The Pollyvote: Applying the Combination Principle in Forecasting to the 2004 Presidential Election

Alfred G. Cuzán, The University of West Florida

J. Scott Armstrong, The Wharton School, University of Pennsylvania

Randall Jones, The University of Central Oklahoma

Abstract

The outcome of the 2004 presidential election was forecast by applying the combination principle, a procedure which in other contexts has been shown to reduce forecast error. This forecasting technique involved averaging within and across four categories of methods (polls, Iowa Electronic Market quotes, quantitative models, and a Delphi survey of experts on American politics) to compute a combined forecast of the incumbent's share of the two-party vote. We called the resulting average the Pollyvote, because it was derived from many methods and applied to a political phenomenon. When tested across the 163 days preceding the election, the mean absolute error of the Pollyvote predictions was only three-fourths as large as the error of the next most accurate method. Gains in error reduction were achieved for all forecast horizons.

Presented at the 25th International Symposium on Forecasting

San Antonio, June 14, 2005

The Pollyvote: Applying the Combination Principle in Forecasting to the 2004 Presidential Election

Alfred G. Cuzán, The University of West Florida

J. Scott Armstrong, The Wharton School, University of Pennsylvania

Randall Jones, The University of Central Oklahoma

In March 2004 we set out to apply the combination principle in forecasting in order to predict President Bush's share of the two-party popular vote (omitting minor candidates) To that end, 268 polls, 10 quantitative models, and 246 days of Bush|Kerry futures contracts quotes in the Iowa Electronic Markets were collected. Also, using the Delphi technique, three surveys in as many months were administered to a panel of 17 American politics experts, asking them for their predictions.[1] In applying the combination principle, we first combined predictions *within* the first three prediction methods, averaging recent polls, averaging the average daily quotes of Bush|Kerry futures contracts for the previous week, and averaging results of the quantitative models. We then averaged the forecast vote *across* all four methods—the combined (averaged) forecasts of the polls, the IEM quotes, and the models, plus the predictions of the experts' panel, assigning equal weights to each. We call the resulting forecast the Pollyvote – "pol" for political and "poly" for many methods. From March to November 2004 the Pollyvote was updated at first once a week and then, as the campaign progressed and more polls were published, twice a week.

The analysis proceeds as follows. First, we discuss the combination principle in forecasting. This is followed by a discussion of the four components that comprise the Pollyvote (polls, quotes in the Iowa Electronic Markets, quantitative models, and predictions of experts),

[1] All data are available on Polly's data page at politicalforecasting.com.

and the proportional contribution of each to the Pollyvote. Next we compare the performance of the Pollyvote relative to its components. We conclude with observations on the value of the combination principle in forecasting elections and, in particular, its application to the 2004 race for the White House.

The Combination Principle in Forecasting

Combinations of forecasts can reduce error in several ways. On one hand, a combined forecast would likely be more accurate than its typical component because biases associated with the data and methods used in various forecasts are likely to differ, particularly if predictions from alternative forecasting methods are being combined. Different methods are likely to have different biases, meaning that their forecast errors would probably be uncorrelated and perhaps offsetting. In addition, combined forecasts are necessarily derived from more information than any one component forecast. More information provides a more complete picture of influences affecting the future. In probability terms, because the "sample" of information underlying a combined forecast is larger than that of a single forecast, it is probable that information used in generating the combined forecast is more accurate than that coming from any single source.

These expectations are supported by empirical validation studies. A meta-analysis of 30 studies showed that the strategy of combining forecasts reduced forecast errors by about 12% when compared with the typical error of the components. Often, though not always, the combined forecast was more accurate than the best individual method. Many of these studies were based on combining only two methods, and most of the combinations were derived from similar methods (such as judgmental forecasts). With every additional method accuracy normally improves, although at a lower rate. Armstrong (2001) recommends using as many as five methods.

Under ideal conditions the gains from combining are expected to substantially exceed the 12% error reduction reported in the meta-analysis mentioned previously. In addition, gains are expected to be higher with increased forecast uncertainty. Thus, combining is especially useful for long forecast horizons. Prior studies also reveal that combining forecasts never harms forecast accuracy and substantially reduces the risk of large forecast errors.

