money" exchanges, such as Foresight Exchange, are naturally in the best position to push the envelope. One often sees quite loosely worded "contracts" such as that a "scientific study will conclude that astrology is a statistically significant predictive method to describe an individual's personality traits."

Finally, even the well designed markets will fail unless a motivation to trade exists. (The failure of the inflation futures market on the Coffee, Sugar, and Cocoa exchange in the mid 1980s is a case in point.) Trade will be motivated by risk concerns if the prediction futures markets are deep enough to allow one to hedge against specific risks. Alternatively, the "play money" exchanges and sports gambling industry both suggest that it may be possible to motivate (small-scale) trading simply through the thrill of pitting one's judgment against others. In the absence of these motivators trade will occur when two parties perceive a profit opportunity from the trade because they disagree about likely outcomes. Disagreement is unlikely among fully rational traders with common information and priors. It is more likely when traders are overconfident in the quality of their private information or their ability to process public information or when they have priors that are sufficiently different to allow them to agree to disagree.

This suggests some basic design principles that can help in formulating successful markets. Securities on events that are already widely discussed are more likely to succeed than those trading on obscure statistics, since trading on them will have higher entertainment value and there will be more information on whose interpretation traders can disagree. Ambiguous public information may be better in motivating trade than private information, especially if the private information is concentrated. It is well known that information asymmetries can lead to markets unraveling, with highly informed traders driving out the partly informed, repressing trade to the point that the market barely exists. Indeed, attempts to set up markets on topics where there are insiders with substantial information advantages have typically failed. For instance, the Tradesports contracts on the next Supreme Court retirement have generated very little trade, despite the inherent interest in the question.

Trade might also be discouraged by the possibility that those with influence over the event may create their own inside information, by establishing a position in the market and then acting accordingly. This is a concern in sports, where athletes are occasionally accused of "point shaving" or losing intentionally. Similarly, it was feared that the DARPA markets would create the opportunity for a terrorist to front run and thus profit from an assassination. With respect to the DARPA markets, this concern may have been misplaced, since the proposed markets were small scale enough that terrorists would not have been able to earn much relative to the presumed going rate for an assassination. Furthermore, much larger opportunities to front run terrorist events exist in traditional financial markets, as was noted after September 11th.

Finally, since the power of prediction markets comes from the aggregation of disperse information, these markets are unlikely to perform well when there is little useful intelligence to aggregate, or when public information is selective, inaccurate, or misleading. Weapons of mass destruction may yet be found in Iraq, but as of this writing, these markets appear to be an example. Since WMD can be non-existent almost everywhere and yet still exist, disperse information about their non-existence was unlikely to overturn the strong case made by the White House, at least initially.

5. Making Inferences from Prediction Markets

Having carefully designed a market to elicit expectations about a certain outcome, how might we then use them in subsequent analysis?

The most direct form of inference involves simply using these predictions directly. For instance, in their experiments at Hewlett Packard, Chen and Plott elicited expectations of future printer sales. These expectations are likely of direct interest for internal planning purposes.

Some analyses have tried to link the time series of expectations elicited in prediction markets with time series of other variables, so as to isolate a causal influence. For instance, in Leigh, Wolfers and Zitzewitz (2003), we interpreted movements in the Saddam Security as an index for the risk of war, and interpreted the comovement with the oil price shown in Figure 4 as a causal relationship, concluding that war led to a $10 per barrel increase in oil prices. A similar analysis suggested that equity prices had built in a 15 percent war discount. Applying a similar methodology, Slemrod and Greimel (1999) linked the price of a Steve Forbes security with municipal bond prices, showing that the threat of a flat tax led to a rising interest rate premium on municipal bonds.

As with any regression context, inferring that these correlation reflect causation requires one to believe that this comovement reflects neither reverse causation (developments in the oil market affecting the likelihood of a successful war in Iraq), a third omitted factor (such as political disruption elsewhere in the Middle East) nor luck (which is why we subject this to statistical testing).

The ultimate aim of this exercise is to make statements about the expected value of oil prices under conditions of either war or peace. The drawback of the time series analysis is that we can only make these inferences if periods of high and low war risk actually occur, so that oil prices at each point in time can be compared. A more direct method would simply ask market participants how they would price oil if war risk were high or low. There is a prediction market analogy to this thought experiment that involves contingent contracts. For instance, we could invent two securities, one which pays $*P* if Saddam is ousted in a year (where *P* is the future oil price), and the purchase price is refunded otherwise, and another that pays $*P* if Saddam is still in power, and again the purchase price is refunded otherwise. The difference in the equilibrium price of these two securities will be the market's expectation of the effects of ousting Saddam on oil prices. The advantage of this inference is that it does not require researchers to wait until sufficient variation in the political situation has accrued for a regression to be estimated. Moreover, it may be that the market's assessment of the oil price impacts of war shift through time, and these changing beliefs can be directly measured through such a conditional market.

