

## 2

# Five Open Questions about Prediction Markets

*Justin Wolfers and Eric Zitzewitz*

Interest in prediction markets has increased in the past decade among participants, private-sector market operators, policymakers, and academics. Markets on the 2004 U.S. election, for instance, were far more numerous and liquid than they were in 2000. Although past media coverage often treated these markets as curiosities, coverage in the 2004 election cycle was far more frequent and more serious than four years earlier.<sup>1</sup>

Academics are using prediction markets to provide a measure of expectations about an event's probability and then using the co-movement of this measure and financial asset prices to extract information about the expected effects of political decisions,<sup>2</sup> in some cases even before the decisions are made.<sup>3</sup> In the last year, this style of analysis has spread beyond academia, most notably in attempts to analyze the consequences of George W. Bush's reelection. Formal analyses of these markets have tended to conclude that their prices can be useful indicators of likely future outcomes.<sup>4</sup> Interest in new applications of prediction markets is focused in three domains: forecasting, decision-making, and risk management.

The success of corporate prediction markets in forecasting printer sales<sup>5</sup> and project management<sup>6</sup> has stimulated interest in their application to other business problems, and firms such as NewsFutures, Net Exchange,

---

We would like to thank Brian Elliot, Brian Galebach, and David Pennock for help with data, and Robert Hahn, Marco Ottaviani, Erik Snowberg, and Paul Tetlock for helpful conversations. Wolfers acknowledges the support of a Geewax, Teriker and Co. Research Fellowship, the Mack Center for Technological Research, and the Zull/Lurie Real Estate Center at Wharton.

and Incentive Markets have sprung up to meet this demand. Among the applications being discussed are “decision markets,” in which securities are traded that pay off based on an outcome (for example, revenue from a product) if, and only if, a particular decision is made (for example, the decision to launch the product). The idea behind these markets is to elicit knowledge from within an organization that might otherwise be lost.

Similar ideas have been discussed in the policy realm. The Policy Analysis Market of the Defense Advanced Research Projects Agency (DARPA) would have launched securities designed to capture the probability of specific events, along with contingent securities designed to capture the outcome of specific policies. During the 2004 primary season, the Iowa markets ran securities designed to predict the general election success of the Democratic contenders, and, later in the year, TradeSports ran similar contracts designed to capture the effects of geopolitical and economic events on the election.<sup>7</sup> Michael Abramowicz<sup>8</sup> and Robert Hahn and Paul Tetlock<sup>9</sup> envision using prediction markets to assist in policymaking by extracting expert opinion in a credible and objective manner. Robin Hanson goes even further, arguing that policymakers should simply define a “GDP+” measure of social welfare, and base policy decisions entirely on market-based predictions of which policies maximize this measure.<sup>10</sup>

Finally, some envision prediction markets as the first step toward markets where participants could hedge their exposure to political and economic events. For example, an employee in a “Bush industry” (for example, traditional energy) could sell TradeSports’s Bush reelection contract to an employee in a “Kerry industry” (for example, alternative energy); both would be hedging their human-capital exposure to the election. The TradeSports Bush reelection contract had trade volume of more than \$15 million during the year and a half it traded, and volume of about \$3 million on election day.<sup>11</sup> Although this is extremely liquid by prediction-market standards, clearly, more liquidity would be needed to allow for meaningful hedging. Hedgestreet.com, a new CFTC-sanctioned derivatives exchange, is attempting to create markets that individuals can use for such hedging, with a focus more on economic risks such as mortgage rates and real estate prices than on politics.

Will these great expectations for prediction markets be fulfilled? This, we argue, depends in large part on developing answers to five open questions:

1. **How can markets attract uninformed traders?** Counter-intuitively, the problem for most prediction markets is attracting sufficient *uninformed* order flow. Markets need uninformed order flow to function; when trading is conducted by rational traders, whose sole motivation is expected returns, the no-trade theorem binds, and the market unravels.<sup>12</sup> Uninformed order flow can have a variety of motivations (entertainment, overconfidence, and hedging, for example), but with the exception of hedging, these are usually noneconomic, putting economists at a comparative disadvantage in predicting which markets will succeed.
2. **How should security design trade off interest and contractibility?** A fundamental problem in mechanism design is that the outcomes of interest are often impossible to write into contracts. Running financial markets on policy outcomes in and of itself does nothing to help with this problem. In fact, using a measure to set payoffs for financial contracts can turn what would otherwise be a good proxy for a policy outcome into a problematic one, as illustrated by the recent experience with the manipulation of cash settlement prices for Standard and Poor's (S&P) and municipal bond futures.
3. **How can markets limit manipulation?** In addition to manipulation of the outcomes on which prediction markets are based, one might worry about manipulation of prediction-market prices themselves, particularly where high-stakes decisions are based on the prices.
4. **Are markets well-calibrated on small probabilities?** Many of the proposed uses of prediction markets will involve the evaluation of small-probability events. A range of behavioral evidence suggests that people are quite poor at distinguishing small probabilities from tiny ones, and even when arbitrage is possible, frictions can cause this miscalibration to carry over into market prices.
5. **How can analysts separate correlation from causation?** Contingent prediction markets allow us to estimate the probability

of an event contingent on another event occurring. Thus, contingent markets provide insight into the correlation of events. Determining whether one event *causes* the probability of another to change is a separate and potentially more important question, however. Many of the proposed uses of decision markets presuppose that the direction of causality can be readily established.

### A Framework

Before discussing these questions, we will introduce a simple model that can serve as a common framework for thinking about most of them. We do this by taking the basic setup from the familiar Kyle model<sup>13</sup> and adding transaction costs, which we believe may be an important factor shaping the efficiency of pricing in small-scale prediction markets, and market-maker risk aversion, which allows uninformed order flow to affect prices.

As in Kyle, we consider three types of agents: perfectly informed traders, uninformed (noise or liquidity) traders, and perfectly competitive market-makers. Trade is in a binary prediction-market security that pays  $y = \$1$  if an event occurs and  $y = \$0$  otherwise.

