Accuracy and Forecast Standard Error of Prediction Markets*

Joyce Berg, Forrest Nelson and Thomas Rietz

Departments of Accounting, Economics and Finance

Henry B. Tippie College of Business Administration

University of Iowa

Working Draft

July 2003

*For many helpful comments and suggestions, we thank Bob Forsythe, George Neumann, Gene Savin, Toni Whited and participants in the Department of Finance Seminar at the University of Iowa Tippie College of Business.

Accuracy and Forecast Standard Error of Prediction Markets

Abstract

"Prediction markets" are designed specifically to forecast events. Though such markets have been conduced for more than a decade, to date there is no analysis of their long-run predictive properties. We provide the first systematic evidence on the long-run predictive power of these markets by studying ex post accuracy and means of measuring ex ante forecast standard errors. Ex post, prediction markets prove accurate at long and short forecasting horizons, in absolute terms and relative to natural alternative forecasts. We use efficient markets theory and some special properties of the markets to develop forecast standard errors. Both time series and inter-market pricing relationships suggest that markets generate efficient random walks in prices. Thus, random walk projections generate reasonable confidence intervals. These confidence intervals differ dramatically from margins of error quoted in polls. We argue this is reasonable because polls do not attempt to, nor can they be expected to, measure the degree of uncertainty about the eventual election outcome conditional on their own results. In contrast, the markets incorporate this uncertainty by design.

Accuracy and Forecast Standard Error of Prediction Markets

"Prediction markets" are designed and conducted for the primary purpose of aggregating information so that market prices forecast future events. These markets differ from typical, naturally occurring markets in their primary role as a forecasting tool instead of a resource allocation mechanism. For example, since 1988, faculty at the Henry B. Tippie College of Business at the University of Iowa have been running markets through the Iowa Electronic Markets (IEM) project that are designed to predict election outcomes.[1] These represent the longest running set of prediction markets known to us. They have proven efficient in forecasting the evening and week before elections. However, no analysis of their long-run forecasting power has been conducted. Here, we analyze these markets to show how prediction markets in general can serve as efficient mechanisms for aggregating information and forecasting events that can prove difficult for traditional forecasting methods. We put special focus on longer-run properties.

Existing evidence (e.g., Berg, Forsythe, Nelson and Rietz, 2003, and references cited therein) shows excellent ex-post predictive accuracy for election prediction markets in the very short run (i.e., one-day-ahead forecasts using election eve prices). While this is an interesting and important result, it does not address the critical question of whether prediction markets can serve as effective long-run forecasting tools (weeks or months in advance). Here, we present the first systematic analysis of election market data on two additional properties that are important for evaluating their long-run efficacy. The first property we study is the longer-run predictive accuracy of markets relative to their natural competitors: polls. This analysis provides the first documented evidence that prediction markets are considerably more accurate long-run forecasting tools than polls across elections and across long periods of time preceding elections (instead of just on election-eve). The second property we study is the forecast standard error of market predictions. This allows us to have a (previously unavailable) measure of confidence in ex-ante market predictions. We study three means of measuring forecast

[1]Since 1993, these markets have expanded to predict many other types of events including other political outcomes, financial and accounting outcomes for companies, national and international economic phenomena, box office receipts for movies, etc.

standard errors. First, we show the difficulty in applying a previously developed structural model designed to explain short-run, ex post accuracy to out-of-sample data. Second, we show that the time series of forecasts from our prediction markets are consistent with efficient market random walks. From this, one can construct forecast standard errors. Third, we show that an efficient inter-market pricing relationship can be exploited to the same end. These estimated forecast standard errors appear somewhat larger, but are not significantly different from the random walk approach. We suggest that both should be used to get a reasonable estimate of forecast standard errors and confidence intervals for prediction markets.

I. Prediction Markets

Since Hayek (1945), economists have recognized that markets have a dual role. They allocate resources and, through the process of price discovery, they aggregate information about the values of these resources. The information aggregation role of some markets seems particularly apparent. For example, corporations cite the value of their stock as the consensus judgment of their owners about the value of the corporation's activities. Increasingly, corporations reward managers based on this value measure. Futures and options markets aggregate information about the anticipated future values of stocks and commodities. If it is true that futures prices are the best predictors of actual future spot prices (as the "expectations hypothesis" asserts), then futures prices constitute forecasts.[2] For example, Krueger and Kuttner (1996) discuss how the Federal Funds futures contract can be used to predict future Federal Funds rates and, hence, future Federal Reserve target rates.

In most markets, if prediction uses arise, they do so as a secondary information aggregation role. However, some recent markets have been designed specifically to exploit their information aggregation

[2] Debate over the ability of futures markets to forecast future prices extends back to Keynes (1930) and Hicks (1946). Many of the arguments result from the secondary nature of information aggregation in these markets. The early "normal backwardization" versus "contago effect" arguments were based on relative power of speculators and hedgers. Today, the idea that "risk neutral" probabilities used to price futures and options differ from the "true" underlying probabilities result from relative levels of hedging demand in the markets. While the IEM markets discussed below may be subject to price deviations due to hedging activities, the narrow scope of the IEM markets, the small size of investments and analysis of individual traders (e.g., Forsythe, Nelson, Neumann and Wright, 1992, and Forsythe, Rietz and Ross, 1999) all lead us to conclude that hedging activities do not affect IEM prices significantly.

characteristics for use as dynamic forecasting systems. Examples of such "prediction markets" include numerous markets run under the Iowa Electronic Markets (designed to predict elections, other political events, movie box office receipts, corporate earnings, returns, stock prices, etc.), similar markets run in other countries (usually designed to predict election outcomes) and markets cited in Plott (2000) (designed to predict sales at a large corporation). While the majority of such markets are run with cash payoffs, some similar Internet "games" have been run using fictitious currency with prize contests as motivation. These include the Foresight Exchange (http://www.ideosphere.com) with "payoffs" tied to a wide range of social, political and scientific events/issues, the Hollywood Stock Exchange (http://www.hsx.com) with "payoffs" tied to movie box office take and the (apparently now defunct) Major League Market (http://majorleaguemarket.com) with "payoffs' tied to the performance of teams and athletes and similar markets with contract "de-listing values" (i.e., liquidating "payoffs") tied directly to predictable events (in contrast to vague notions of "popularity" of contracts).

Prediction markets, such as the Iowa Electronic Markets (IEM), represent an important advance in forecasting. The idea is simple: trade contingent claims in a market where the claims pay off as a function of something one is interested in forecasting. If structured correctly, the prices should reflect the expected payoffs to the claims. This relationship can be used for forecasting. For example, the IEM's vote-share markets trade contracts with payoffs that equal $1 times the relative percentages of the vote taken by candidates in an upcoming election. Prices should converge to the market's expectation of relative vote shares. Though simple in concept, such markets act as complex, dynamic, interactive systems that incorporate information in new ways. Through the actions of traders, prediction markets aggregate information from individuals, incorporate polls and other sources of information and weight all of this information through the price formation process. They compete directly with, and potentially use as information, traditional methods of forecasting such as polls, econometric modeling and marketing surveys.

We pose two questions about prediction markets. (1) Are prediction markets accurate for forecasting purposes in an ex post sense far in advance? (2) Can we develop an ex ante means of assessing the likely predictive accuracy of, or develop confidence intervals for, prediction markets? Both are important for using prediction markets as long-run forecasting tools and have not been addressed in the prior literature. In the rest of the paper, we use data from IEM political markets as examples of prediction markets to study these questions. In the next section, we discuss general principles of forecasting and properties of good forecasting tools.

II. The Forecasting Problem

Consider the problem of forecasting some measurable future outcome. Denote the actual outcome, which will occur at date T, by S_T. Examples of S_T include sales for a company, the price of a good, the temperature or, for the particular forecasts studied here, the outcome of an election. A useful forecasting tool should give two things as of date t (where t is some date before the outcome date T): (1) a point estimate of S_T conditional on information available on date t (i.e., $s_t := \mathrm{E}(S_T|I_t)$ where s_t is the forecast and I_t represents information available at date t) and (2) some measure of confidence in the estimate. Typically, the measure of confidence is a standard deviation conditional on information as of date t (i.e., $\sigma_{s,t} := \sqrt{E[(S_T - s_t)^2 \mid I_t]}$) from which confidence intervals can be constructed around the point estimate. We will refer to $\sigma_{s,t}$ as the forecast standard error.

Commonly used forecasting techniques include (1) time series models (e.g., historical values of the outcome to be forecast with trends, auto-regressive and moving average components when needed); (2) structural models (e.g., regression models based on input variables estimated on past observations); (3) sampling (e.g., surveys or polls) and (4) less formal methods such as focus groups, interviews of knowledgeable parties and expert panels. Given a sufficient number of observations under essentially identical conditions and sufficient stationarity, both time series and structural models can give standard deviations for model forecasts based on the law of large numbers or the small sample properties of the models. However, frequently the prediction problem lacks sufficient data, stationarity or both. For example, sales in a rapidly changing market or of a new product may prove difficult for such models to predict. Given a random sample and a sufficiently static environment, surveys or polls can generate confidence intervals using sampling theory and the law of large numbers. However, obtaining a truly random sample can be difficult (e.g., the Truman/Dewey race) and often the environment can change quickly. For example, political campaigns are designed to influence how people will vote in an upcoming election. If they are effective, one cannot reasonably expect the opinions of voters to remain static in the presence of two or more well-funded and well-run campaigns.

Here, we ask whether prediction markets can serve as effective forecasting tools for problems that prove difficult for more traditional methods. As an example of the process, we study the behavior of four markets from the IEM designed to predict US Presidential election outcomes and compare them to the obvious alternative: polls. We believe that prediction markets like the IEM should predict complex phenomena such as election outcomes accurately for several reasons. First, the market design forces traders to focus on the specific event of interest. For example, traders in the IEM election markets can reap profits if they can predict well the specific upcoming election. This requires more than simply building a model based on past elections (because of the large differences across elections) and more than simple consideration of a fictitious election "if it were to be held today" (as polls ask respondents to consider). Second, to voice their opinions, traders must open a position in the market, putting money at stake. Presumably, the more confident that they are in their predictions, the more money they will be willing to risk. Third, the market aggregates the diverse information of traders in a dynamic and, hopefully, efficient manner.

Ex post evidence suggests that prediction markets can be good at forecasting in the very short run. Berg, Forsythe, Nelson and Rietz (2003) summarize the evidence from 49 IEM election markets run between 1988 and 2000. Election-eve average absolute prediction errors average 1.37% for US Presidential elections, 3.43% for other US elections and 2.12% for non-US elections. They also find that the election-eve market forecasts generally predict better than the latest major national polls. New evidence presented in the current paper shows that markets are generally much better predictors than polls months in advance of the elections.[3]

While the point predictions for vote share prove accurate ex post, some measure of forecast standard error is necessary for making prediction markets useful forecasting tools. Here, we show how to develop ex ante measures of standard error for market forecasts. This is very different from developing a margin of error for a poll. To appreciate this, consider the difference between a poll's "margin of error" and a forecast standard deviation. These are two very different kinds of uncertainty. A poll is designed to measure and report the *current* distribution of voter responses, the average response and a confidence interval for that average. It is a snapshot. A poll's margin of error describes the degree of uncertainty surrounding the estimate of the *current* average response. Simply polling more voters will reduce this uncertainty (given a random sample). However, this is not the only uncertainty that affects the reliability of a forecast. The degree to which current responses reflect actual future actions also matters. Even if one believes that a voter's poll response is truthful (and reflects the most likely future action), it does not measure degree of conviction (or, alternatively, the chances a voter will change his or her mind, or that masses of voters will). Other than looking at the number of undecided voters, a poll says little about how responses, preferences or decisions of voters may change across time. Further, using a larger sample cannot reduce this type of uncertainty. Therefore, polls say little or nothing about the degree of uncertainty inherent in forecasting the eventual election outcome based on poll responses. In short, the margin of error is <u>not</u> the forecast standard error of the poll and does not pertain to predictive accuracy.

Now consider the problem for markets designed to predict election outcomes. The markets differ from polls in that they are designed specifically to forecast actual election outcomes. To profit, traders must take into account what may or may not happen between the current date and the election. Thus, by their design, markets

[3]Prediction markets also perform well in other areas. The IEM runs markets for a variety of different prediction problems (see the IEM website http://www.biz.uiowa.edu/iem for details). Plott (2000) shows that prediction markets can be used to accurately forecast sales for a company. While they are not prediction markets as defined here, evidence from traditional laboratory markets shows that laboratory markets can aggregate diverse information efficiently if traders have enough experience and the information structure is simple enough and the structure is commonly known. See Sunder (1995) for a summary of the evidence.

force traders to consider election dynamics. As we will show below, the price in a market designed to predict relative vote shares is a point estimate of the expected future vote-share outcome. By itself, it does not convey the degree of uncertainty about that outcome. Further, assessing the degree of uncertainty poses a difficult problem. One cannot treat the data as a cross sample (as polls would) because the market price is not an equally weighted average of a random sample. The market prices do not constitute an ordinary time series of fundamental variables. Instead, they are a time series of forecasts for a single future outcome. Because of this, we will rely primarily on properties of efficient markets and efficient forecasts along with inter-market pricing relationships to develop methods of measuring forecast standard error.

Here, we propose methods of determining forecast standard errors relying on efficient market theory, the analysis of market price/prediction dynamics and inter-market pricing relationships to measure the likely predictive accuracy ex ante for prediction markets. We use the four IEM Presidential markets run to date as examples. First, we apply a structural model based on market microstructure factors and calibrated on previous markets and election results. Then, we discuss a time series approach based on the behavior of prices in a given election market. Finally, we discuss an "implied volatility" approach that parallels option pricing techniques and uses data from simultaneous winner-takes-all election markets. The latter two approaches are consistent with each other and imply that a time series approach based on market prices and efficient markets theory can be used to generate forecast standard errors for prediction markets.