Thus far very few of these contingent markets have been constructed, although this year's Iowa Electronic Market on the 2004 Presidential election is instructive. Table 4 shows the prices of a series of contracts linked to the two-party vote share of each Democratic hopeful; there exist an equal number of securities linked to the vote share of President Bush, conditional on President Bush facing each specific candidate. Unlike the contingent contract described above, these contracts pay zero if the specific matchup does not occur.

Table 4: Contingent Markets

Candidate	Candidate Vote Share *A*	Bush Vote Share given this Candidate *B*	Prob. this Candidate wins Nomination *A+B*	Expected Vote Share if Nominated *A/(A+B)*
Howard Dean	30.5%	27.9%	58.4%	52.2%
Wesley Clark	8.3%	8.5%	16.9%	49.3%
Richard Gephardt	7.0%	5.0%	12.1%	58.1%
John Kerry	4.3%	4.8%	9.1%	47.4%
Hillary Clinton	3.2%	3.1%	6.3%	50.3%
Joe Lieberman	1.0%	1.1%	2.1%	47.6%
Other Democrats	1.6%	1.6%	3.2%	49.6%

Source: Volume-weighted average prices 11/1/2003 – 11/16/2003, Iowa electronic markets.

Because the expected Democratic and Republican shares of the two-party vote must sum to one no matter who the candidate is, adding the prices of the securities shown in columns A and B yields the probability that each candidate wins the Democratic nomination (shown in column C).

The final column calculates the implied expected vote share for each candidate, conditional on that candidate winning the nomination. Robin Hanson (1999) has suggested that these expectations should be used to guide decision making. As such, delegates to the Democratic convention interested in selecting the strongest candidate would simply compare the ratios in the final column, and accordingly vote for Gephart. Thus, these conditional prediction markets might instead be called prediction markets.

While we are optimistic that these data can be used to inform decision-making, some care is required. In making statements about the comovement of two variables social scientists have long struggled with distinguishing correlation from causation, and these decision markets do not resolve this issue. For instance, the reason that the Gephardt security is trading at a much higher price than Gephardt|Bush, might be that he is perceived as a thin-skinned candidate, who simply will drop out of the nomination race if President Bush looks strong in the polls. Alternatively the markets may believe that Gephardt won't win the nomination unless the country makes a dramatic shift to the left, but that if this does happen, it is likely that Gephardt will win both the nomination and the election. Or finally, the market might be entirely unsure as to whether Gephardt is a superstar or a dud, but they believe that the Democratic convention knows the truth, and Gephardt will only win the nomination if he is in fact a star.

The analogy with the usual regression problem of distinguishing correlation with causation is fairly direct. One could imagine that traders hold a frequentist view of probability and that they price the securities in Table 4 by simply inventing a hundred (or more!) possible scenarios, and prices simply reflect average outcomes across these scenarios. Equally, the traders could code the results of these hundred imagined histories and pass them on to an econometrician to analyze. The econometrician would note a robust correlation between Gephardt winning the nomination and the Democrats winning the presidency. But a careful econometrician would be reluctant to infer correlation, noting that there are important "selection effects" at play, as the scenarios in which Gephardt wins the nomination are not random.

With some of these selection effects, there is a ready solution: just as an econometrician uses a selection model to correct for selection bias (Heckman, 1979), we could add another market price or contingency reflecting the variable driving the selection of Democratic candidates. Thus adding a contract that pays off if a candidate drops out of the nomination race early would allow a somewhat more fine-grained assessment of candidate ability. But to the extent that many of these selection effects operate on unobservable characteristics, it may impossible to ever fully rely on decision markets to guide voters to the candidate with the greatest vote-winning potential.

Thus, we return to the point that we started with: prediction markets are extremely useful for estimating the market's expectation of certain moments; simple market designs elicited expected means or probabilities, more complex markets elicited variances, and contingent markets can be used to elicit the market's expectations of covariances and correlations (and hence market perceptions of specific regression coefficients), although as with any estimation context, further identifying assumptions are required before a causal interpretation can be made. These simple contingent markets, as well as more complex combinatorial markets, are as yet virtually untested, and a useful focus for further research. There may be important and interesting applications in domains where selection problems are absent.

6. The Future: Innovative Applications

The research agenda on these markets has reflected an intriguing interplay between theory, experiments, and field research, drawing on scholars from economics, finance, psychology and computer science. This research program has established that prediction markets provide three important roles, providing incentives to seek information, incentives for truthful information revelation, and an algorithm for aggregating diverse opinions. Of course there exist both a much wider range of incentive mechanisms for truthful revelations, and any number of ways of aggregating opinions. The literature on prediction markets has yielded an active research program within experimental economics that has attempted to evaluate some of these alternative mechanisms. Yet while this research holds the promise of establishing even more effective institutions, it nonetheless remains the case that practice lags, and relatively primitive mechanisms, such as the ubiquitous staff meeting, are still remain the most widely used mechanisms within organizations.