The probability that the event occurs is given by  $q$ ; this probability is observed by the market-makers. The perfectly informed traders have inside information and know whether the event will occur. The uninformed traders have a noisy subjective expectation of the event probability given by  $q$  plus a noise term  $\eta$ . All traders have log utility and trade to maximize their subjective expected utility. As such, they take positions:

$$x = \frac{w}{p(1-p)}(e - p),$$

where  $w$  is their wealth,  $p$  is the price of the security,  $e$  is their subjective expectation that the event occurs, and these parameters yield demand for  $x$  prediction securities.<sup>14</sup>

In addition, we can allow the uninformed traders to derive a direct utility benefit from holding a particular position, perhaps for gambling or entertainment reasons. If this direct benefit has a per-unit certainty equivalent of  $g$ , then the uninformed trader will trade as if  $e = q + \eta + g$ .

Alternatively, if the market price is being used in a decision, traders may have other (external) reasons to trade in order to affect the price. As such,  $g$  may represent the gains to market manipulators from their effect on the equilibrium price. Thus, for manipulators,  $g$  would be equal to the product of the marginal price impact of their trading and the outside benefit they receive from moving the price.

This framework also lends itself quite naturally to considering a hedging motive for trade. Risk-averse traders have a hedging demand when their wealth depends on whether the event occurs. We can view this as beginning trading with an endowment of contracts, and optimal hedging simply involves deciding how much of this endowment to sell at the market price. In this setup, if the trader's wealth is  $H$  lower if the event happens, the trader will trade as if  $e = q + \eta + g + H^*p^*(1-p)/w$ .

So, to summarize, traders trade based on the sum of objective probability ( $q$ ), expectation errors ( $\eta$ ), gambling and manipulation motives ( $g$ ), and a hedging motive related to the risk in the endowment of traders ( $H$ ). These different beliefs and motives are summarized in the variable  $e$ , which is drawn from the distribution  $F(e)$ .

If traders face a per-contract transaction cost,  $t$ , they will buy if  $e - t > p$  and sell if  $e + t < p$ .

A competitive group of market-makers will post bids and offers such that the marginal utility of an additional trade, if made, is zero. These conditions are:

$$P_{bid} = E(y|e + t < P_{bid}) - t_{MM} - cx_{MM}$$

$$P_{ask} = E(y|e - t > P_{ask}) + t_{MM} - cx_{MM}$$

where  $c = p(1-p)/w$  captures the risk-averse market-makers' desire to shade prices lower when they built up a large exposure ( $x_{MM}$ ) in this market.

With no informed traders, the market-makers' *expectations* do not change with the order flow, so  $P_{bid} = q - t_{MM} - cx_{MM}$  and  $P_{ask} = q + t_{MM} - cx_{MM}$ . (Note, however, that due to their desire to limit exposure, market-makers will change *prices* in response to new orders.) Traders only buy or sell when their subjectively held beliefs differ from the market-maker's prices by at

least the sum of their own and the market-maker's transaction costs (when  $|e - (q - cx_{MM})| \geq t + t_{MM}$ ). The equilibrium bid-ask midpoint must satisfy:

$$p_{mid} = q + c[X_{buy}(p_{mid} + t + t_{MM}) + X_{sell}(p_{mid} - t - t_{MM})]$$

$$X_{buy}(p) = \int_p^1 x(e - p) \cdot dF(e)$$

$$X_{sell}(p) = \int_0^p x(e - p) \cdot dF(e)$$

where  $X_{buy}$  and  $X_{sell}$  are total demand from buyers and sellers, respectively.

Transaction costs cause trading to be done entirely by the traders with the most noisy observations of the event probability or, alternatively, with the greatest external motivations (gambling, hedging, or manipulation) for holding the securities. The market-maker's observation of the objective probability does help hold market prices close to their efficient levels ( $q$ ), but market-maker risk aversion, combined with an asymmetry in the distribution of  $e - q$ , can cause prices to deviate from objective probabilities.<sup>15</sup>

Adding perfectly informed traders has two effects. The perfectly informed traders buy when  $y = 1$  and sell when  $y = 0$ , affecting market-maker inventory in a way that pushes the final bid-ask midpoint toward  $y$ . At the same time, these traders force the market-maker to add an adverse-selection component to the bid-ask spread, which causes the uninformed traders to be further restricted to those with extreme values of  $e$ .<sup>16</sup> Too much informed trade can cause markets to unravel, and this is especially likely when high transaction costs are already limiting uninformed trade.<sup>17</sup>

The discussion above suggests that low transaction costs are essential to liquid, active, and informative prediction markets. Adding a fixed pool of perfectly informed traders leads these insiders effectively to levy a tax on the trading of market-makers and traders, and greater expected volume can lower the contribution toward this fixed cost required from each trade. High- and low-liquidity equilibria may be simultaneously possible. This brings us to our first open question: Will prediction markets attract necessary uninformed trade?

### Attracting Uninformed Traders

An important implication of the model sketched above is that the success of the prediction market in generating trade depends critically on attracting uninformed traders.<sup>18</sup> Prediction markets are too small for meaningful hedging, and are likely to remain so for some time, so the primary motivations for uninformed order flow are likely to be entertainment and overconfidence. Economists may be at a comparative disadvantage in analyzing these motivations, but we can look at what has worked.

Risk-love, or the “thrill of a gamble,” provides obvious motives for uninformed traders, and both TradeSports and Betfair have successfully attracted many sports bettors to their markets. TradeSports runs a variety of nonsports markets, but the contracts that attract significant volume on TradeSports are typically those with popular currency (the presidential election, the fates of Saddam Hussein and Martha Stewart, outcomes on *American Idol* and *The Apprentice*), suggesting primarily an entertainment motive for trading. Beyond sports bettors, TradeSports has been quite successful at marketing its platform to those employed in financial markets, a pool of risk-acceptant traders.

At the same time, contracts on more “serious” topics (economic numbers, financial market outcomes, and contingent political contracts) attract volume that compares favorably with volumes on the successful Iowa vote-share markets. These traders may have been attracted more by the substantial free media attention TradeSports has received, and uninformed trading in these contracts may be more of a combination of entertainment and overconfidence motives. As the wonkishness of the contract rises, however, volume and liquidity fall rapidly. The few TradeSports markets of the sort planned by the DARPA Policy Analysis Market (for example, will there be a Palestinian state by 2005?) have not been very successful.