Existing research suggests that to combine forecasts, one should

1. use different methods or different data sets, or both
2. include up to five methods when possible
3. combine forecasts mechanically, according to a predetermined procedure
4. apply equal weights to the various components that comprise the forecast, unless there is strong prior evidence of relative accuracy of the components.

Given that previous research has established that combining produces more accurate forecasts than the typical component, our purpose in this project is to assess the extent of improvement obtained with this method in forecasting the 2004 presidential election. Election forecasting provides an ideal setting in which to apply the combination principle. There is uncertainty as to which prediction method is best; many forecasts are produced by different methods; and the techniques of the various methods are substantially different. Therefore, we were surprised at being unable to find any studies on the use of combining *across* methods in election forecasting, although Bartels and Zaller (2001) combined 48 regression models to obtain an improved *ex post* forecast of the 2000 presidential election.

Combining Within Method

We now describe the four components of the combined Pollyvote: trial-heat polls, the Iowa Electronic Market, quantitative modeling, and Delphi surveys of a panel of American

politics experts. Also, with the first three, where it is appropriate, we demonstrate the value of combining within methods. Be it noted, though, that when we last performed the computations, Bush's share of the two-party vote was estimated at 51.3%. Since then this has been set at 51.24%. Therefore, our calculations for the errors are off by 0.06%.

The polls. Trial heat polls, revealing public support among candidates in an election, are the traditional means of forecasting election results. The technique was pioneered in the late 1930s by George Gallup. His firm conducted in-person interviews across the nation, providing the first assessments of public opinion using accepted probability sampling procedures. Over the intervening years polls taken late in presidential campaigns have been reasonably accurate. Today, many organizations conduct polls, and nearly all do so by telephone interviews with respondents at their residences. The underlying approach, however, remains the same. A probability sample of interviewees is asked which candidate they would vote for if the election were being held at that time. Although the survey results are not predictions – only assessments of current opinion or "snapshots" – consumers of polls routinely project the results to election day.

The early days of polling were marked by a 1936 *Literary Digest* Poll failure, widely cited as a major turning point in the development of polling. The magazine had a perfect record in predicting the winners of U. S. presidential elections since 1920, and its forecast was off by only 1% in 1932. But in 1936 the *Literary Digest* predicted a landslide victory for Landon over Roosevelt, 55% to 41%. The actual result was that Roosevelt garnered 61%, compared to Landon's 37%. Thus, there was a 20% error in predicting Roosevelt's vote. Squire (1988) provided the first empirical study of this case, drawing upon a May 1937 Gallup survey on the causes of the *Literary Digest's* failure. Was the failure due to sampling bias (the most popular

hypothesis, judging from prior discussion of this case), or due to non-response bias (a less popular hypothesis)? Squire's analysis attributes part of the error to sampling. Nevertheless, this problem could not have been sufficient to cause the incorrect prediction. The most significant error, he concluded, was non-response bias. After 1936, polling procedures kept improving over the years, notwithstanding the Truman-Dewey failure in the close election of 1948. Perry (1979: 323) reports that the error incurred in American national elections declined steadily, from 1.7% in the 1950s to 1.5% in the 1960s and 1.0% in the 1970s.

Nevertheless, early research by Campbell and Wink (1990), as extended by Campbell (1996; 2000; 2004a) and Jones (2002), suggests that trial heat polls conducted before September are inadequate as predictors of presidential election outcomes. For elections from 1948 through 2000, the mean absolute error (MAE) for trial heats in June compared to the election result was greater than 7%. For July the MAE exceeded 6%. By early September, around Labor Day, the error had dropped to about 4%, and by mid-October to about 3% (Campbell 2004a: 764). Labor Day seems to be a critical point in the campaign. Campbell reports that since 1948, 11 of the 14 frontrunners in trial heats near Labor Day won the popular vote (2004a: 765). Although historically the forecast error of polls taken during the fall campaign has been low, during the six to eight weeks before the election the candidates' standings in the polls have varied, sometimes substantially. Yet the public usually returns to the candidate preferred around Labor Day. This phenomenon is explored by Gelman and King (1993) in an article with the intriguing title, "Why Are American Presidential Election Campaign Polls So Variable When Votes Are So Predictable?"