III. The Iowa Political Markets

The Iowa Political Markets, a subset of the Iowa Electronic Markets (IEM), are designed specifically to predict election outcomes. Started in 1988, they are the longest running set of formal prediction markets known to us. We use them to show how measures of forecast standard error can be developed for prediction markets.

The IEM political markets are computerized, electronic, real-time exchanges where traders buy and sell futures contacts with payoffs based on election outcomes.[4] Because real money is used, traders are subject to the monetary risks and returns that result from their trading behavior. Contracts in the political markets are designed to make two kinds of forecasts in separate markets: forecasts of winners and forecasts of vote shares. As with other financial futures markets, prices should represent consensus forecasts of future values adjusted for the risk free interest rate and the risk premium associated with the aggregate risk factor of the futures position. In the IEM, no risk adjustment is required because (1) the risk free rate in the market is zero and (2) neither an aggregate risk factor nor a premium for one can exist. Thus, the prices should reflect only expectations about the election outcomes.[5]

In IEM markets designed to forecast winners ("winner-takes-all" markets), contracts pay \$1 or \$0 conditional on the winner of the popular vote. Contracts span the space of possible outcomes by having one designated for each major candidate and an "other" contract for the rest of the field. Table 1 gives the specific contracts for the presidential winner-takes-all markets run to date. Simple no-arbitrage pricing arguments imply that prices of these contracts should equal the traders' consensus forecast of the probabilities of each candidate winning the election (see footnote 5). Thus, winner-takes-all markets forecast information about the distribution of outcomes, specifically the probabilities of each candidate winning.

[4]Because these are real futures contracts, the IEM is under the regulatory purview of the Commodity Futures Trading Commission (CFTC). The CFTC has issued a "no-action" letter to the IEM stating that as long as the IEM conforms to certain restrictions (related to limiting risk and conflict of interest), the CFTC will take no action against it. Under this no-action letter, IEM does not file reports that are required by regulation and therefore it is not formally regulated by, nor are its operators registered with, the CFTC.

[5]This can be shown in a variety of ways. See Malinvaud (1974) for the general equilibrium proof. One can also price the contracts as assets using CAPM and APT models. In each, $P_t = E(P_{t+h})/(1+k)^h$, where k (the required expected return) is the sum of the risk free rate and compensation for aggregate risk factors. Since the risk free rate is zero and there are no aggregate risk factors, the expected return on any given asset is zero. Alternatively, given that the expected market portfolio return is constrained to be zero by design, any factor risk premiums must be zero. Again, this makes for a zero expected return on any given asset. As a result, $P_t = E(P_{t+h})/(1+k)^h = E(P_{t+h})$. Even though traders cannot make the appropriate risk free hedges here (because they cannot trade the underlying fundamental asset), one might be tempted to use the modern portfolio theory futures pricing relationship: $F_{t+h} = E(P_{t+h}) x (1+r_f)^t/(1+k)^t$, where F_{t+h} is the time t futures price for delivery at date $t+h$, $E(P_{t+h})$ is the expected future spot price of the underlying fundamental, r_f is the risk free rate and k is the required expected return determined by the risk of the futures position. Again, both the risk free rate and the required expected return are constrained to be zero. This gives: $F_{t+h} = E(P_{t+h})$.

In IEM markets designed to forecast vote shares ("vote-share" markets), contracts pay an amount equal to the fraction of the popular vote received by a candidate times $1.[6] Table 1 gives the specific contracts for the presidential vote-share markets run to date. Appropriate contract specification and normalization insured that the contract payoffs always summed to $1.[7] Again, simple no-arbitrage arguments imply that market prices should reflect the traders' consensus forecast of the vote shares taken by each candidate (see footnote 5). Thus, vote-share markets provide point predictions about candidate vote shares. If, in advance of determining the election outcome, one thinks of each candidate's vote share as a random variable with some distribution, the vote-share market gives the mean of that candidate's vote-share distribution while the winner-takes-all market gives the aggregate probability of a given range of the distribution (in which that candidate receives the highest vote share).[8]

As a prediction system, the IEM differ from expert panels and polls in a number of respects. Instead of being a randomly selected, representative sample or a deliberately chosen panel, IEM traders are self-selected. People who are not interested either do not sign up or drop out. Further, the market does not equally weight traders' opinions in the price formation process. Instead, the market price is a weighted average which, through trading behavior and market dynamics, depends upon the traders' forecasts and the levels of confidence they have in their forecasts as well as an untold number of factors like aggressiveness, risk aversion, timing, wealth, etc. Unlike polls or expert panels in which participants are asked for their independent opinions, each trader in the market sees the net effect of the belief of all other traders, and the time series of changes in those beliefs, and can alter his own perceptions accordingly. This makes the market more than a static, one-time prediction but rather a dynamic system that can respond instantaneously to the arrival of new information. Unlike polls that ask each respondent how he or she would vote if the election were held today, the market asks traders to forecast how everyone will vote in the actual upcoming election.

[6]In 1988, the contracts paid the vote share times $2.50.
[7]In 1988, the total payoff was $2.50.
[8]This is a direct result of the argument in footnote 5. Problems that arise because actual traded contracts do not cover all aspects of the joint vote share distribution for multiple candidates are discussed later.

As an example of these differences, consider the demographics of IEM traders. A good poll would strive to collect responses from a random, representative sample. In contrast, IEM traders are self-selected and differ greatly from a representative sample of voters. In 1988, traders included only interested members of the University of Iowa academic community. In the other elections, traders included interested individuals from around the world. In the 2000 vote-share market, 20% of the traders were from Iowa while Iowa only accounted for 1% of the nation's population in 2000. Men constituted 75% of the active traders but only 49% of the overall population (and slightly less of the voting population). Our traders are typically young, white, well educated and have high family incomes. Thus, IEM predictive accuracy relies heavily on a sample (in practice, a non-representative sample) of interested traders forecasting the behavior of the voting population at large.

IV. Predictive Accuracy in Four Presidential Elections

The IEM has conducted markets on four US presidential elections. Tables 1 and 2 summarize each of these markets. In 1988, a vote-share market predicted the popular vote shares taken by Bush, Dukakis, Jackson and rest-of-the-field. In 1992, a winner-takes-all market predicted the probabilities of winning for the Democrat (Clinton), the Republican (Bush) and other candidates (primarily Perot). The vote-share market was split between two sub-markets. One sub-market predicted the vote split between the Democrat (Clinton) and the Republican (Dole). A second sub-market predicted the split between the two major parties and Perot. In 1996, a winner-takes-all market predicted the probabilities of winning for Clinton, another Democrat, the Republican (Dole) and rest-of-the-field. The 1996 vote-share market predicted the vote split between the Clinton as the Democratic nominee and Dole as the Republican nominee.[9] Finally, in 2000, winner-takes-all and vote-share markets forecast the election for the Democratic, Reform and Republican nominees (Gore, Buchanan and Bush, respectively). In this paper, we focus on the vote-share markets and the vote splits among the leading candidates

[9] Early in the 1996 market, separate sets of contracts predicted the vote splits for Clinton versus other possible Republican nominees including Alexander, Forbes, Gramm and rest-of-the-Republican-field.

because these are the most directly comparable to polls. We judge the accuracy of these market forecasts by comparing them to the actual election outcomes.

Table 2 shows statistics for the Presidential markets in the 1988 through 2000 elections. In 1992, 1996 and 2000, traders could participate in both vote-share and winner-takes-all markets. The number of active traders ranged from 155 in 1988 to 1151 in the 1996 winner-takes-all market. Overall volumes ranged from 15,826 contracts worth $8,123 in 1988 to 652,165 contracts worth $137,386 in the 1996 winner-takes-all market. The winner-takes-all markets appear more popular with traders. When one exists, the volume in the winner-takes-all market significantly exceeds the vote-share-market. This appears especially true in the last week of the elections, with winner-takes-all dollar volumes running 15-20 times vote-share volumes.

Figures 1 through 4 show the time series of (non-normalized) closing prices for the final sets of contracts in each vote-share market.[10] Horizontal lines represent the actual election outcomes. Dashed vertical lines show significant dates. For comparison purposes, the scales are the same across all four graphs. Figure 5 shows the winner-takes-all markets run in the 1992, 1996 and 2000 elections. These graphs highlight the variation across the elections. Descriptions of major events and price movements are given in the appendix for the interested readers.

A. IEM Predictions versus Polls

How does the IEM compare to other forecasting methods? Polls form the natural alternative.[11] Figure 6 contains graphs of spread predictions from major polls and market prices for the four presidential elections since

[10] The closing prices are the last trade prices before midnight each day. Though unit portfolio values should sum to 1 (because the payoffs to contracts always will sum to 1), individual closing prices may not because of non-synchronous trade. To account for this when making predictions, we typically normalize these prices by dividing each by their sum. This insures that the predicted vote shares sum to 1 and adjusts for non-synchronous trades. We present these graphs for the reader who may be interested in the non-normalized prices, but will continue to make predictions and judge accuracy using normalized prices.

[11] However, pollsters are averse to interpreting polls as forecasts. Worcester (1996) states, “Polls ... are useless in telling us much about the outcome of an election weeks or even months or years in the future. Nevertheless, the voting intention question is valuable for what it summarizes about people's attitudes and values at the moment.” Certainly election eve polls are regarded as forecasts of the election outcome. And whether or not the pollsters would have us give them this

1988. Market predictions are generated from closing prices (the last trade price before midnight each day). Poll outcomes represented in the graphs appear on the last day that polling took place for that particular poll, which is typically a day earlier than the release of the poll. Letters distinguish polls, indicating either the polling organization or the agent that requested and published the poll. For both market prices and polls, the outcomes are plotted as the normalized two-party vote margin. Thus, for example, the market outcome in Figure 6 for 1996 is computed as $s^{VS}_{Clinton-Dole,t} := \frac{p^{VS}_{Clinton,t} - p^{VS}_{Dole,t}}{p^{VS}_{Clinton,t} + p^{VS}_{Dole,t}}$, where s designates the normalized spread and p's designate closing market prices for candidates at time t.[12] The poll outcome is computed as $s^{Poll}_{Clinton-Dole,t} := \frac{r^{Poll}_{Clinton,t} - r^{Poll}_{Dole,t}}{r^{Poll}_{Clinton,t} + r^{Poll}_{Dole,t}}$, where s designates the normalized spread and r's designate the percentages of poll respondents for candidates at time t.[13] In all four graphs, vertical lines indicate significant mid-campaign events (party conventions and debates) and a horizontal line shows the actual election outcome.

Several things are obvious from the four graphs. First, the markets present a very different picture of the elections than the polls. What the polls are measuring as voter sentiment at any particular point in time frequently differs greatly from what the market predicts will actually occur in the election. The market prediction often stays well above or below all contemporaneous polls for extended periods of time. During these periods, the market is typically closer to the final outcome than contemporaneous polls. Second, all three graphs reveal a striking volatility in poll outcomes, both in absolute terms and in comparison to the market. Polls on the same day by different organizations or subsequent polls by the same organizations frequently differ dramatically, generating differences that fall outside the quoted margins of error. In each election, we observe the well-known poll phenomenon of "convention bounce" (the tendency for a party to rise in the polls during that party's

interpretation, the media and the public by and large do take polls to be some kind of forecast. Given this popular usage, it is natural to compare forecasts arising from the IEM with concurrent opinion polls.

[12] Closing prices are the last trade price before midnight each day. If no trade occurs in a day, the previous day's closing prices are carried over. Normalization adjusts for the possibility that non-synchronous trades lead to predictions that do not sum to 1 by adjusting each observation proportionately.

[13] The effect of this normalization on poll results is to allocate non-responses across the two candidates in proportion to the share of respondents choosing those candidates.

convention and then fall). These strong effects do not appear in the markets. Third, the market appears to forecast the election outcomes more closely than polls months in advance.

Table 3 presents a simple test of whether the market or polls predict the election outcome more closely. We conduct binomial tests comparing the poll predictions with market predictions for relative predictive accuracy on the last day the polls were in the field. Since poll results are not announced until at least the day after the completion of data collection and the market prices are as of midnight that night, the poll outcomes cannot be incorporated into market prices on the same date. The binomial variable is assigned a value of 1 if the market prediction is closer to the actual outcome than the poll prediction and 0 otherwise. The tests generally reject resoundingly the null that poll and market predictions are equally accurate in favor of the alternative that the market predictions are closer to the final outcomes.[14]

In summary, the evidence here shows that the markets are accurate months in advance and do a markedly better job than polls at longer horizons. However, to date, no “margin of error” or forecast standard error has been developed for the markets. We will address this in the next section.

V. Generating Forecast Standard Errors for Prediction Markets

As discussed in Section I, we believe that prediction markets such as political stock markets should predict phenomena accurately because of (1) the specific forecasting nature of such markets, (2) the financial incentive of such markets and (3) the information aggregation process of such markets. The previous research along with evidence presented in the last section shows the accuracy achieved ex post. Here we ask whether we can evaluate a market's likelihood to predict accurately ex ante. We discuss three possible means of establishing confidence intervals around the forecasts of the markets. First, we apply a short-run model based on market microstructure factors and calibrated on previous markets and election results. Then, we discuss a longer-run time series approach based on the behavior of prices in a current election market. Finally, we discuss an "implied

[14] The only exceptions are when few polls cause loss of power (during the last five days of each election) and during the

volatility" approach that parallels option pricing techniques and uses data from simultaneous winner-takes-all election markets.