The legal environment has forced onshore prediction markets to make compromises that have limited their attractiveness. The Iowa Electronic Markets agreed to limit positions to \$500 in order to receive a “no action” letter from the Commodity Futures Trading Commission (CFTC).<sup>19</sup> This compromise limits the scope and depth of their markets, and possibly their efficiency. For example, the “Will Bush win the popular vote?” contracts on TradeSports and IEM often trade at levels different enough to imply

arbitrage opportunities, albeit opportunities to win only small amounts, given the position limits on the Iowa exchange.<sup>20</sup>

Other markets made other compromises. Platforms such as NewsFutures, the Hollywood Stock Exchange, and the Foresight Exchange operate using play money, albeit play money that can sometimes be converted into prizes. Many economists may be skeptical that play-money markets solve the cheap-talk problem that prediction markets are meant to solve, and, indeed, play-money and real-money market prices can and do diverge. That said, a comparison of NewsFutures and TradeSports prices for securities predicting NFL victories for the 2003 season found that although the two markets often yielded different predictions, they were about equally well-calibrated.<sup>21</sup>

The consulting firms running corporate prediction markets have taken the same approach as experimental economists on campuses: They have subsidized participation, allowing everyone to leave a winner, albeit to varying degrees. The Policy Analysis Market was also supposed to have involved a subsidy for participation. If the information to be generated by the market is important enough to justify the subsidy, and a willing financier can be enlisted, then this approach can work. It is important, however, that subsidies not be designed in such a way that they can be gamed, as one could imagine with per-trade subsidies or with some schemes for subsidization via a money-losing market-maker.

Finally, two markets that we are aware of, Economic Derivatives and Hedgestreet, have obtained regulatory approval to operate as futures exchanges. Both have avoided political contracts and have focused instead on economic events, probably to avoid being considered betting venues as opposed to financial markets. Hedgestreet has only recently started up, but Economic Derivatives, a joint venture by Deutsche Bank and Goldman Sachs that runs markets in economic numbers, has been operating since 2002. Its traders are all institutional, and its markets have attracted volumes of hundreds of millions of dollars. Traders reportedly trade for hedging reasons.<sup>22</sup> In addition, a combination of overconfidence and career concerns may provide a motivation for trading—no self-respecting economic analyst would want to admit that he knows only as much as his average colleague.

Career concerns may conceivably provide a motivation to participate in some of the business and policy applications that are being considered.

If the decision to trade is motivated by career concerns, however, the trading itself may reflect objectives other than maximizing trading profits. For instance, although rank-order tournaments can create efficient incentives in many career models, in the trading context they may lead to “doubling up,” or related high-variance strategies.

To summarize, we have suggested that three routes to attracting order flow have been successful thus far: offering sports betting, subsidization, and, possibly, exploiting career concerns. Each has its own drawbacks, however.

### **Balancing Interest and Contractibility**

Observable indicators of contractibility have thus far done little to explain the success of a prediction market. While contracts on the IEM have multiple prospectuses clearly outlining how the market will settle under a number of contingencies, contract “prospectuses” on TradeSports are often limited to the single sentence below the contract name on the trading screen. In some cases, this sentence is supplemented by a short memo announcing the contract, but even then issues that have material effect on event probabilities are only subsequently clarified in response to questions. For example, TradeSports offered a contract on whether Yasser Arafat would depart the Palestinian state by the end of 2005. When he became ill in late 2004, there was mild controversy on the TradeSports forum (<http://forum.tradesports.com>) about whether seeking medical treatment in Paris or dying was considered departing Palestine. TradeSports tightened the contract definition to include the latter but not the former.

TradeSports traders have thus far been willing to trust the exchange to make a good-faith determination, and this trust is helped because TradeSports does not itself hold positions in any of its contracts, and because it has a clear need to maintain a reputation for fairness among its traders. But as the sums traded become more substantial, traders may begin to worry more about these issues. As the example of the Arafat Security illustrates, in some cases no definition captures the event of interest perfectly.<sup>23</sup>

This issue is likely to become even more salient once contracts start trading on policy outcomes. Many policy outcomes are notoriously difficult

to reduce to a simple set of measures. To take one of the less-problematic examples, suppose one runs a prediction market on the crime rate in Baghdad. One presumably cares about the actual crime rate, but a contract is more likely to be written on the reported crime rate. The reported crime rate may have historically been a reasonable proxy for the actual crime rate, but this relationship may not hold up for several reasons. Reported crimes may be an inverted U-shaped function of true crimes—at some point, the reported rate may decline as the true rate rises. Furthermore, if individuals can influence the reported crime rate and have incentives to do so, either arising from, or independent of, the prediction market, the two rates can diverge.

These issues may also arise in business contexts. One might view a tracking stock as analogous to a prediction-market security on the future stream of a business unit's profits. Tracking stocks have been greeted with some skepticism, given the incentive of the parent company to influence the business unit's profits through devices such as transfer pricing.

### **Managing Manipulation**

Just as thinking about the contractibility of the outcomes of interest raises questions about outcome manipulation, the prospect of basing decisions on prediction markets raises questions about the potential for price manipulation.

In some of the applications that have been proposed for prediction markets, one might expect these outside interests to be the dominant determinant of traders' positions. For example, if a market were run among industry participants on which technological standard the industry should adopt, one could imagine that trading would be driven primarily by an interest in influencing the outcome. With no firm-level position limits, one could imagine such a market turning into an auction; with position limits, it could turn into voting. Either mechanism might be a satisfactory way of choosing a standard (ignoring potential antitrust concerns), but, arguably, simply running an auction would be more transparent than (and thus preferable to) running a prediction market that approximates an auction.

Manipulation can be made more expensive by allowing free entry into these markets and by providing a means for entrants to invest in becoming

informed.<sup>24</sup> Known attempts to manipulate public prediction markets have largely failed. In 1996, Pat Buchanan's supporters reportedly attempted to bid up his price on the Iowa markets; and in the 2004 election cycle, several large sales of the TradeSports Bush reelection contract temporarily pushed down its price, on September 14 from 63 to 49, and on October 16 from 53 to 10. In these cases, however, the price impact of the trades was reversed within twenty-four hours. Strumpf ran an explicit experiment of randomly placing \$500 trades on the Iowa markets and likewise found that the price impact was only temporary.<sup>25</sup>

Although free entry can offset the effects of manipulators, it may be undesirable in corporate contexts where the firm wishes to appropriate returns to the information aggregated in the market. Beyond secrecy concerns, acquiring the information to become an informed trader may be costly enough that free entry does not offset the trading activity of manipulators.