In 2004, polls conducted by reputable survey organizations at about the same time revealed considerable variation in results. For example, in mid-September Gallup showed Bush

ahead by 14%, whereas on the previous day Harris reported a 1% Kerry lead. Various explanations for discrepancies such as these have been proposed. Among them are the exclusive use of cellular telephones among many young adults who do not have landline phones; increased use of call screening aimed at bothersome telemarketers, fund raisers, and campaign workers; the difficulties in identifying likely voters; and the practice of counting undecided voters who may "lean" toward a candidate at the time of the interview, but whose weak commitment may lead them to change their preferences later (Asher 2004).

One way to address this problem is to apply the combination principle, averaging poll results or taking the median score. In an example of the latter method, Colley and Gott (2004) correctly predicted the winner in each state, except Hawaii, by taking the median of the most recent 30 days of polls. Our project illustrates the former method. From March to October the Pollyvote was revised weekly or twice weekly as new polls were published. Through August, on average about five to six trial heat polls were published per week, usually one every 2 or 3 days. On the assumption that the more recent polls contained more useful information, from the beginning we decided to average the three most recent polls to construct this component of the Pollyvote. Doing so was a pragmatic choice. Well over a dozen pollsters were represented in the data base, so any one three-poll average was likely to include a different combination of polls. However, on election eve we faced a problem. Fourteen polls were published between Saturday and Monday morning. There was no non-arbitrary way of deciding which three polls were the most recent. So, we averaged all 14 polls.

Table 1 displays this component of the Pollyvote as it was shown on the morning of the election in Polly's Table 1 at politicalforecasting.com. All poll results reported there are based on interviews with likely voters. It is apparent that most polls were close to the election result.

The mean absolute error of the typical poll was just under 1.0%, which is consistent with the errors reported by Perry (1979). Still, the error is 4/5 larger than that of the combined poll shown in the bottom row of Table 1. In other words, by applying the combination principle to the polls on the eve of the election, the forecast error was reduced by 45%, from 0.91% to 0.50%.

. . . . . .

Table 1 about here

. . . . . .

Comparable calculations during the last week of August and the last week of September yield a reduction in MAE of about 10%. Over the August 1- November 1 period, the error reduction is much greater: 1.6% in the typical poll vs. 0.02% in the combined poll.

The Iowa Electronic Markets. Prediction markets, also known as betting markets, information markets, and events futures, are becoming an important method in forecasting. Betting on the outcome of American elections is nothing new. Between the end of the Civil War and World War II, "large and often well-organized markets for betting on presidential elections" correctly picked the winner in every case but 1916; also, "they were highly successful in identifying those elections—1884, 1888, 1892 and 1916—that would be very close" (Rhode and Strumpf, 2004: 127). More recently, in the four elections prior to 2004, the Iowa Electronic Markets (IEM), a teaching, research, and forecasting tool created by the College of Business faculty at the University of Iowa, has performed better than polls in predicting the margin of victory for the presidential election winner. "In the week leading up to the election, these markets have predicted vote shares for the Democratic and Republican candidates with an average absolute error of around 1.5 percentage points. By comparison, over the same four elections, the final Gallup poll yielded forecasts that erred by 2.1 percentage points" (Wolfers and Zitzewitz,

2004: 112; see also Berg et. al., forthcoming. On the performance of election markets in Australia, see Wolfers and Leigh, 2002).

We applied the combination principle to prices of futures contracts of Bush over Kerry on the IEM by taking the average of the daily averages over the week leading to the most recent poll, which was approximately the same time when voters were being interviewed for the poll. We expected that averaging over a week would adjust for variations that occur merely because bettors are influenced by the actions of other bettors even when the other bettors have no new information (Bikhchandani et. al., 1998).

. . . . . .

Table 2 about here

. . . . . .