A. A Model Based on Market Micro Structure Factors

Berg, Forsythe and Rietz (1997) develop regression models that explain the relative predictive accuracy of the 16 vote share markets run on the IEM between 1988 and 1994. Explanatory variables were inspired by considering market microstructure factors that should indicate levels of certainty, disagreement and/or information aggregation in markets. Their Model I uses three independent variables: (1) the number of party/candidate associated contracts traded in the market, (2) the total dollar volume of trading during the seven days before the election and (3) the average difference in weighted bid/ask queues at midnight before the election.[15] The estimated model is:

$$AAE = 0.693734 \times n - 0.009712 \times Vol + 0.0515215 \times DQueue,$$

where AAE is the market's average absolute prediction error, n is the number of party/candidate contracts traded, Vol is the total seven-day dollar volume and DQueue the average difference in weighted bid/ask queues. In sample, the standard deviation of the error in predicting average absolute error is 0.86%.

Since Berg, Forsythe and Rietz (1997), the IEM has run two presidential election markets. As a first measure of how likely these markets were to have predicted well, we generate the predicted absolute error using this model. Table 4 gives this information along with the information for the markets in Berg, Forsythe and Rietz for comparison purposes. The 1996 and 2000 election markets are typical in having 2 and 3

August 15 to October 15 period of 1988.

[15]The first two variables are measured in the obvious manner. The third is measured as follows: The weighted bid queue is given by the average (across contract types) of the sum of all closing bids weighted by (1) the dollar quantity committed at that bid and (2) the bid as a percentage of the best bid. This weighting reflects the commitment and seriousness of the bid by including its size and closeness to the market. The weighted ask queue is given by the average (across contract types) of the sum of all closing asks weighted by (1) the dollar quantity committed at that ask and (2) one minus the ask over the one minus the best ask. (The latter weighting reflects closeness to the market.) The difference in weighted queues is the absolute difference in these two queue measures. See Berg, Forsythe and Rietz (1997) for justification and details regarding these measures.

party/candidate contracts, respectively. They are relatively high volume markets, especially the 2000 market. This had the highest final week volume of any US vote-share market. These factors serve to reduce the predicted error relative to the average market. However, the differences in election eve weighted queues are relatively high. This pushes up the predicted error.

While the model predicts about average performance for the 1996 election market, that market was the second worst of the 18 US election market run to date in terms of election eve predictive accuracy. The actual average absolute prediction error of 4.53% differed from the predicted error of 2.55% by 2.30 times the standard error of the model. So, this market did significantly worse than predicted by the model. In contrast, the model predicted poor performance in 2000, while the market actually performed about average. The actual average absolute prediction error of 1.96% differed from the predicted error of 8.58% by -7.70 times the standard error of the model. So, this market did significantly better than predicted by the model.[16]

Thus, the out-of-sample performance of this model is poor. In addition, this model was developed using election-eve data and, as a result, cannot be used to assess likely predictive accuracy more in advance of the election. Next, we turn to other means of measuring forecast standard error.

B. A Time Series Model

Here, we build time series models of the evolution of prices to estimate the forecast standard error for the election outcome given prices as of date t. To do this, consider the vote-share prediction of candidate i at date t, $p_{i,t}^{VS}$. Since there is no aggregate uncertainty in the market and the risk free rate is zero, the capital asset pricing model, arbitrage pricing theory and general equilibrium state contingent claim pricing all give the same result that this price is the market consensus expected future vote share.[17] Similarly, the spread between any two

[16]However, notice that this prediction of poor performance comes primarily from the large difference in weighted queues. This may have been caused by a system change between the 1996 and 2000 elections. A change in the market software system resulted in the 2000 election markets having a higher default expiration time for limit orders than in previous elections.

[17]See footnote 5 above for justification.

candidates *i* and *j* on any given day *t*, denoted $s^{VS}_{i-j,t} = p^{VS}_{i,t} - p^{VS}_{j,t}$, is the market consensus expected future spread. Because of problems caused by the bid/ask bounce (as discussed below), we calculate the prediction as the normalized mid-point of the midnight (i.e., "closing") bid and ask.

If the markets are efficient, current prices incorporate all information as of date *t* and, therefore, should be sufficient statistics for forecasting future prices and spreads. This makes for a very simple, efficient markets time series model for forecasting the future outcome. Since $s^{VS}_{i-j,t} = E_t\left(s^{VS}_{i-j,t+h} \mid I_t\right)$, we can rewrite the relationship as:

$$s^{VS}_{t+h} = s^{VS}_{t} + e_{t,h} \Leftrightarrow s^{VS}_{t+h} - s^{VS}_{t} = e_{t,h},$$

where $e_{t,h}$ is a mean zero error term that represents the *h* step ahead forecasting error at date *t* and we have dropped the contract subscripts for notational convenience. Thus, the evolution of spread forecasts should follow a first order auto-regressive AR(1) process (set *h*=1). In theory, it should have a unit root (i.e., the coefficient on the lagged spread should be one).[18]

Assuming a stationary time series allows us to estimate the variance of the forecasting error. In particular, decompose the h-step-ahead forecast error into the sum of one-step errors:

$$s^{VS}_{t+h} - s^{VS}_{t} = e_{t,1} + e_{t+1,1} + \ldots + e_{t+h-1,1}.$$

Efficient markets would require the errors to be independent. If the errors are independently and identically distributed, then the mean and standard deviation of the distribution can be estimated from sample data and the forecast and forecasting error at the election date, *T*, can be easily computed. Following Judge, et al (1982, ch. 25), the standard error of the forecast is:

$$\sqrt{E_t\left(s^{VS}_{t+h} - s^{VS}_{t}\right)^2} = \sqrt{h} \times s_{\,e},$$

[18] These results can also be derived from the law of iterated expectations since the information available at date t+h (I_{t+h}) includes all the information available at date t (I_t). That is $I_t \subseteq I_{t+h}$. If, at each point in time, the vote share represented by prices equals the expected final vote shares given the current information set, then the law of iterated expectations states that: $E(s_{t+h}|I_t) = E[E(S_T|I_{t+h})|I_t] = E(S_T|I_t) = s_t$.

where σ_e is the standard deviation of the one step ahead forecast error. Given this, we can compute +/-1.96 standard error confidence intervals around the forecast.

However, there is no guarantee that the markets will be efficient random walks. Instead, prices may show an over-reaction to news. De Bondt and Thaler (1985, 1987 and 1990) argue that both ordinary and expert traders tend to overreact to information. This would lead to negative correlations of price changes across time and mean reversion in prediction markets. Alternatively, prices may underreact to news. Abarbanell and Bernard (1992) and Chan, Jegadeesh and Lakonishok (1996) argue that naturally occurring financial markets adapt to new information slowly. This would lead to the opposite effect of underreaction in prediction markets. So, theory predicts a unit root in the evolution of spread forecasts while evidence from naturally occurring markets suggests that the coefficient on lagged spread forecasts may be less than one (overreaction) or greater than one (underreaction). We will ask whether we can distinguish coefficients that differ from one and, if so, adjust the forecasts and forecast standard error estimates accordingly. We will also ask whether correlated errors add moving average components to the evolution of spread forecasts from the market.

In practice, there are several problems in estimating the process underlying the evolution of spread forecasts from our markets. First, the markets are not of equal duration and, early in each market, there is low volume and excess volatility resulting from thin markets. We will make all the markets comparable by analyzing only the last one hundred days before each election. By this point in time, each market was operating with a reasonable number of traders and reasonably thick bid/ask queues. Second, Nankervis and Savin (1988) point out the difficulty of accurately assessing confidence intervals in small samples for AR(1) processes. We will solve this problem by bootstrapping as suggested by Savin and Nankervis (1996). Last, the bid/ask bounce will cause some negative auto-correlation and, therefore the appearance of mean reversion, in these markets. We will address this problem using what has become standard in the finance literature. We use as the market prediction the midpoint of the closing bid and ask instead of closing prices.[19] Thus, we will compute the spread forecasts based on the normalized difference in midpoints.[20]

Panel A in Table 5 shows the results of AR(1) regressions of the model $s_{t+1} = \alpha + \beta s_t + e_t$ for the last 100 days of each election. The intercept is never significantly different from zero. While the coefficient on the lagged spread term, β, falls below the null of 1 in 1992 according to robust confidence intervals, β is never significantly different from the null of 1 according to bootstrapped confidence intervals. Likelihood ratio tests to determine if adding a second autoregressive term or a moving average term improves explanatory power show no significance. Overall, spreads evolve according to AR(1) processes that appear to have some mean reversion according to point estimates. However, none can be distinguished statistically from efficient market, random walks.

Panels B and C in Table 5 show the forecasts of the elections at 1, 7, 14, 28, 56 and 84 days before each election along with *T*-*t*-step ahead forecasting standard errors for each horizon. Figures 7 through 10 show the forecast spread and confidence intervals based both on the estimated AR(1) process and a random walk. (Figures 8 through 10 also show a set of implied volatility confidence intervals and a fitted line through them that will be discussed in the next section.) If one were to "call" the 1988 election based on the point estimates of the AR(1) process and the forecasting standard errors (Panel B in Table 5 and the smallest confidence intervals in Figures 7 through 10), one would have called it with 95% confidence for Bush beginning 22 days before the election. One would have called it on the same date using the random walk (Panel C in Table 5 and the next smallest confidence intervals in Figures 7 through 10). One would have called the 1992 election for Clinton

---

[19] The closing midpoint is the midpoint between the best outstanding bid and ask at midnight.

[20] To see the size of this problem, consider the first-order autocorrelation of price and midpoint changes. Negative correlations could result from mean reversion or a bid/ask bounce. The correlation coefficients for both measures (over the last 100 days before the election) are:

Measure	1988	1992	1996	2000
Price Changes	-0.2598*	0.2021*	-0.2927*	-0.2150*
Midpoint Changes	0.0475	-0.1801	-0.1150	0.0782

where and "*" denotes a value that differs significantly from zero at the 95% level of confidence.

during 14 of the last 18 days using the AR(1) criterion and 6 of the last 17 days using the random walk criterion. One would have called the 1996 race for Clinton during the entire 100 days using the AR(1) criterion and from day 52 on using the random walk criterion. In 2000, one would have called the race for Gore from day 55 to day 49 using the AR(1) criterion. This is primarily due to the mean reversion in the point estimate reducing the forecast standard error. One would have called the race for Bush only at 1, 5 and 9-day horizons. Using the random walk criterion, one would have called the race for Bush only at 1 and 9 day horizons. Thus, according to the level of predictive uncertainty measured by the evolution of prices, 2000 was the closest race followed by 1992, 1988 and 1996. In fact, the actual spreads (Democrat minus Republican normalized vote-shares) were 0.20%, 6.92%, -7.91% and 9.23% in 2000, 1992, 1988 and 1996.

Next, we derive an alternative measure of predictive uncertainty and ask whether our traders act as if the evolution of prices follows a random walk by looking at the relationship between prices in the vote-share and winner-takes-all markets.

C. An Implied Volatility Model

In many elections, including the last three Presidential elections, we conduct both vote-share markets and winner-takes-all markets. As discussed above, the vote-share markets trade contracts with payoffs equaling the percentage of the vote taken by the associated candidates/parties. A price or a bid/ask midpoint can be interpreted as a point prediction of the vote share. The winner-takes-all markets trade contracts that pay off $1 if the associated candidate/party wins the popular vote and $0 otherwise. Thus, a price or bid/ask midpoint in the winner-takes-all market can be interpreted as the probability that the associated candidate receives the majority (or plurality) of the vote. These contracts constitute binary option contracts on the underlying vote share. They can also be used to infer the volatility surrounding the vote-share market's forecast using techniques similar to the implied volatility research on options markets in finance.

In principle, it is simple to use the vote-share and winner-takes-all markets together to determine the implied volatility of the vote-share forecast. If one considers the forecast of the (future, actual) vote share to be a distribution, the vote-share market gives its mean and the winner-takes-all market measures the probability of a specified range of outcomes (those in which the candidate wins). For a specified distribution, this allows one to infer characteristics of the distribution. (This is the same intuition that drives implied volatility research for options.) For example, consider a two-candidate race. The price (serving as the prediction) for that candidate in the vote-share market is the mean of the distribution of that candidate's forecast vote share. The price (again serving as the prediction) for that candidate in the winner-takes-all market is the probability that the vote share will exceed 50%. If the expectations about the vote share are (approximately) normally distributed, then an inverse normal function will give the variance that is consistent with the mean (from the vote-share market) and probability of exceeding 50% (from the winner-takes-all market). A similar inverse normal will give the variance that is consistent with a predicted vote spread (from the vote-share market) and probability of a positive or negative spread (from the winner-takes-all market). These variances can be used to assess the level of confidence that the market places in its own estimates.

We implement the implied volatility forecasts as follows. Consider forecasting the spread on date *t* (before the election, which occurs at time *T*) in a two-party race. Suppose that, conditional on information at time *t*, the vote-share spread at the time of the election, $s^{VS}_{i-j,T}$, is distributed $N(\mu_{s,t},\sigma_{s,t})$ and, so, $\frac{s^{VS}_{i-j,T}-\mu_{s,t}}{\sigma_{s,t}}$ is distributed N(0,1). The probability of candidate *i* winning the two-way popular vote race is given by:

$$P_t(s^{VS}_{i-j,T} > 0 \mid I_t) = P_t\left(\frac{s^{VS}_{i-j,T}-\mu_{s,t}}{\sigma_{s,t}} > \frac{-\mu_{s,t}}{\sigma_{s,t}} \middle| I_t \right).$$

This implies that:

$$\sigma_{s,t} = \frac{-\mu_{s,t}}{\Phi^{-1}\left[1-P_t\left(s^{VS}_{i-j,T} > 0 \mid I_t\right)\right]},$$

where $\Phi^{-1}[.]$ is the inverse standard normal function. The vote-share market provides an estimate of the expected spread: $\hat{\mu}_{s,t} = s^{VS}_{i-j,t}$. The winner-takes-all market provides an estimate of the probability of a Democratic win:

$\hat{P}_t\left(s^{VS}_{i-j,T} > 0 \mid I_t\right) = p^{WTA}_{i,t}$. Using these, we estimate standard deviation of the forecast vote spread distribution as:

$$\hat{s}_{s,t} = \frac{-s^{VS}_{i-j,t}}{\Phi^{-1}\left[1 - p^{WTA}_{i,t}\right]}.$$ [21]

In practice, several difficulties arise. First, bid/ask bounce can affect the implied volatilities. We will account for this as we did above, by using normalized spreads determined by bid/ask midpoints as our predictions.