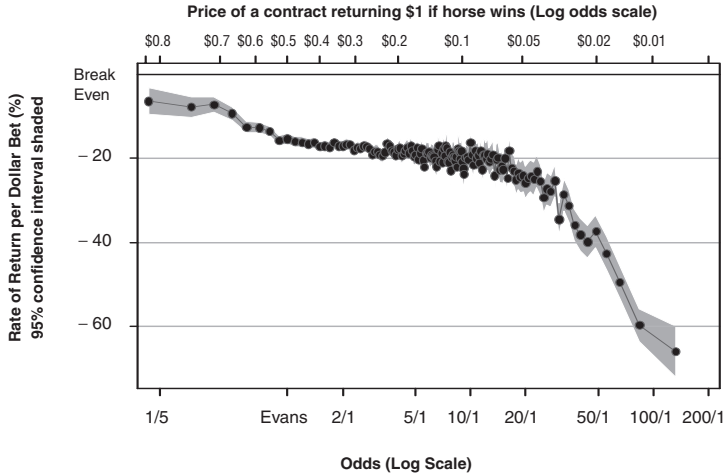
### **Ensuring Calibration on Small Probabilities**

An exception to the generally good predictive track record of betting and prediction markets is their performance on low-probability events. The best-documented example is the favorite/long-shot bias in horse racing.<sup>26</sup> Figures 2-1 and 2-2 on the next page illustrate this point, based on data from a large sample of North American horse races and from the Iowa markets, respectively. Wolfers and Zitzewitz discuss a similar mispricing from S&P 500 index contracts on TradeSports.<sup>27</sup> In both examples, market prices overestimate the probability of unlikely events. In horse racing, the overestimation is greatest for events priced below 5 percent; in the Iowa example, it is present for events priced between 0.2 and 0.3.

A range of experiments by psychologists suggests that the inability to distinguish small from tiny probabilities is not specific to prediction markets. It is, rather, a result of behavioral biases.<sup>28</sup>

A complementary view also suggests that specific types of trading behavior may lead to an overestimation of small probabilities. Charles Manski considered a setting in which prediction-market traders set their bet sizes to maintain a constant downside risk.<sup>29</sup> If this is the case, they

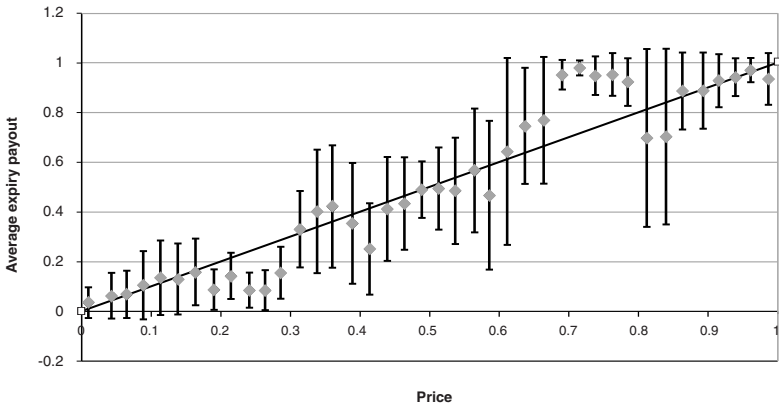
FIGURE 2-1  
**FAVORITE-LONGSHOT BIAS: RATE OF RETURN AT DIFFERENT ODDS**



SOURCE: Trackmaster, Inc.

NOTE: Sample is all horse races in the United States, 1992–2002, n=5,067,832 starts in 611,807 races.

FIGURE 2-2  
**IOWA ELECTRONIC MARKETS: PRICES AND EXPIRY PAYOUTS**



Graph plots the average expiry price of a winner-take-all contract on the Iowa market, conditional on its current price. Data is divided by current price into groups that are 2.5 percentage points wide. Error bars are 95 percent confidence intervals of the estimate of the mean expiry price, calculated from standard errors that are adjusted for sampling the same contract type multiple times. Consecutive groups, however, contain many of the same contracts and thus group means are not independent of each other.

SOURCE: Iowa Electronic Markets, <http://www.biz.uiowa.edu/iem/>.

demand twenty-four times as many contracts when going long at \$0.04 than when shorting the same contract, and, even absent transaction costs, a market price of \$0.04 will reflect the beliefs of traders at the ninety-sixth percentile of the beliefs distribution.<sup>30</sup> Thus, if errors in subjective probability assessments are symmetric, markets may overestimate low probabilities. Wolfers and Zitzewitz analyzed a broader class of models and found these biases to be generally quite small.<sup>31</sup>

The framework discussed earlier suggested a reason these biases could persist in equilibrium. The intuition for the result is that when transaction costs are positive, prices near zero or one are set by those with extreme opinions. To give a concrete example, suppose the objective probability for an event is 3 percent and the market-maker's transaction cost ( $t_{MM}$ ) is 1 percent. In the absence of informed trading or a net market-maker position, a market-maker will be willing to buy at \$0.02 and sell at \$0.04. If traders also face a transaction cost ( $t$ ) of \$0.01, then they will buy if their subjective probability is greater than 5 percent and sell if it is less than 1 percent. The former involves an overestimate by a factor of 1.67, the latter an underestimate by a factor of 3. Given these magnitudes, overestimates may be more common than underestimates, leading a risk-averse market-maker to set a bid-ask midpoint greater than the objective probability.