We believe that the early evidence on the forecasting ability of prediction markets is mostly – although not uniformly – strong. Equally we have suggestive evidence from a number of domains that human foibles distort market prices. Thus perhaps the best conclusion is that in specific circumstances, and with careful design, prediction markets have substantial potential to help organizations make better informed decisions. We are already seeing increasing interest in these markets in the private sector, with early experiments at HP now being supplemented with new markets on pharmaceuticals and the likely success of future technologies.

Equally, while DARPA's ill-fated attempt at establishing a Policy Analysis Market ultimately failed, it seems likely that private-sector firms will continue to innovate, and policymakers will still be able to turn to prediction markets run by innovative firms like Tradesports, Net Exchanges and NewsFutures. That these exchanges will be run by publicly-regulated, private-sector firms likely reflects a sensible political outcome. Nonetheless, to the extent that the benefits from the information generated by trade in these markets is not fully internalized by these private firms, prediction markets are also likely to be underprovided.

6. References

Aït-Sahalia, Yacine, Yubo Wang, and Francis Yared (2001), "Do options markets correctly price the probabilities of movement of the underlying ssset?" *Journal of Econometrics*, 102, 67-110.

Athanasoulis, Stefano, Robert Shiller, and Eric van Wincoop (1999), "Macro markets and financial security," *Economic Policy Review*, 21-39.

Bates, David (1991) "The Crash of '87: Was It Expected? The Evidence from Options Markets," Journal of Finance 46(3), 1009-44.

Berg, Joyce and Thomas Rietz (2003), "Prediction Markets as Decision Support Systems", *Information Systems Frontiers*, 5(1).

Berg, Joyce, Robert Forsythe, Forrest Nelson and Thomas Rietz (2001) "Results from a Dozen Years of Election Futures Markets Research," forthcoming in Charles Plott and Vernon Smith (Eds) *Handbook of Experimental Economic Results*.

Chen, Kay-Yut and Charles Plott (2002), "Information Aggregation Mechanisms: Concept, Design and Field Implementation for a Sales Forecasting Problem."

Cramerer, Colin (1996) "Can Asset Markets be Manipulated? A Field Experiment with Racetrack Betting," Cal Tech working paper.

Hanson, Robin (1999) "Decision Markets," *IEEE Intelligent Systems*, 14(3), 16-19.

Hanson, Robin (2003) "Combinatorial Information Market Design," Information Systems Frontiers, 5(1), 105-119.

Heckman, James J. (1979) “Sample Selection Bias as a Specification Error,” *Econometrica* 47(1), 153-61.

Leigh, Andrew, Justin Wolfers and Eric Zitzewitz (2003), “What do Financial Markets Think of War in Iraq?”, *NBER Working Paper 9587.*

Leigh, Andrew and Justin Wolfers (2002), “Three Tools for Forecasting Federal Elections: Lessons from 2001”, *Australian Journal of Political Science* 37(2).

Ortner, Gerhard (1998), “Forecasting Markets – An Industrial Application”, *mimeo*, Technical University of Vienna.

Pennock, David, Steve Lawrence, C. Lee Giles, and Finn Arup Nielsen (2001), “The Real Power of Artifical Markets,” *Science*, 291, 987-988.

Plott, Charles and Shyam Sunder, (1982), “Efficiency of experimental security markets with insider information: An application of rational-expectations models,” *Journal of Political Economy*, 90(4), 663-98.

Plott, Charles and Shyam Sunder, (1988), “Rational Expectations and The Aggregation of Diverse information in Laboratory Security Markets,” *Econometrica*, 56, 1085-1118.

Rubenstein, Mark (1994), “Implied Binomial Trees,” *Journal of Finance*, 49, 771-818.

Shiller, Robert (2003), *The New Financial Order: Risk in the Twenty-first Century*, Princeton University Press: Princeton, NJ.

Shleifer, Andrei and Robert Vishny (1997), “The Limits of Arbitrage,” *Journal of Finance* 52(1), 35-55.

Slemrod, Joel and Timoth Greimel (1999), “Did Steve Forbes Scare the Municipal Bond Market?”, *Journal of Public Economics*, 74(1), 81-96.

Spann, Martin and Bernd Skiera (2003), “Internet-Based Virtual Stock Markets for Business Forecasting”, *Management Science*, 49, 1310-1326.

Data Sources

Austrian Electronic Markets

UBC Election Stock Market

Iowa Electronic Market

Hollywood Stock Exchange

Foresight Exchange

www.Tradesports.com

www.Centrebet.com.au

British bookies

www.newsfutures.com

www.probabilitysports.com

www.economicderivatives.com
(We could make a table from this.)