Second, in markets with more than two contracts, the multivariate distributions of forecasts will depend on the correlation structure across candidates. Since we have only one observation of the mean for each candidate (from the vote-share market) and one other point on the distribution for each candidate (from the winner-takes-all market), we cannot pin down all of the parameters of the distribution without making additional assumptions. Of the three markets we consider, only the 1996 Presidential race has two contracts in both the vote-share and the winner-takes-all markets. In the 1992 race, there were two contracts in the major party vote-share market, one for Clinton and one for Bush. However, in the winner-takes-all market, there were three contracts, one each for Clinton, Bush and Perot. To accommodate the problem in this market, we make a simple assumption: the relative chances of Bush or Clinton receiving more than 50% of the Bush/Clinton vote was proportional to their relative chances of winning in the three way race. In the 2000 race, there were three contracts in both markets, one each for Bush, Gore and Buchanan (as the Reform party candidate). To accommodate the problem in this market, we look only at the Bush/Gore vote. We assume that their relative vote shares and relative probabilities of winning are independent of the Buchanan vote. This is not unreasonable given that Buchanan was predicted to take, and actually took, an extremely low percentage of the vote and, as a

[21] Note that this estimator becomes unstable as the expected vote shares or probabilities of winning approach 50%/50%. We will address this problem by fitting confidence intervals to the sample of implied volatilities below and return to this problem in the summary and conclusions.

result, had almost no chance of winning. A stronger third-party candidate would have created much more difficulty here.

Third, as with simple spread predictions, asynchronous movements in bids and asks across contracts and markets may affect the estimates. We attempt to control for this partially by normalizing predictions in each single market as stated above. However, because implied volatility relies on prices across two separate markets, it may be even more sensitive to these factors. The marginal traders in these two markets may differ from each other and, as a result, may reflect different expectations. Further, there is no natural way of normalizing across markets. Estimates may be especially sensitive to these factors in close races because the implied volatility estimator becomes unstable as the probabilities of winning approach 50/50. One might mitigate these effects by averaging over time. However, averaging suffers from its own problems when market prices are changing because “stale” bids and asks may be more influential in one market than another. We will address this problem using a unique method for fitting a process to the observed implied volatilities as discussed shortly.

Fourth, the implied volatility estimator is undefined for predicted probabilities of exactly 0.5 for each candidate and when one candidate is projected to receive more than 50% of the vote in the vote-share market while the other candidate is projected the more likely winner in the winner-takes-all market.[22] Using midpoints helps mitigate both problems. We have to drop observations to address remaining problems. We drop observations when the winner-takes-all midpoints are within one mil of each other because our pricing grid only goes down to mils. This eliminates one observation (September 3, 2000). We also drop 15 observations in 1992, 0 in 1996 and 10 in 2000 (out of 100 in each year) because the estimator is undefined due to inconsistencies in the vote-share and winner-takes-all markets. In the next paragraph, we describe a unique fitting method that averages out the remaining instability resulting from winner-takes-all predictions that approach 50%.

[22]This is inconsistent with a symmetric distribution. Though it might be accounted for by the right kinds of asymmetries in forecast distributions, it can also arise easily from asynchronous trading when prices are moving.

As discussed above, Figures 8 through 10 show the predicted spread (Democratic vote share minus Republican vote share) for the 1992, 1996 and 2000 Presidential election markets using bid/ask midpoints to generate the spread predictions. In addition to the AR(1) and random walk confidence intervals discussed above, they also show "confidence" intervals calculated by (1) finding the implied standard deviation of the spread given the prices in the two markets and (2) adding to or subtracting from the spread prediction 1.96 times this standard deviation. The graphs illustrate the relatively large deviations in day-to-day implied volatility forecasts (denoted with "+" signs). The problem worsens for close races. It results from small changes in relative prices when the inverse normal function in the denominator of the implied volatility formula approaches zero as discussed above.[23]

Because of the day-to-day variation in the implied volatility forecasts, we have "averaged" them by fitting a line through the implied volatility confidence intervals. The line is fitted using a non-linear least squares estimation (NLLS) technique to estimate the parameters of an AR(1) process using the following steps. (1) Assume that the spreads follow an AR(1) process with an unknown autoregressive parameter and unknown standard deviation for the daily errors. That is, $u_{t+1} = qu_t + f_t$, where the u's represent the (unobservable) anticipated spread forecasts used by traders in the winner-takes-all market and f_t, is an error term for the unknown process. (2) Compute the T-t-step ahead forecasting variance of this AR(1) process conditional on q, s_f and t. Denote it by $s^2_{q,f,t} := s^2_f \sum_{h=1}^{T-t} q^{2(h-1)}$. (3) Hypothesize that this equals the implied volatility variance, denoted $s^2_{IV,t}$. (4) Use $s^2_{IV,t} := s^2_f \sum_{h=1}^{T-t} q^{2(h-1)} + n_t$ to estimate $s_f{}^2$ and q through NLLS regression.

The result of this regression is the fitted AR(1) process that gives standard deviations of forecasting errors that most closely resemble the volatilities revealed by traders according to the implied volatility model.

[23] We also note that each of the four elections apparently results in increasing uncertainty in the last week or two before the election according to implied volatility forecasts. This may reflect residual uncertainty as the election approaches that is not allowed by an AR(1) parameterization. A complete analysis of this phenomenon is beyond the scope of this paper, but we hope to investigate it in further research.

Confidence intervals from these fitted AR(1) processes are also plotted in Figures 8 through 10. The fitted ? parameters are 1.0149, 1.0028 and 1.0093 for 1992, 1996 and 2000, respectively. The NLLS standard errors of these estimates are 0.0068, 0.0014 and 0.0073, respectively. Bootstrapped NLLS standard errors are 0.0029, 0.0014 and 0.2507, respectively.[24] Notice that the estimated parameters, s_f^2 and q, depend only on the implied volatilities, not directly on the observed vote share price process. A natural question is whether these estimated processes differ significantly from the observed processes in the corresponding vote share market. One can construct z-statistics for the hypothesis that these estimates differ from the estimated AR(1) process given in Table 5. Using the bootstrapped standard deviations for both estimates, these test statistics are -0.98, -0.75 and -0.23, respectively. Thus, trader behavior across the two markets leads to implied volatilities that are consistent with the volatilities estimated from the actual time series estimates of the spreads. So, one can use the probability predictions of the winner-takes-all markets as reasonable indicators of the level of confidence that traders place in the vote-share forecasts. Further, ex-post, these accord with the actual uncertainty revealed as the market develops.

VI. Summary and Conclusions

Whether markets can be designed specifically and used successfully as forecasting tools forms the basis for an exciting new line of economic market research. The evidence to date suggests that such "prediction markets" can be used to aggregate diverse and complex information to predict specific phenomenon such as election outcomes or corporate sales. This can be an important advance in forecasting because markets aggregate information in a complex, dynamic way that traditional forecasting methods cannot.

[24] The high standard error in the bootstrapped 2000 data comes from a few outliers. These result from estimating the volatility when the vote share is very close to 50%/50%, where the estimator blows up. Dropping the 4 largest (outlier) observations drops the bootstrapped standard error 0.0135. (The z-statistic remains insignificant.) This suggests one needs to add additional WTA markets with different cutoff vote shares to get a more precise estimate of volatility in close races.

However, two things are necessary for a good forecasting tool. First, the forecasts should be relatively accurate and unbiased. The prediction markets for elections studied here give accurate forecasts at both short and long horizons and in both absolute terms and relative to polls. While prior evidence shows the very short run accuracy of the markets (e.g., Berg, Forsythe, Nelson and Rietz, 2003), the important contribution of this paper is to document the long-run accuracy (weeks and months in advance). Here, we also show that these markets yield price-based predictions consistent with efficient markets and efficient forecasting. The time series of prices cannot be distinguished from efficient random walks and inter-market pricing relationships are also consistent with efficient random walks. Therefore, the price on any given day serves as the best predictor of future prices and the ultimate outcome.

Second, one should be able to measure the forecast standard error of any given forecast. In contrast to more traditional forecasting methods, measures of forecast standard error are not apparent for prediction markets. The difficulty arises because prices (and, therefore, predictions) are not simple averages of random samples. Therefore, cross-sectional techniques (along the lines of polls) cannot be used. Market prices/predictions are also not typical time series of fundamental variables. Instead, they are a sequence of forecasts of a single future outcome. We exploit properties of efficient markets and efficient forecasts to generate the first known useful measures of forecast standard errors for prediction markets. The first measure comes from analyzing the time series of forecasts from prices and projecting forecast standard errors from the inter-day volatility of the forecast sequence. A second measure derives from an inter-market pricing relationship that gives implied volatilities. These can be computed along lines similar to those used for ordinary financial options. They can also serve to project the forecast standard errors of prediction markets. In the markets studied here, these two estimates of forecast standard error do not differ significantly. We suggest using both to assess the forecast standard error and, therefore, the likely predictive accuracy of prediction markets.

Finally, the behavior of the implied volatility forecasts from the particular markets studied here leads to an additional recommendation. Recall that the vote-share markets should give the means of the forecast distributions of vote shares in the upcoming elections. The winner-takes-all markets give the probabilities of taking the majority of the popular vote. Thus, the markets studied here estimate the variance of the forecast distribution by identifying the mean of the distribution and the area in a tail. When the cutoff for the tail equals the mean, the variance of the distribution cannot be identified and, when it is close, the variance cannot be estimated with much precision. We suggest that more precision could be attained by running additional markets designed to identify more points on the distribution.[25]

[25]We note the IEM ran markets in 1992 designed to predict the probabilities that the vote shares would fall in particular ranges (e.g., greater or less than 55% of the vote). However, these markets were exceptionally thin, resulting in extremely low volumes and, likely, unreliable prices. In future elections, we hope to design markets that will generate sufficient volumes and thick enough queues to test whether we can get additional precision by identifying additional points on the forecast distribution.

References

Abarbanell, JS, and VL Bernard, 1992, “Tests of Analysts’ Overreaction/Underreaction to Earnings Information as an Explanation for Anomalous Stock Price Behavior,” Journal of Finance, 47, 1181-1207.

Berg, JE, R Forsythe and TA Rietz, 1997, “What Makes Markets Predict Well? Evidence from the Iowa Electronic Markets,” in W Albers, W Güth, P Hammerstein, B Moldovanu and E van Damme (eds.), Understanding Strategic Interaction: Essays in Honor of Reinhard Selten, New York: Springer, 444-463.

Berg, JE, R Forsythe, FD Nelson and TA Rietz, 2003, “Results from a Dozen Years of Election Futures Markets Research,” forthcoming in CA Plott and V Smith (Eds) Handbook of Experimental Economic Results.

Chan, LKC, N Jegadeesh and J Lakonishok, 1996, “Momentum Strategies,” Journal of Finance, 51, 1681-1713.

Forsythe, R, FD Nelson, GR Neumann and J Wright, 1992, “Anatomy of an Experimental Political Stock Market,” American Economic Review, 82:1142-1161.

Forsythe, R, TA Rietz and TW Ross, 1999, “Wishes, Expectations and Actions: Price Formation in Election Stock Markets,” Journal of Economic Behavior and Organization, 39: 83-110.

Hayek, FA, 1945, “The Use of Knowledge in Society,” The American Economic Review, 35, 519-530.

Hicks, JR, 1946, Value and Capital, 2nd ed., London: Oxford University Press.

Judge, GG, RC Hill, WE Griffiths, H Lütkepohl and TS Lee, 1982, Introduction to the Theory and Practice of Econometrics, John Wiley and Sons, New York.

Keynes, JM, 1930, Treatise on Money, 2nd ed., London: Macmillan.

Krueger, JT and KN Kuttner, 1996, “The Fed Funds Futures Rate as a Predictor of Federal Reserve Policy,” Journal of Futures Markets, December 1996.

Malinvaud, E, 1974, "The Allocation of Individual Risks in large Markets," in J. H. Dréze, Ed., Allocation under Uncertainty: Equilibrium and Optimality, MacMillan Press, London, England, 110_125.

Nankervis, JC, and NE Savin, 1988, “The Student’s t Approximation in a Stationary First Order Autoregressive Model,” Econometrica, 56, 119-145.

Plott, CR, 2000, “Markets as Information Gathering Tools,” Southern Economic Journal, 67, 1-15.

Savin, NE and JC Nankervis, 1996, “The Level and Power of the Bootstrap t-Test in the AR(1) Model with Trend,” The Journal of Business and Economic Statistics, 14, 161-168.

Sunder, S, 1995, "Experimental Asset Markets: A Survey," in JH Kagel and AE Roth, eds, Handbook of Experimental Economics, Princeton NJ: Princeton University Press, 445-500.

Worcester, R, 1996, “Political Polling: 95% Expertise and 5% Luck," Journal of the Royal Statistical Society, A, 159, Part 1, pp.5-20.

Appendix

This appendix contains descriptions of the individual election market dynamics for the interested readers. Figures 1 through 4 contain the raw price series for each of the four Presidential election markets.