### Separating Correlation and Causation

Decision-market securities are designed to allow one to estimate how expectations of policy outcomes vary with the policy chosen. For example, the Iowa markets ran contracts that paid one penny for each percentage point of the two-party vote share won by the Democrats, conditional on the trader also correctly picking the winner of the Democratic nomination race. The ratio of the price of this contract to the price of a contract that pays \$1.00 if the candidate is nominated yields an estimate of the expected vote share of each candidate, conditional on the candidate being nominated (table 2-1 on the next page). In a previous paper, we reported that the expected vote shares on January 29, 2004, were 55 percent for John Edwards, 50 percent for John Kerry, and 46 percent for Howard Dean—if the relevant candidate were to win the nomination.<sup>32</sup>

TABLE 2-1  
CONTINGENT MARKETS: 2004 PRESIDENTIAL ELECTION

Contract Pays Conditional on Specific Democratic Candidate	Democratic Candidate Vote Share (Contract price, \$) A	Republican Vote Share against This Candidate (Contract price, \$) B	Implied Probability This Candidate Wins Nomination (%) C = A + B	Expected Share of Popular Vote if Nominated(%) D = A/C
John Kerry	0.344	0.342	68.6	50.1
John Edwards	0.082	0.066	14.8	55.4
Howard Dean	0.040	0.047	8.7	46.0
Wesley Clark	0.021	0.025	4.6	45.7
Other Democrats	0.015	0.017	3.2	46.9

SOURCE: Iowa Electronic Markets, closing prices, January 29, 2004.

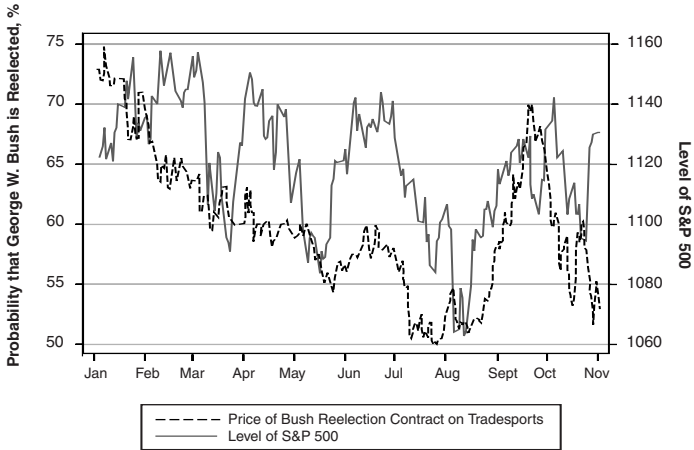
NOTE: Columns A and B show the prices of contracts that pay \$0.01 for each percentage point of the two-party popular vote won by Democrats or Republicans respectively, conditional on picking the winner of the Democratic nomination. (Contracts pay \$0.00 if the selected candidate does not win the Democratic nomination.)

It is tempting to draw a causal interpretation from these results: that nominating John Edwards would have produced the highest Democratic vote share. Indeed, as the name “decision markets” implies, this is the inference that we are intended to draw, and, in many circumstances, the most likely source of a correlation will be causal. (Berg and Rietz provide a related account of markets on the 1996 Republican nomination.<sup>33</sup>)

But alternative explanations exist. For example, on January 29, Edwards was behind in the delegate count, and only rated a 15 percent probability of winning the nomination. A come-from-behind victory would arguably have required a very good campaigning performance by Edwards or a shift in public sympathies toward his “Two Americas” message, either of which would have boded well for his general election performance. The decision market tells us that in the state of the world in which Edwards wins the nomination, he will also probably do well in the general election. This is not the same as saying that he will do well if, on the basis of the decision market, Democrats nominate Edwards.<sup>34</sup>

A related example of the difficulties of separating correlation from causation comes from the analyses in the financial press of the correlation of

FIGURE 2-3  
**BUSH'S REELECTION PROSPECTS AND THE STOCK MARKET, 2004**



SOURCE: Snowberg, Wolfers, and Zitzewitz, "Partisan Impacts on the Economy: Evidence From Prediction Markets and Close Elections."

President Bush's reelection chances with the performance of the stock market.<sup>35</sup> Figure 2-3, from Snowberg, Wolfers, and Zitzewitz, shows the close relationship between the level of the stock market and the TradeSports contract tied to Bush's reelection.<sup>36</sup> Regressing the (log) level of S&P on the level of Bush price yields a coefficient of 0.204, which led several commentators to infer that Bush's reelection would raise the S&P by 20.4 log percentage points. Difference specifications, which are appropriate given that these are asset prices, yield an estimated effect of 8–16 percentage points.

Only a few careful commentators noted the potential of an endogeneity issue: Bad economic news was hurting both financial markets and Bush's reelection prospects. The disappointing nonfarm payrolls report released three weeks prior to the election provides a stark example of the importance of this omitted causal link: Immediately upon the report's release, the Bush reelection contract fell by 1.5 percentage points (and the S&P 500 future fell by about 0.5 percent).<sup>37</sup>

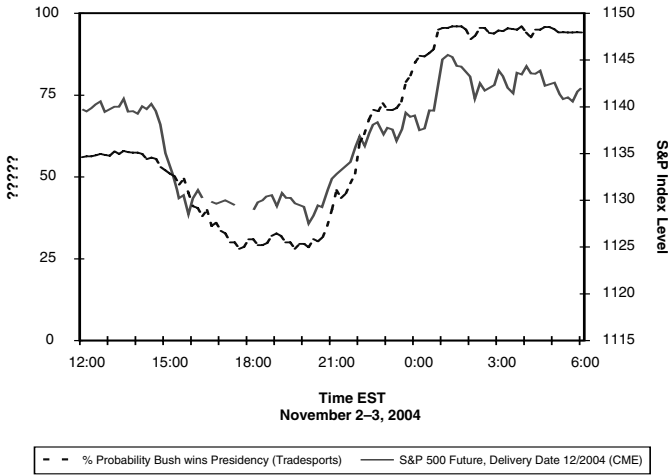
A common approach to endogeneity issues is to use instrumental variables (IV) estimation. In this context, a valid “instrumental event” must affect Bush’s reelection probability, and it affects the stock market only through Bush’s reelection probability. The third debate provides one such example.<sup>38</sup> In the three-hour period consisting of the debate and one and a half hours of post-debate spin, the Bush reelection security fell 3 percentage points, and the S&P 500 rallied by 10 basis points. The standard deviation of a three-hour S&P move at that time of day is about 9 basis points, so a 95 percent confidence interval around the estimated market effect of the debate would range from  $-8$  to  $+28$  basis points. Scaling this up to an estimate of the effect of the reelection of President Bush relative to a Kerry counterfactual suggests that the effect of Bush’s reelection can be bounded by  $-9.3$  and  $+2.7$  percentage points. Wherever the true causal effect lies in this interval, we can be confident that the naïve inference that a Bush White House would raise the value of U.S. equities by 20 percent was false.

As Snowberg, Wolfers, and Zitzewitz report, an even starker experiment was provided on election night 2004. Figure 2-4 shows the price of the Bush security on TradeSports, sampled every ten minutes, and the S&P 500 (or the December S&P future when the New York markets closed). Clearly, there is strong co-movement in the chart, but the magnitudes are particularly interesting. From 3:00 p.m. to 6:00 p.m., various websites leaked early exit polls that suggested a likely Kerry presidency. These revelations were accompanied by a 1 percent decline in the S&P and roughly a 25-percentage-point decline in the Bush security (from 55 to 30). Later that evening, election returns overturned the exit polls, and from 8:30 p.m. to 1:00 a.m. eastern time, the Bush security rallied from 30 percent to more than 90 percent, and the stock market increased by about 1.5 percent.<sup>39</sup> The former movement scales to an estimated Bush effect of  $+4$  percentage points ( $= -1/-25$ ), the latter to an estimate of  $+2.5$  percentage points ( $= 1.5/60$ ).<sup>40</sup> Again, these experiments strongly suggest a different causal effect of a Bush presidency on the stock market than a naïve interpretation of figure 2-3 suggested.

Finally, an instrumental variables approach could potentially be extended to decision markets. For instance, we argued earlier that the price of the “Democrat Vote Share if Edwards” contingent security provided only an estimate of the correlation between Edwards’s winning the nomination

FIGURE 2-4

TITLE



SOURCE: Snowberg, Wolfers, and Zitzewitz, “Partisan Impacts on the Economy: Evidence From Prediction Markets and Close Elections.”

and the Democrats’ vote share. An ideal instrument for Edwards’s receiving the nomination would be an event that affected the likely winner of the Democratic primary, but not Edwards’s general election chances. One such event might have been if a sex scandal hurt his only rival.<sup>41</sup> And, indeed, rather than waiting for the event to occur (or, as Matt Drudge did, pretending it did occur), we could run prediction markets to derive the market’s expectations of the relevant moments required for an IV estimator. That is, an “IV prediction market” would require five contracts, from which one could derive the two moments required for an IV estimate:

1. A contract paying \$1.00 if Edwards is nominated,  $P_1 = p(\text{Edwards})$
2. A contract paying \$1.00 if Edwards is nominated and a scandal occurs,  $P_2 = p(\text{Edwards} \ \& \ \text{Scandal})$
3. A contract paying \$1.00 if a scandal occurs,  $P_3 = p(\text{Scandal})$

4. A contract paying the Democratic vote share if Edwards is nominated and a sex scandal occurred,  $P_4 = E[\text{Democrat Votes} \mid \text{Edwards \& Scandal}] * p(\text{Edwards \& Scandal})$
5. A contract paying the Democratic vote share if Edwards is nominated,  $P_5 = E(\text{Democrat Votes} \mid \text{Edwards}) * p(\text{Edwards})$

The ratio  $P_2/P_3$  yields the probability of Edwards's nomination conditional on a scandal,  $p(\text{Edwards} \mid \text{Scandal})$ . Furthermore, subtracting the unconditional probability of Edwards winning the nomination ( $P_1$ ) yields the increase in probability resulting from the scandal:  $P_2/P_3 - P_1 = p(\text{Edwards} \mid \text{Scandal}) - p(\text{Edwards})$ . This difference is analogous to the first-stage regression coefficient in a Wald estimator. Likewise, the ratio  $P_4/P_3$  yields Edwards's expected general election vote-share condition on a scandal:  $E\{[E(\text{Democrat Votes} \mid \text{Edwards}) * p(\text{Edwards})] \mid \text{Scandal}\} * p(\text{Edwards} \mid \text{Scandal})$ . Subtracting the expected Democratic vote share if Edwards is nominated ( $P_5$ ) yields the effect of the scandal on Edwards's expected general election vote share:  $P_4/P_3 - P_5 = E[\text{Democrat Votes} * p(\text{Edwards}) \mid \text{Scandal}] - E[\text{Democrat Votes} * p(\text{Edwards})]$ . This difference is analogous to the coefficient from the reduced-form regression. The ratio of the reduced-form and the first-stage regression coefficients yields the "prediction IV" Wald estimator, which in this case would be the ratio of the scandal's effect on Edwards's expected general election vote share and its effect on his chances of winning the nomination, or:

$$\beta^{\text{Prediction IV}} = \frac{P_4/(P_3 - P_5)}{P_2/(P_3 - P_1)}.$$

The numerator of this ratio is the increase in Edwards's expected general election vote share resulting from the scandal, and the denominator is the increase in his probability of being nominated. Given our identifying assumption that the scandal does not affect Edwards's general election chances, this ratio provides an estimate of Edwards's general election performance in states of the world in which he would not have won the nomination but for an exogenous event.<sup>42</sup> It seems plausible that "IV prediction markets" like this better identify causal parameters and hence yield estimates that are more directly applicable for decision-making.<sup>43</sup>

Of course, although we are optimistic that “prediction IVs” are feasible, they run into the same five open questions that we asked earlier in this chapter:

**1. Would traders be willing to trade the relevant securities?**

This is particularly problematic given that few traders have a need to hedge against the types of possible events that make for plausibly exogenous experiments.

**2. Are the relevant outcomes contractible?**

**3. Are these markets manipulation-proof?** Even though the prediction IVs isolate the causal parameter of interest, as is well known, even small changes in the first-stage estimate (or market price) yield large changes in the estimated causal parameter, suggesting that the returns to manipulation may be large.

**4. Would the market be well-calibrated on the small probabilities in these multistate contingent contracts?**

**5. Can we ever fully separate correlation from causation?** As with a lot of other empirical work, it can be difficult to find experiments that truly do identify the causal parameter. The one reason for real optimism on this front is that we no longer need wait for natural experiments, but rather can trade on the likely effects of experiments that may never happen.