In 1988 (Figure 1), market prices implied an early Dukakis lead followed by a relatively close race through to the last debate. (However, early numbers are based on very low volumes.) Bush took a slight lead in early September and pulled well ahead in mid-October on heavy volume. Volume spiked with the second presidential debate. Shortly, Bush prices rose and Dukakis prices fell to near the actual election outcomes and remained there throughout the rest of the market. The average absolute prediction error never exceeded 1.5% from October 14 through the election.

In 1992, the election predictions showed much greater volatility (Figure 2). Early volatility and volume may have stemmed from uncertainty over the Democratic nominee early in the market. During the first week of June, volume spiked while Bush prices fell and Clinton prices rose. We cannot identify a specific cause for this. Ross Perot was studying a possible run during this time, but announced he would not run on the last day of the Democratic convention. Apparently, this surprised the market. However, Perot's price began an immediate slow recovery and barely moved when he re-entered the race. The Bush-Clinton race was relatively close, but more volatile than 1988, through the conventions and up to the debates. Clinton pulled ahead during the debates. As in 1988, prices approached the actual outcomes shortly after the last debate and remained close through the election. While there was more volatility than in 1988, the average absolute prediction error (across both markets) averaged less than 2% over the last three weeks of the election. The winner-takes-all prices (Figure 5) mirrored these movements, with rising Clinton prices from the conventions through the election with the exception of a downturn followed by a recovery just before the election.

The 1996 race (Figure 3) proved relatively dull. Volumes were very low throughout the race. Early volatility and volume may have arisen from uncertainty over the Republican nominee. From the point Dole locked up the nomination, Clinton led. Other than two movements on very low volume in late July, prices remained quite stable from March through the election. In fact, due to a few late trades, the closing prices on election eve had the highest forecasting error of *any* closing price since February 28. Again, in the weeks leading up to the election, the average absolute prediction error was small. In this case, it averaged 2.3% over the final three weeks. Throughout the race, Clinton led and generally gained ground in the winner-takes-all market (Figure 5).

The 2000 race (Figure 4) was the tightest race of the four in terms of the market prediction across time and the actual election outcome. During the early primaries, the race appeared close. Volume spiked when McCain suspended his campaign (March 9) and Gore pulled ahead until McCain announced his support for Bush (May 9). The race appeared extremely close from then until September, when Gore opened a slight lead. Through the debates, Gore's lead evaporated and Bush pulled ahead slightly. In the final days before the election, the race became extremely close again. Three days before the election, the market was predicting a slight lead for Bush. Two days before the election, the lead switched to Gore. On election eve, the lead switched back to Bush. In contrast, polls showed large leads for Bush from April through mid August, a relatively large Gore surge until mid September and a relatively large Bush resurgence from then through the election. Again, the market proved accurate in the weeks leading up to the election. With one exception (caused by the slight upward spike in the Reform contract on November 4), the average absolute prediction error never exceeded 2% from September 20 through the election. The winner-takes-all prices (Figure 5) also showed the closeness of this race. Predictions for both major candidates hovered near 50% until mid-September. Then a Gore lead preceding the debates was followed by a Bush surge through the debates and a lead after them. However, even at its greatest, the separation here was much smaller than in the other winner-takes-all markets. Thus, throughout the election, the IEM showed a close race in absolute terms and closer race than polls did. In the end, Gore took the popular vote by 0.2% (according to payoff-determining preliminary returns) in the closest race since the 1960 race between Kennedy and Nixon.[26]

[26]The IEM contracts are based purely on popular vote as reported by the New York Times shortly after the elections. In the

Table 1
Contracts Traded in Presidential Election Markets

Year	Winner-Takes-All Market	Vote-share Market[1,2]
1988	None	Bush
		Dukakis
		Jackson
		Rest-of-field
1992	Democrat (P.CL)	Democrat (D.CL)
	Republican (P.BU)	Republican (R.BU)
	Other (P.PE)	Perot (PERO)
		Democrat and Republican (D&R)
1996	Clinton (CLIN)	Democrat (V.CLIN)
	Other Democrat (OTDEM)	Republican (V.DOLE)
	Republican (DOLE)	
	Other (ROF96)	
2000	Democrat (DemVS)	Democrat (Dem)
	Reform (ReformVS)	Reform (Reform)
	Republican (RepVS)	Republican (Rep)

[1]This table lists the vote-share contracts outstanding on election eve. When vote-share markets first open, they may include considerably more contracts representing other candidates that subsequently drop out of the race.
[2]The structure of the vote-share market changed in 1992 due to difficulty in verifying third party votes.

Table 2: Summary of Market Activity

	1988	1992[1]	1992[1]	1996[1]	1996[1]	2000[1]	2000[1]
	Vote-share	Vote-share	Winner-Takes-All	Vote-share	Winner-Takes-All	Vote-share	Winner-Takes-All
Market Duration							
Opening Date	6/1/88	1/10/92	7/10/92	2/4/96	11/15/94	12/29/99	5/1/00
Election Date	11/8/88	11/3/92	11/3/92	11/5/96	11/5/96	11/7/00	11/7/00
Weeks Open	23	43	17	39	103	45	27
Trader Investments							
Total	$4,976	$79,356		$200,000[2]		$148.000	
Minimum	$5	$5		$5		$5	
Median	N.A.	$25		$100		$10	
Maximum	$420	$500		$500		$1005	
Activity over Entire Market							
No. of Active Traders	155	592	471	264	1151	802	965
Trading Volume							
Contract Volume	15,826	78,007	215,585	23,093	652,165	46,820	262,587
Dollar Volume	$8,123	$21,445	$51,316	$3,628	$137,386	$17,576	$130,058
Activity over Last Week of Each Market							
No. of Active Traders	54	114	237	41	216	104	355
Trading Volume							
Contract Volume	962	1,389	59,836	592	48,243	4,192	44,752
Dollar Volume	$1,924	$569	$10,858	$312	$6,027	$1,396	$21,395

[1]Traders may be active in both markets.
[2]Estimated.

Table 3: Binomial Tests for Relative Accuracy of the Market and Contemporaneous Poll Predictions.

Poll predictions come from major polls taken during the election and are the normalized two-party vote shares. The market predictions are the normalized two-party vote share predictions on the last day each poll was in the field collecting data. The binomial variable takes the value 1 if the market prediction is closer the actual election outcome and 0 otherwise. Each p-value is the exact binomial probability of a number of 1s that large or larger, given that number of trails and a hypothesized probability of 0.50.) The number of observations is the number of polls in the sample period. If multiple polls are released on the same day, the same market price is compared to each poll.

Days included in sample	Item	1988	1992	1996	2000	all years
All (from the beginning of the market	Number of polls	59	151	157	229	596
	poll "wins"	25	43	21	56	145
	market "wins"	34	108	136	173	451
	% market	58%	72%	87%	76%	76%
	p-value (1sided)	0.148	0.000	0.000	0.000	0.000
Last 100	Number of polls	45	82	124	180	431
	poll "wins"	24	23	18	54	119
	market "wins"	21	59	106	126	312
	% market	47%	72%	85%	70%	72%
	p-value (1sided)	0.724	0.000	0.000	0.000	0.000
Last 65	Number of polls	34	62	91	141	328
	poll "wins"	19	15	15	52	101
	market "wins"	15	47	76	89	227
	% market	44%	76%	84%	63%	69%
	p-value (1sided)	0.804	0.000	0.000	0.001	0.000
Last 31	Number of polls	21	40	58	84	203
	poll "wins"	7	7	13	26	53
	market "wins"	14	33	45	58	150
	% market	67%	83%	78%	69%	74%
	p-value (1sided)	0.094	0.000	0.000	0.000	0.000
Last 5	Number of polls	6	6	11	25	48
	poll "wins"	0	1	4	8	13
	market "wins"	6	5	7	17	35
	% market	100%	83%	64%	68%	73%
	p-value (1sided)	0.016	0.109	0.274	0.054	0.001

Table 4: Berg, Forsythe and Rietz (1997) Model I Predictions
and Actual Average Absolute Prediction Errors

The fitted regression model from Berg, Forsythe and Rietz (1997) is:

$$AAE = 0.693734 \times n - 0.009712 \times Vol + 0.0515215 \times DQueue,$$

Where AAE is the average absolute prediction error across candidates/contracts, n is the number of contracts in the market, Vol is the dollar volume over the last week and Dqueue is the difference in the election eve weighted bid/ask queues. The first two variables are measured in the obvious manner. The third is measured as follows: The weighted bid queue is given by the average (across contract types) of the sum of all closing bids weighted by (1) the dollar quantity committed at that bid and (2) the bid as a percentage of the best bid. This weighting reflects the commitment and seriousness of the bid by including its size and closeness to the market. The weighted ask queue is given by the average (across contract types) of the sum of all closing asks weighted by (1) the dollar quantity committed at that ask and (2) one minus the ask over the one minus the best ask. (The latter weighting reflects closeness to the market.) The difference in weighted queues is the absolute difference in these two queue measures. See Berg, Forsythe and Rietz (1997) for justification and details regarding these measures

Market	n	Vol	DQueue	Actual AAE	Predicted AAE	Difference
1992 Illinois Primary	6	76.13	29.8	5.30	4.96	0.34
1992 Michigan Primary	6	64.58	100.58	8.59	8.72	-0.13
1992 Presidential (Perot versus Others)	2	205.88	38.43	0.34	1.37	-1.03
1992 Presidential (Major Party Candidates)	2	256.97	15.64	0.06	_0.31	0.37
1994 AZ Senate	3	7.30	13.22	1.57	2.69	-1.12
1994 NJ Senate	3	19.03	7.12	2.26	2.26	0.00
1994 NY Governor	4	11.15	5.1	4.09	2.93	1.16
1994 PA Senate	3	3.43	5.72	1.11	2.34	-1.23
1994 TX Governor	3	55.80	5.76	1.09	1.84	-0.75
1994 TX Senate	3	23.75	1.66	3.27	1.94	1.33
1994 US House Seats	3	18.02	36.04	3.07	3.76	-0.69
1994 US Senate Seats	3	4.81	25.25	3.25	3.34	-0.09
1994 UT House	4	17.12	2.57	3.56	2.74	0.82
1994 VA Senate	5	93.32	3.6	1.99	2.75	-0.76
1990 IA Senate	2	250.82	30.92	1.16	0.54	0.62
1990 IL Senate	2	20.89	56.72	5.21	4.11	1.10
In Sample Averages	3.375	70.56	23.63	2.87	2.87	0.00
1996 US Presidential	2	311.80	81.26	4.53	2.55	1.98
2000 US Presidential	3	598.82	239.03	1.96	8.58	-6.62

Table 5 AR(1) Regression Models for the Evolution of the Predicted Vote Spreads over the last 100 Days Preceding the Elections

The OLS regressions are on the model: $s_{t+1} = \alpha + \beta s_t + \varepsilon_t$, where s represents the vote spread measured by the difference in the bid/ask midpoints for the contracts on the Democratic and Republican vote shares. The robust standard errors use the Huber-White correction and the bootstrap standard errors are based on 1000 repetitions.[*] The likelihood ratio tests use ARMA models with one moving average term. The first tests the significance of adding a second moving average term and the second tests the significance of adding an autoregressive term. There are no substantive differences between results using the Huber-White standard errors and Newey-West standard errors with lags of up to one week.

Panel A: Regression Estimates

	Election Year			
Item	1988	1992	1996	2000
Constant	-0.0011	0.0031	0.0077	0.0001
Robust Std. Err.	0.0007	0.0014	0.0056	0.0009
Lagged Spread	0.9874	0.9282[**]	0.9312	0.9480
Robust Std. Err.	0.0189	0.0316	0.0547	0.0411
Bootstrap Std. Err.	0.0185	0.0882	0.0948	0.0870
N	100	100	100	100
Adj. R^2	0.9562	0.8852	0.8457	0.8774
Model Sigma	0.0060	0.0151	0.0074	0.0083
LR-Test for MA Term	0.985	3.31	0.89	1.03
LR-Test for 2nd AR Term	0.986	1.78	0.73	0.93

Panel B: Forecasts and Forecast Standard Errors using Point Estimates of the AR(1) Process

	Election Year							
Days to Election	1988		1992		1996		2000	
	s_t	σ_t	s_t	σ_t	s_t	σ_t	s_t	σ_t
1	-7.72%[+]	0.60%	6.07%[+]	1.51%	13.83%[+]	0.74%	-3.31%[+]	0.83%
7	-6.53%[+]	1.54%	4.09%	3.42%	15.86%[+]	1.58%	-2.87%	1.87%
14	-6.10%[+]	2.08%	11.17%[+]	4.12%	14.70%[+]	1.83%	-1.41%	2.23%
28	-2.07%	2.71%	5.05%	4.53%	10.57%[+]	1.93%	0.97%	2.43%
56	-1.22%	3.31%	2.57%	4.62%	9.97%[+]	1.95%	3.90%	2.47%
84	0.22%	3.57%	-1.20%	4.62%	9.14%[+]	1.95%	0.15%	2.47%

Panel C: Forecasts and Forecast Standard Errors using a Random Walk

	Election Year							
Days to Election	1988		1992		1996		2000	
	s_t	σ_t	s_t	σ_t	s_t	σ_t	s_t	σ_t
1	-7.72%[#]	0.60%	6.07%[#]	1.51%	13.83%[#]	0.74%	-3.31%[#]	0.83%
7	-6.53%[#]	1.59%	4.09%	4.07%	15.86%[#]	1.99%	-2.87%	2.25%
14	-6.10%[#]	2.25%	11.17%	5.75%	14.70%[#]	2.82%	-1.41%	3.18%
28	-2.07%	3.18%	5.05%	8.14%	10.57%[#]	3.99%	0.97%	4.50%
56	-1.22%	4.50%	2.57%	11.51%	9.97%	5.64%	3.90%	6.37%
84	0.22%	5.51%	-1.20%	14.09%	9.14%	6.91%	0.15%	7.80%

[*]There are no substantive differences between results using the Huber-White standard errors and Newey-West standard errors at lags up to a week.