## Conclusions

In a previous essay, we reviewed much of the accumulated evidence on the operation of prediction markets, concluding on an optimistic note.<sup>44</sup> In this essay, we pose five questions which, we argue, require answers for prediction markets to reach their considerable potential. The research agenda suggested by these questions spans several fields of economics and encroaches significantly on related disciplines.

The first question arguably falls more in the field of marketing than economics: How can one attract uninformed order flow to markets? This

question is important because these traders provide the potential profit-motivating informed groups to trade. Lower transaction costs, in both the monetary and the convenience sense, are important, but inherent interest or buzz is clearly an important determinant.

Our second question concerns contractibility: How does one trade off interest with outcome contractibility? Of course, the domain of contracts has expanded considerably over the past few centuries, and there may be lessons in the history of contracts that prediction-market designers would be well-advised to follow.

Third, many of the corporate and policy decision markets that have been proposed raise questions about manipulation, especially in environments where free entry by arbitrageurs cannot eliminate the problem. With position limits, a prediction market may turn into voting; with no position limits, it may turn into an auction. If the markets literally turn into a complicated version of a vote or an auction, presumably we are better off with the simple version. But one might imagine situations in which they turn into a hybrid, and more work on understanding when those hybrids are sensible mechanisms could be productive.

Fourth, as psychologists and insurance salespeople have known for years, most people are badly calibrated when evaluating small probabilities, and some of their difficulties appear to be spilling over into prediction-market pricing. Given the current scale of prediction markets, the relevant limit to arbitrage is probably not the agency problems discussed in Shleifer and Vishny,<sup>45</sup> but rather transaction costs. We are optimistic that both declining transaction costs and carefully framed prediction-market contracts will yield more accurate responses.

Finally, there is the issue of separating causation and correlation, which is a difficult one, but more or less on empirical economists' home turf. For settings where interpretation of correlations is problematic, we propose an "instrumental events" approach for analyses based on time-series movements and an analog in contingent markets—"prediction IVs"—which may help isolate causal parameters.

## Notes

1. For example, a Factiva search for mentions of “Iowa Electronic Markets” from January 1 to November 1 yielded twenty-seven hits in 2000 and one hundred in 2004. Adding the largest private-sector market, TradeSports.com (and its politics-only subsidiary Intrade) to the search increases the number of 2004 hits to 237.

2. Joel Slemrod and Timothy Greimel, “Did Steve Forbes Scare the Municipal Bond Market?” *Journal of Public Economics* 74, no. 1 (1999): 81–96; Brian Knight, “Are Policy Platforms Capitalized into Equity Prices? Evidence from Equity Markets during the Bush/Gore 2000 Presidential Election,” *Journal of Public Economics* (2006, forthcoming).

3. Andrew Leigh, Justin Wolfers, and Eric Zitzewitz, “What Do Financial Markets Think of War in Iraq?” (Working Paper 9587, National Bureau of Economic Research, Cambridge, Mass., 2003).

4. See, for example, references cited in notes 2 and 3, above.

5. Kay-Yut Chen and Charles Plott, “Information Aggregation Mechanisms: Concept, Design and Implementation for a Sales Forecasting Problem” (Working Paper 1131, California Institute of Technology, Pasadena, 2002)

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

7. Justin Wolfers and Eric Zitzewitz, “Experimental Political Betting Markets and the 2004 Election,” *Economists’ Voice* 1, no. 2 (2004) <http://www.bepress.com/ev/vol1/iss2/art1/> (last accessed January 26, 2006).

8. Michael Abramowicz, “Information Markets, Administrative Decisionmaking, and Predictive Cost-Benefit Analysis,” *University of Chicago Law Review* 71 (2004): 933–1020.

9. Robert W. Hahn and Paul C. Tetlock, “Using Information Markets to Improve Decision Making,” *Harvard Journal of Law and Public Policy* 29, no. 1 (Fall 2005): 213–89.

10. Robin Hanson, “Shall We Vote on Values, but Bet on Beliefs?” (working paper, George Mason University, 2003).

11. These figures include the trading on a symmetric “Kerry to win” contract.

12. Paul Milgrom and Nancy Stokey, “Information, Trade, and Common Knowledge,” *Journal of Economic Theory* 26 (1982): 17–27.

13. Albert Kyle, “Continuous Auctions and Insider Trading,” *Econometrica* 53, no. 6 (1985): 1315–36.

14. To derive the above, solve the traders’ problem: Set  $x$  to maximize subjective expected utility  $E[U] = e^* \ln[w + (1 - p)x] + (1 - e^*) \ln(w - px)$ .

15. For more information, see Justin Wolfers and Eric Zitzewitz, “Interpreting Prediction: Market Prices as Probabilities” (working paper, University of Pennsylvania, Philadelphia: 2005).

16. Suppose that  $\bar{I}$  traders are informed, each informed trader buys one contract, and  $I - \lambda$  traders are uninformed. Then the market-makers' bids and asks must satisfy

$$P_{ask} = \frac{\lambda q + (1 - \lambda) X_{buy} (P_{ask} + t) q}{\lambda q + (1 - \lambda) X_{buy} (P_{ask} + t)} + t_{MM} - c X_{MM} \quad \text{and}$$

and

$$P_{bid} = \frac{-(1 - \lambda) X_{sell} (P_{bid} - t) q}{\lambda(1 - q) - (1 - \lambda) X_{sell} (P_{bid} - t)} t_{MM} - c X_{MM},$$

where the first term in each expression is the expectation of  $y$  conditional on receiving a buy or sell order. With  $\lambda = 0$ , this first term is always equal to  $q$  (the market-maker's unbiased prior), but as  $\lambda$  increases,  $P_{ask}$  increases and  $P_{bid}$  decreases, leading to wider spreads and less uninformed trading.

17. In the expressions in the above footnote, it is easy to show that  $P_{ask}$  is increasing and  $P_{bid}$  is decreasing in  $t$  and  $t_{MM}$ . It is likewise straightforward to show that when  $\lambda > 0$ , then for distributions of  $e$  that have  $\min(e) > 0$  and  $\max(e) < 1$ , there will be a level of  $t$  that leads to market unraveling, that is, to  $P_{ask} = 1$  and  $P_{bid} = 0$  being the only values that satisfy the expressions in the above footnote.