[**]Significantly different from the null (constant=0, coefficient on lagged spread = 1, equivalent log-likelihoods) according to robust confidence intervals.

[+]Spread significantly different zero according to AR(1) point estimate confidence intervals.

[#]Spread significantly different zero according to random walk confidence intervals.

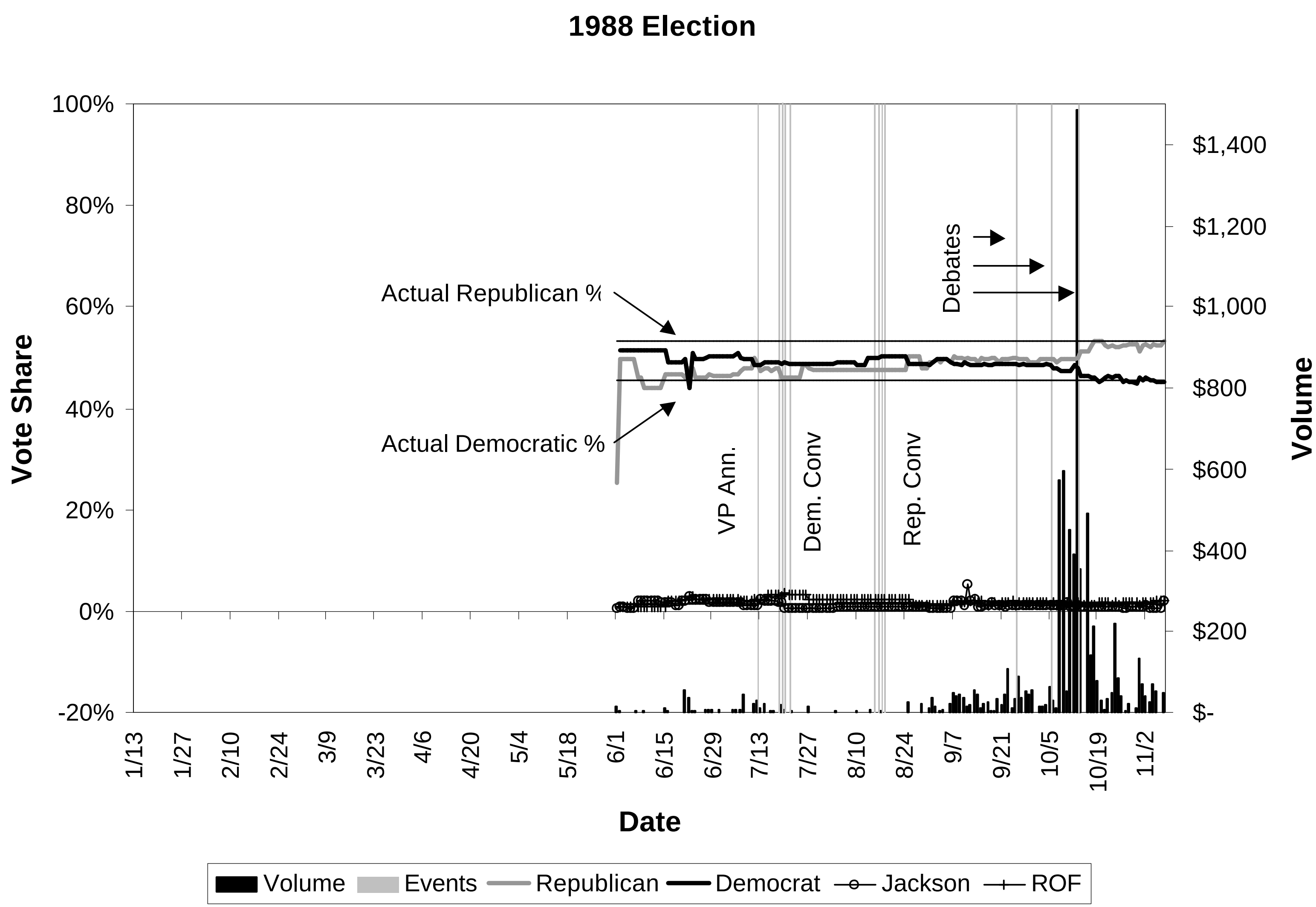

Figure 1: Predicted Vote Shares (Non-normalized) and Volumes in the 1988 Presidential Vote-share Market

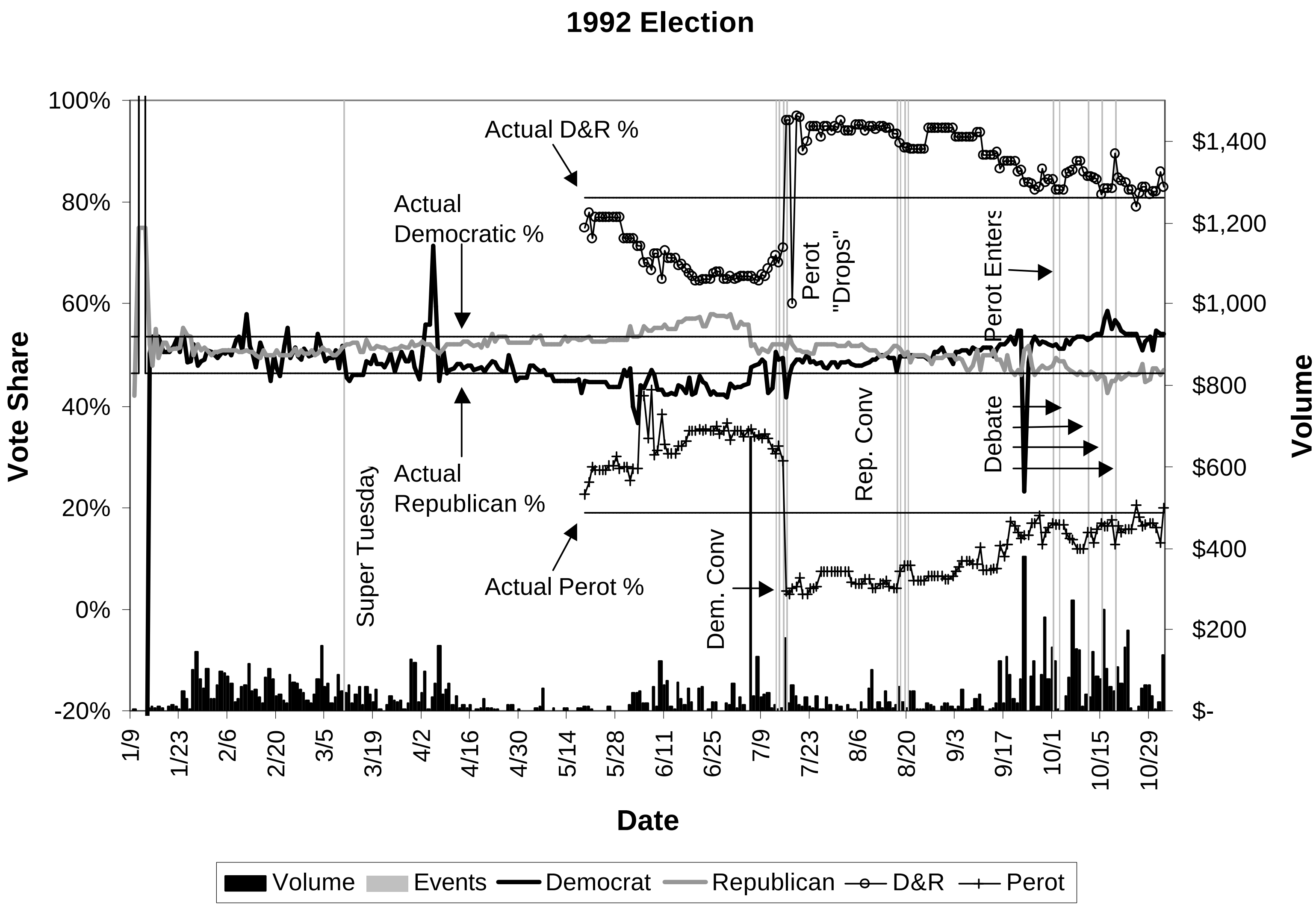

Figure 2: Predicted Vote Shares (Non-normalized) and Volumes in the 1992 Presidential Major Party Vote-share Market

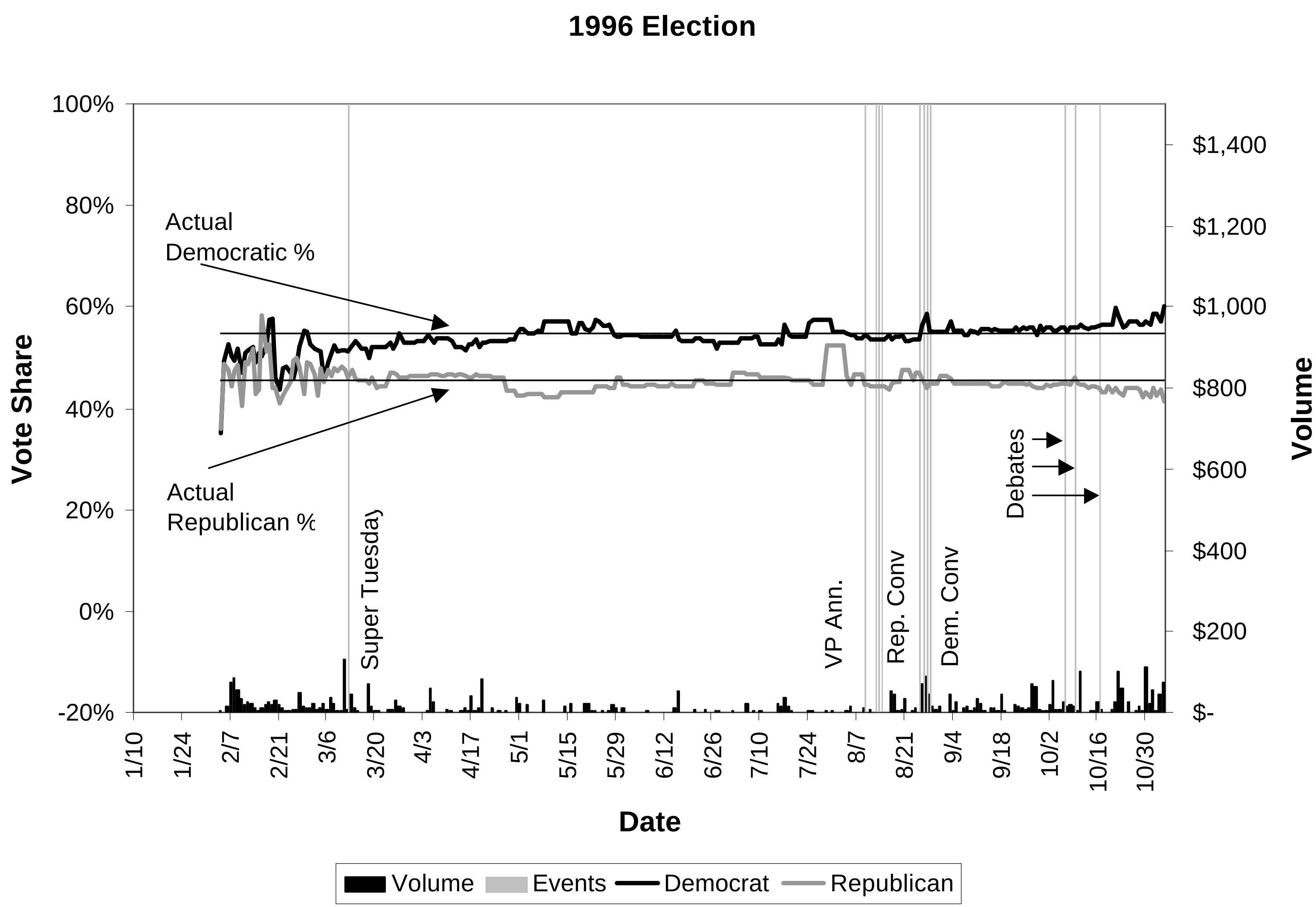

Figure 3 Predicted Vote Shares (Non-normalized) and Volumes in the 1996 Presidential Vote-share Market

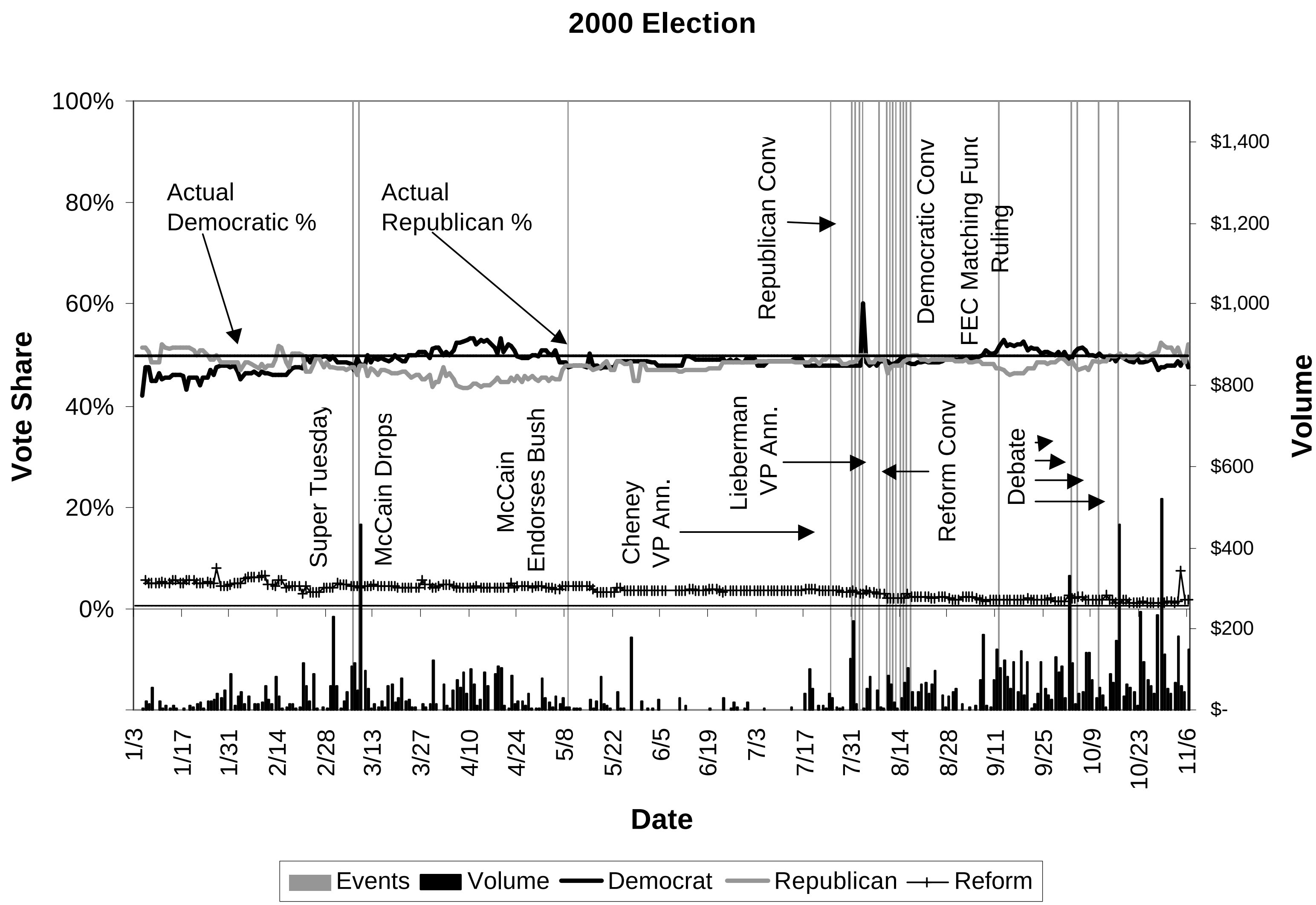

Figure 4 Predicted Vote Shares (Non-normalized) and Volumes in the 2000 Presidential Vote-share Market

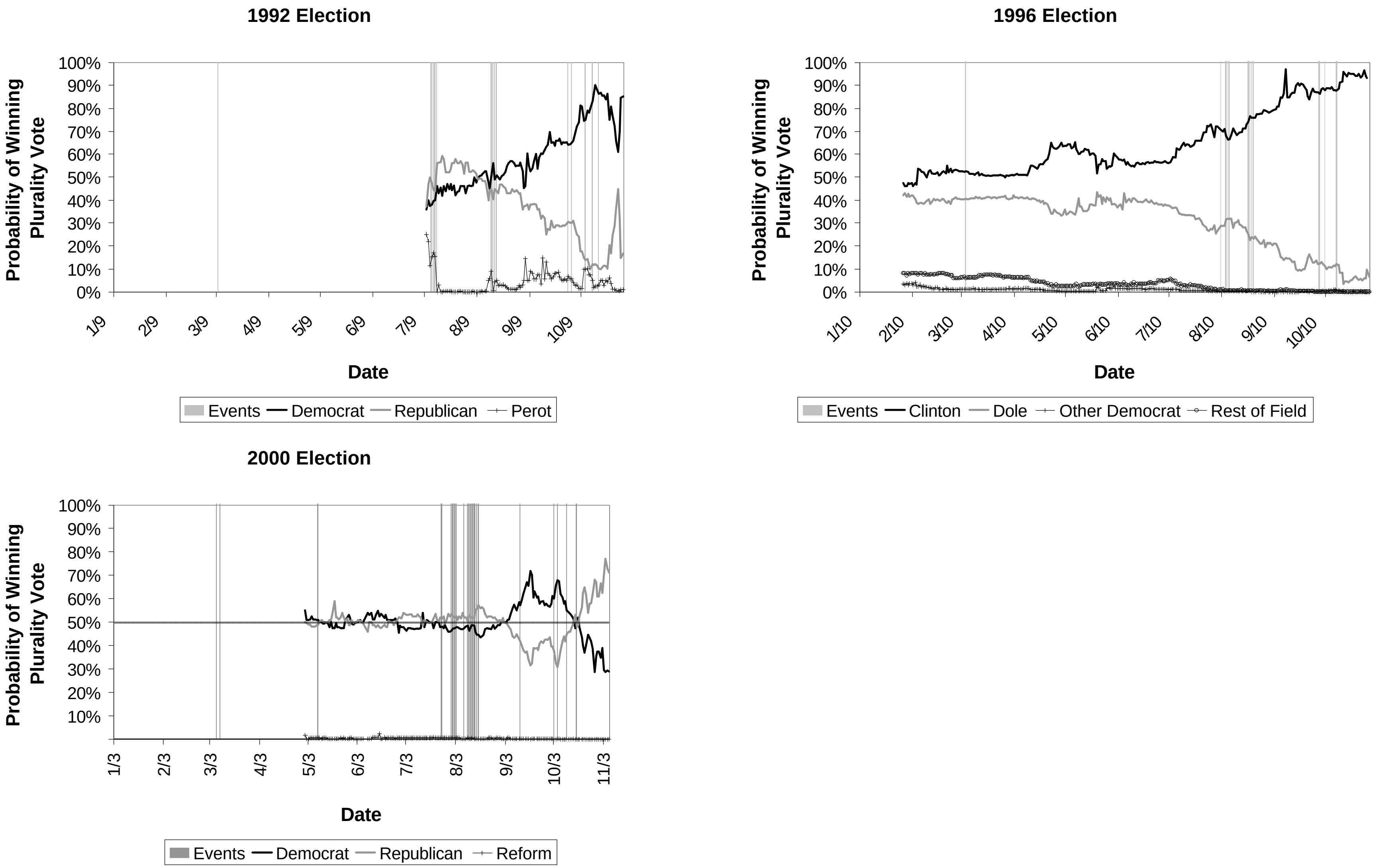

Figure 5: Presidential Winner-Takes-All Market Predictions on Probabilities of Winning the Popular Vote for the 1992, 1996 and 2000 Elections

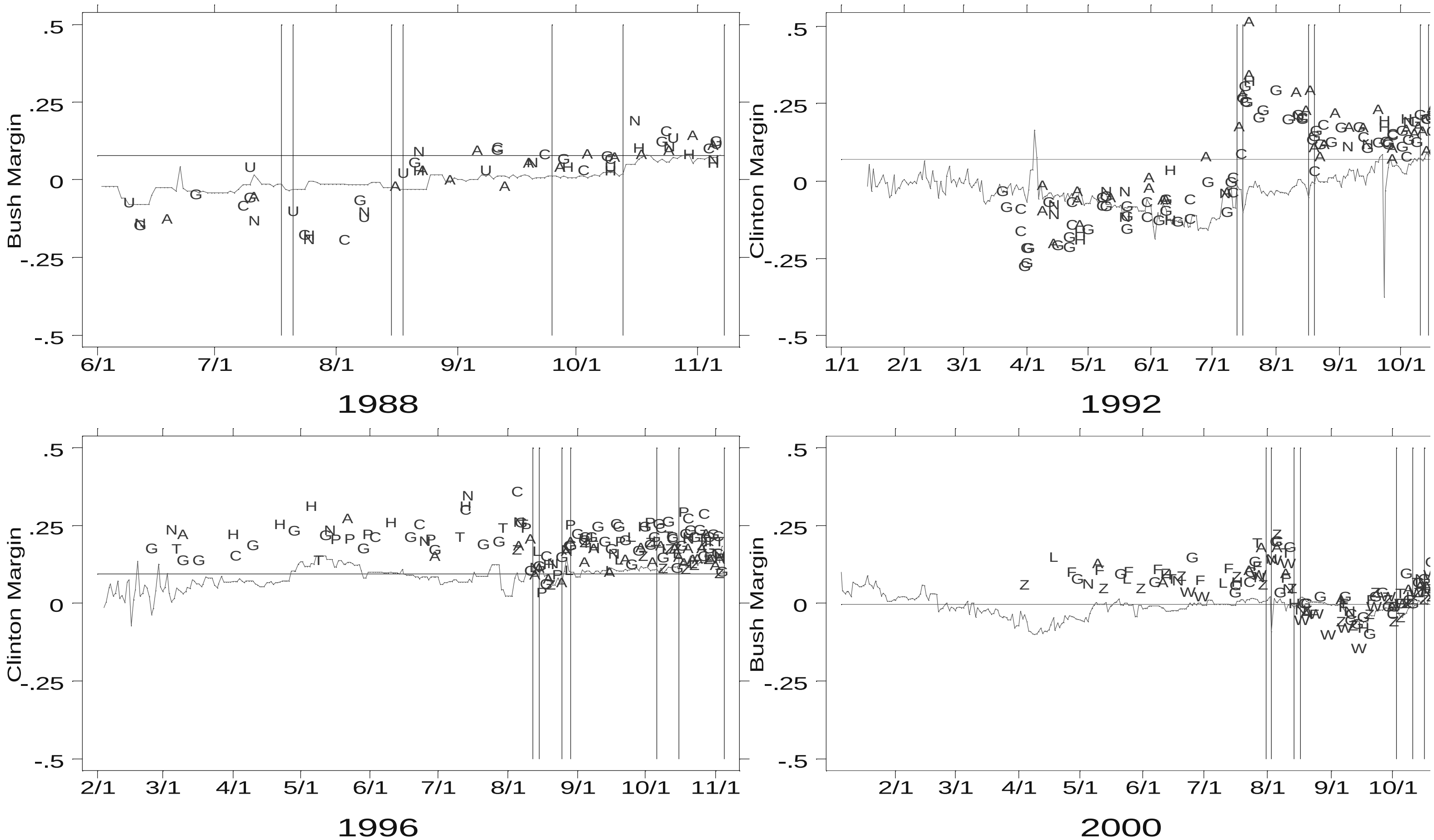

Figure 6: Implied Vote Share Margins for Market and Polls. The vertical axis is the vote margin for the winning candidate. The horizontal axis is the date. The solid moving line is the (normalized) margin from market prices. Letters represent the (normalized) margins implied by polls (A=ABC, C=CBS, T=CNN, F=Fox, G=Gallup, H=Harris, L=Hotline, N=NBC, W=Newsweek, Z=Zogby). The horizontal line is the actual election margin. The vertical lines show the dates of convention, debates and Election Day.

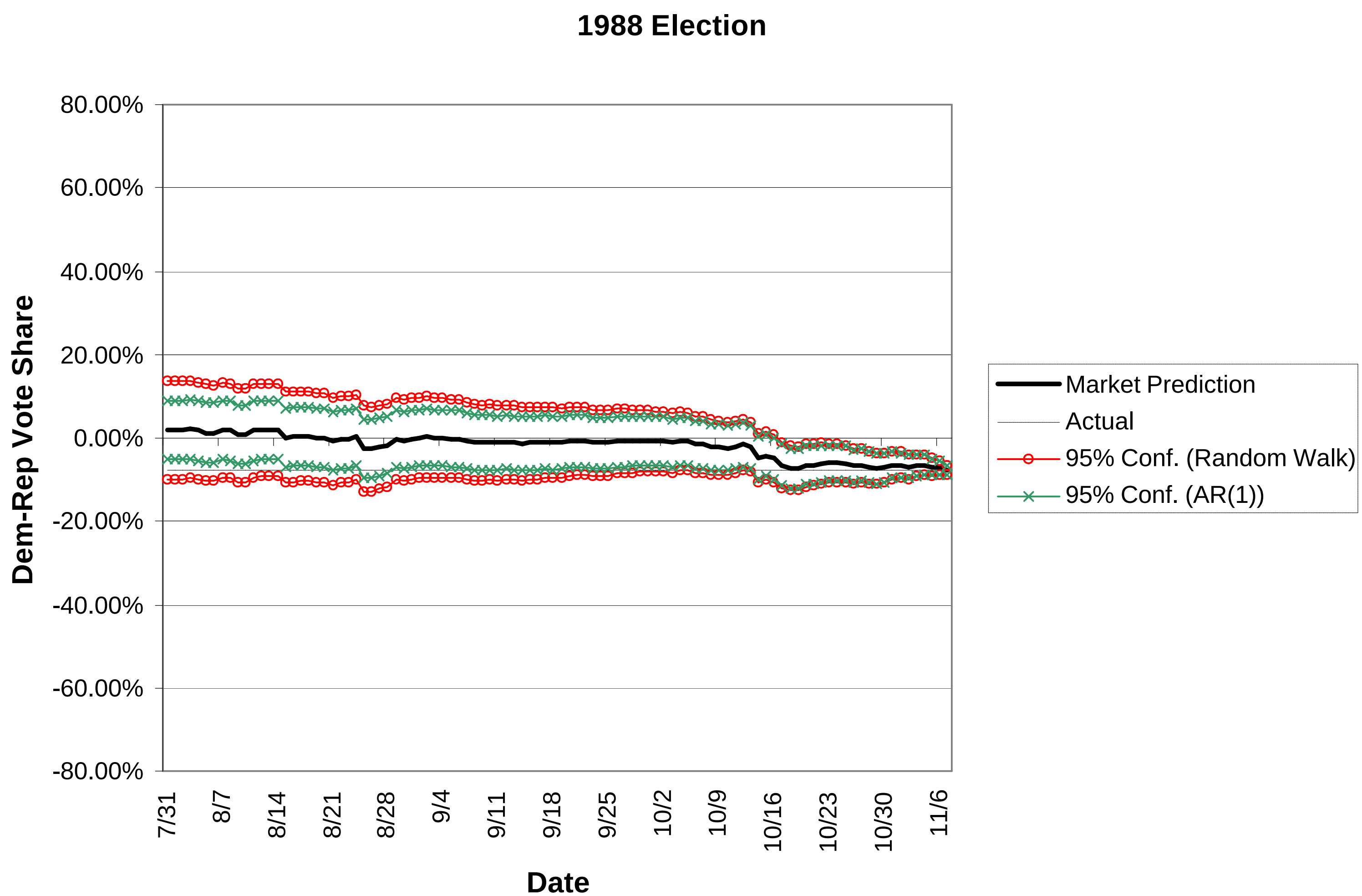

Figure 7: Spread (Democratic – Republican vote share) predictions using bid/ask midpoints for the 1988 Presidential election market along with 95% confidence intervals computed from (1) an estimated AR(1) process and (2) a random walk.