18. The term “uninformed” overstates the requirement, which is that the market must attract traders whose estimate of the true probability is less precise than the market-maker's, or whose hedging demand causes them to trade even when their beliefs coincide with those of the market-maker.

19. The Iowa markets are run as a research project by academics at the Tippie College of Business at the University of Iowa. They are described in Robert Forsythe, Forrest Nelson, George R. Neumann, and Jack Wright, “Anatomy of an Experimental Political Stock Market,” *American Economic Review* 82, no. 5 (1992): 1142–61.

20. For example, this contract averaged fifty-two on Iowa during the month of August 2004 and about forty-six on TradeSports.

21. Emile Servan-Schreiber, Justin Wolfers, David Pennock, and Brian Galebach, “Prediction Markets: Does Money Matter?” *Electronic Markets* 14, no. 3 (2004): 243–51.

22. Refet Gürkaynak and Justin Wolfers, “Macroeconomic Derivatives: An Initial Analysis of Market-Based Macro Forecasts, Uncertainty and Risk,” forthcoming, in *NBER International Seminar on Macroeconomics*, ed. Christopher Pissarides and Jeffrey Frankel (MIT Press: Cambridge, 2005).

23. Another example of a TradeSports contract whose definition was later clarified is the “Will Bush win the electoral votes of Maine?” contract. One might have supposed this meant all four electoral votes, but the exchange determined after the contract had been trading for several months that three out of four

would suffice. In the same vein as Leigh, Wolfers, and Zitzewitz, “What Do Financial Markets Think of War in Iraq?” we commented on the judgment that would have been required to settle the “Will Saddam Hussein be out of office?” contract had a contract been expiring on April 9, 10, or 11; the active trading on this contract suggests that many traders were comfortable with this possible ambiguity.

24. Robin Hanson and Ryan Oprea, “Manipulators Increase Market Accuracy” (working paper, George Mason University, Fairfax, Va., 2004).

25. Koleman Strumpf, “Manipulating the Iowa Political Stock Market” (working paper, University of North Carolina, Chapel Hill, 2004).

26. Richard H. Thaler and William T. Ziemba, “Anomalies: Parimutuel Betting Markets: Racetracks and Lotteries,” *Journal of Economic Perspectives* 2, no. 2 (Spring 1988): 161–74.

27. Justin Wolfers and Eric Zitzewitz, “Prediction Markets,” *Journal of Economic Perspectives* 18, no. 2 (2004): 107–26.

28. Daniel Kahneman and Amos Tversky, “Prospect Theory: An Analysis of Decision under Risk,” *Econometrica* 47, no. 2 (1979): 263–92.

29. Charles Manski, “Interpreting the Predictions of Prediction Markets” (Working Paper 10359, National Bureau of Economic Research, Cambridge, Mass., 2004).

30. Whether traders use such a budgeting rule is, of course, an empirical question that remains to be answered.

31. Justin Wolfers and Eric Zitzewitz, “Interpreting Prediction Market Prices as Measures of Beliefs” (working paper, University of Pennsylvania, Philadelphia, 2005).

32. Wolfers and Zitzewitz, “Prediction Markets.”

33. Joyce Berg and Thomas Rietz, “Prediction Markets as Decision Support Systems,” *Information Systems Frontiers* 5, no. 1 (2003): 79–93.

34. This issue is referred to as “decision selection bias” in Robin Hanson, “Decision Markets,” *IEEE Intelligent Systems* 14, no. 3 (1999): 16–19; Robin Hanson, “Impolite Innovation: The Technology and Politics of ‘Terrorism Futures’ and Other Decision Markets,” in *Promoting the General Welfare: American Democracy and the Political Economy of Government Performance*, ed. Eric Patashnik and Alan Gerber (Washington, D.C.: Brookings Institution Press, 2005).

35. See, for example, E. S. Browning, “As Bush Goes So Goes the Market,” *Wall Street Journal*, September 20, 2004, C1.

36. Erik Snowberg, Justin Wolfers, and Eric Zitzewitz, “Partisan Impacts on the Economy: Evidence From Prediction Markets and Close Elections” (working paper, Stanford University, Stanford, Calif., 2006).

37. Authors’ calculation from financial data.

38. The other debates are unfortunately of less use. For the first debate there were no instant polls to determine a winner objectively; it took at least one news

cycle for a consensus to emerge that Kerry had won, and this leaves us with too long an event window. The second debate was on a Friday evening, when futures markets were closed, and the vice-presidential debate was considered to be approximately a draw.

39. The slow incorporation of information in prediction markets is illustrated by figure 2-4. The bulk of the exit poll–related movement occurred between 2:50 p.m. and 3:30 p.m. eastern time in the stock market, whereas it lasted until 5:30 p.m. eastern time in the TradeSports reelection market. Likewise, the subsequent rally began about ten to twenty minutes earlier on the Chicago Mercantile Exchange than on TradeSports.

40. Standard deviations of S&P changes during the 2:50 p.m.–3:30 p.m. and 8:30 p.m.–1:00 a.m. windows are twenty-six and twelve basis points, respectively. These imply standard errors of roughly 1.0 and 0.2 percentage points, respectively.

41. Of course, a sex scandal involving a senior Democrat might affect the general public's views of all Democrats, which would potentially invalidate this instrument.

42. Note that as with most applications of IV methods, we identify a Local Average Treatment Effect (LATE). Naturally, whether this is the causal parameter of interest depends on the context. For example, if Edwards is of higher average quality in states of the world in which he can take advantage of a small scandal than in states of the world where he can take advantage of a large scandal, then a LATE estimator based on a small scandal may overstate the performance one could expect from exogenously choosing him as the nominee.

43. Naturally, this example shows only one of many such market designs that can be used to recover the causal parameter of interest.

44. Wolfers and Zitzewitz, "Prediction Markets."

45. Andrei Shleifer and Robert Vishny, "The Limits of Arbitrage," *Journal of Finance* 52, no. 1 (1997): 35–55.