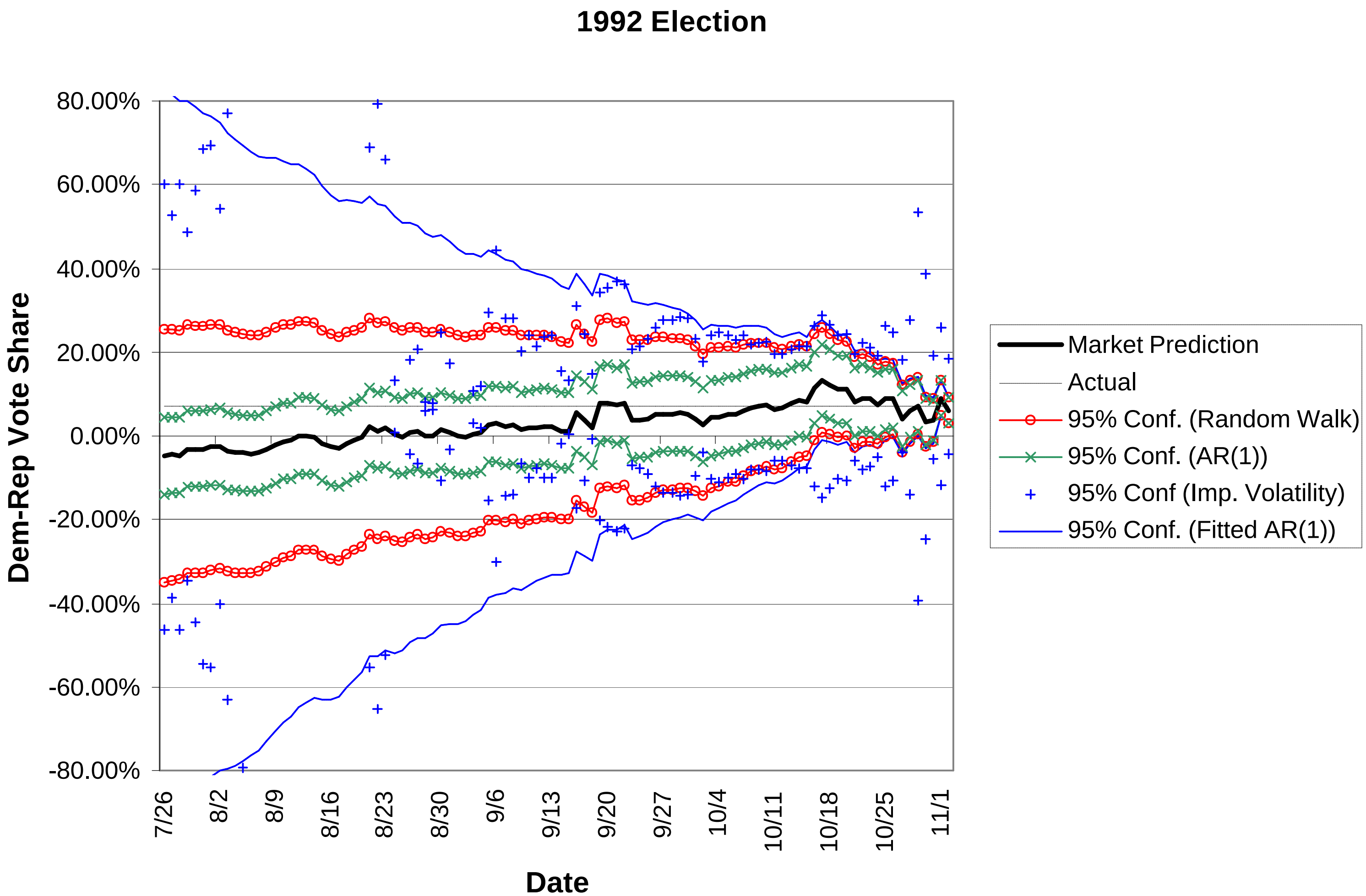

Figure 8: Spread (Democratic – Republican vote share) predictions using bid/ask midpoints for the 1992 Presidential election market along with 95% confidence intervals computed from (1) an estimated AR(1) process, (2) a random walk., (3) implied volatilities computed from simultaneous midpoints in the winner-takes-all market and (4) a fitted AR(1) process that minimizes the sum of squared differences between the implied volatility standard errors and the fitted AR(1) standard errors.

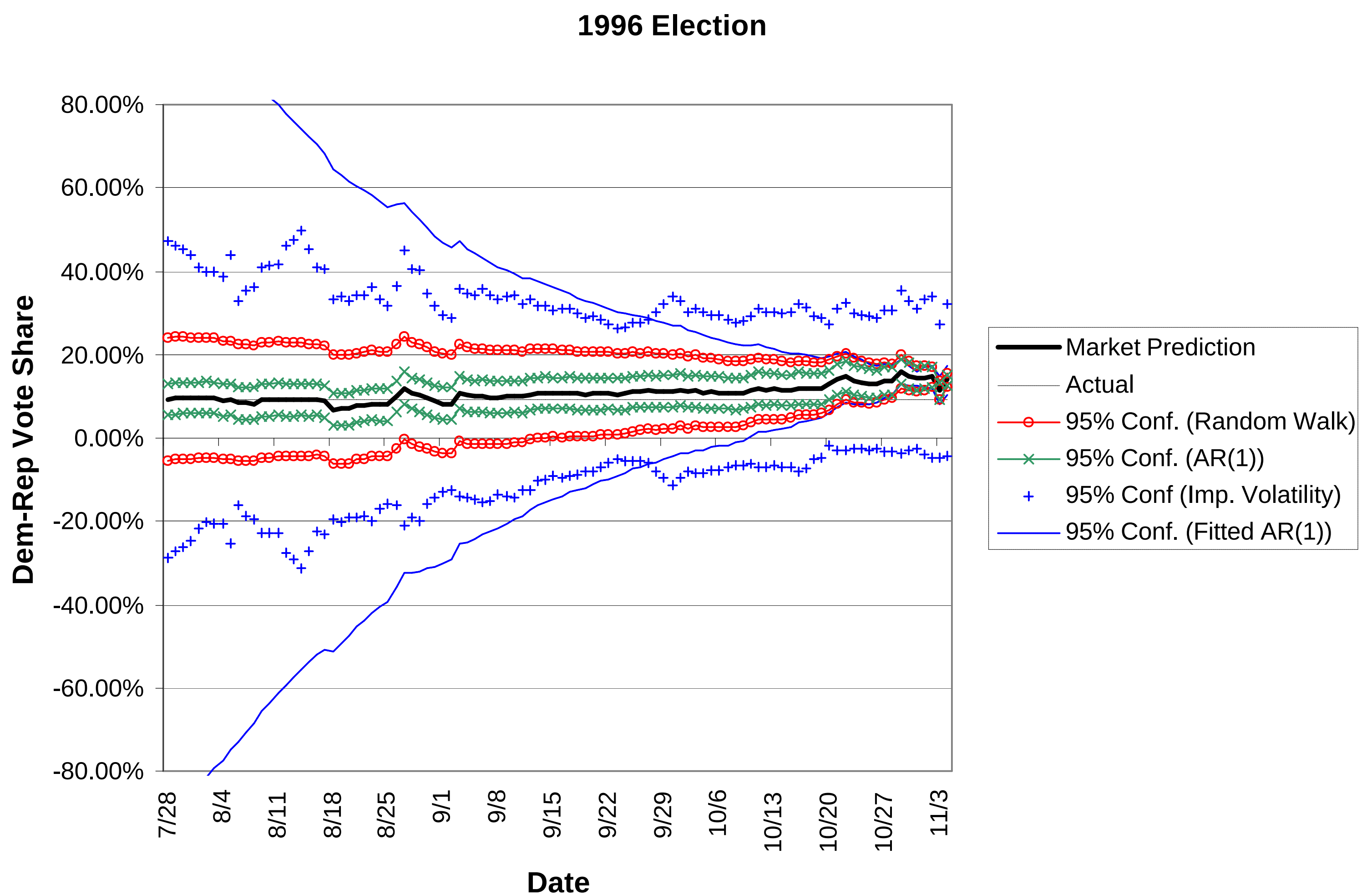

Figure 9: Spread (Democratic – Republican vote share) predictions using bid/ask midpoints for the 1996 Presidential election market along with 95% confidence intervals computed from (1) an estimated AR(1) process, (2) a random walk., (3) implied volatilities computed from simultaneous midpoints in the winner-takes-all market and (4) a fitted AR(1) process that minimizes the sum of squared differences between the implied volatility standard errors and the fitted AR(1) standard errors.

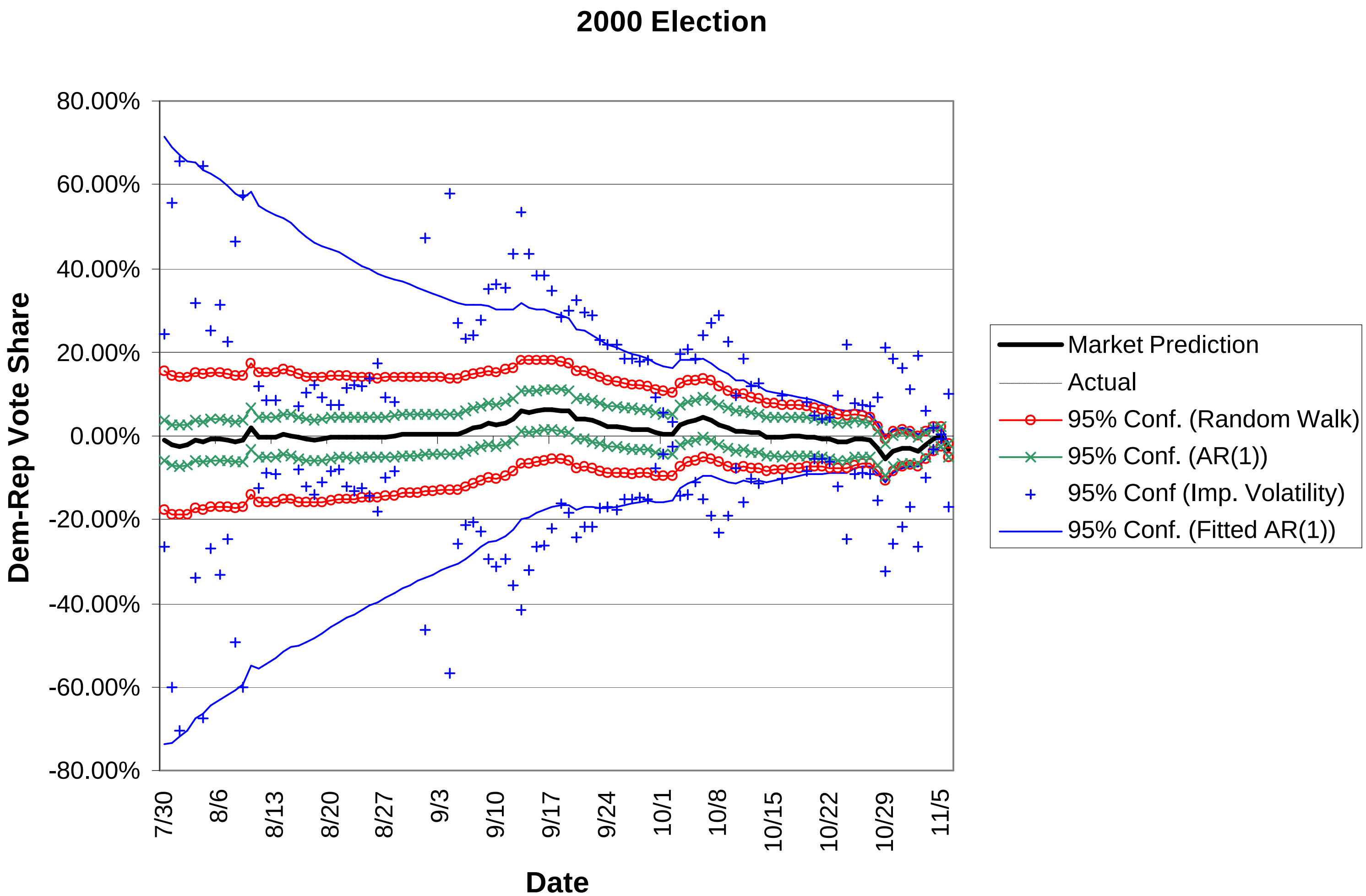

Figure 10: Spread (Democratic – Republican vote share) predictions using bid/ask midpoints for the 2000 Presidential election market along with 95% confidence intervals computed from (1) an estimated AR(1) process, (2) a random walk., (3) implied volatilities computed from simultaneous midpoints in the winner-takes-all market and (4) a fitted AR(1) process that minimizes the sum of squared differences between the implied volatility standard errors and the fitted AR(1) standard errors.