In Table 2 we compare the daily average values of Bush contracts during the last week of trading, ending on November 1st, with the composite values. In predicting the election result the error of the composite quoted prices computed over the week between October 26th and November 1st is only 2/3 as large as that of the typical quoted price (or "quote"). Thus, in the week leading up to the election, when error would be expected to be small, it was reduced by 35% (from 0.46 to 0.30) when the combination principle was applied to the data. The error reductions were of 0% (i.e., no reduction) in the last week of August and 11% in the last week of September. Over the August 1-November 1 period, the error reduction again was much greater: 0.73 in the typical quote vs. 0.36 in the combined quote.

Quantitative Models. Over the last several election cycles, political scientists and economists have employed regression models of past elections to forecast the percent of the two-

party vote going to the incumbent party candidate in the next election. Most models consist of between two and seven variables and are estimated for elections since World War II.

. . . . . .

Table 3 about here

. . . . . .

Most quantitative models include at least one measure of economic conditions, although no two employ the same exact indicator. Also, most models have at least one public opinion variable, typically trial heat polls or presidential approval ratings. The individual track records of quantitative models is mixed, although Bartels and Zaller (2001: 14, 19) note that the weighted average of results from numerous models performs according to expectations from "political scientists'understanding of presidential elections." Among the best-known models are those by Abramowitz, Campbell, and Fair. Most models have undergone some revision since their first appearance, particularly after a forecast has been wide of the mark.[2]

Up to ten quantitative model forecasts were included the Pollyvote. Most were not available until August and some of them, such as those by Fair and Lewis-Beck and Tien, were revised in response to the latest economic estimates. The forecasts that were posted in Polly's Table at politicalforecasting.com on the eve of the election are displayed in Table 3. Again, note that the error obtained with the composite forecast (2.5%) is smaller than the typical forecast error (2.8%). The reduction in error that resulted from applying the combination principle to the quantitative models was 11%.

[2] For descriptions and comparisons of forecasting models, see Jones (2002) and Cuzán and Bundrick (2005). See, also, the contributions by most of the modelers whose names appear in Table 3 in the October, 2004, issue of *PS: Political Science and Politics* 37: 733-767.

Panel of Experts. The Delphi technique, developed by the Rand Corporation in the 1950s, encourages a consensus on a specified subject among experts, who may be widely scattered geographically. A Delphi survey includes four features: "anonymity, iteration, controlled feedback, and the statistical aggregation of group response" (Rowe and Wright, 1999: 354). Typically, one obtains estimates from five to twenty experts. The experts are contacted at least twice. Having been provided with summary statistics for the group and reasons offered for individual responses (while preserving anonymity), the participants are asked to revise their original estimates. This process may be repeated for several rounds, after each of which the responses are compiled into a group estimate. Because panelists do not meet in person, the possibility of biased responses due to the influence of strong personalities or individual status is eliminated. (Software at forecastingprinciples.com explains the process and may be used to aid in the construction of surveys and their analysis.)

We have found no previous published use of the Delphi technique for election forecasting. Yet, Delphi seems appropriate for predicting elections because experts have knowledge of current information such as polls and are aware of the impact of debates, conventions, and issues. Finally, they might be able to adjust for excluded variables, such as the effect of a candidate's gaff. In implementing the Delphi technique for predicting the 2004 election, we contacted some two dozen American politics experts. Most were from the ranks of academia, though some were at Washington think tanks, or in the media, or were former politicos. We deliberately excluded anyone who does election forecasting, because that method is represented as a separate component in the Pollyvote. In the end, seventeen experts, whose names appear in the Appendix, participated in at least one of three surveys, each consisting of two rounds. Results were obtained

and posted on Polly's Page in August after the Democratic National Convention, in September after the Republican National Convention, and in October after the debates.[3]

In each survey we asked panel members for their estimates of Bush's share of the two-party vote on election day, along with an explanation of their predictions. After the first round, summary statistics for the group, along with reasons offered for the estimates, were distributed to the panelists, without identifying the authors. The experts then were asked to offer another estimate or to reconfirm their original one. Interestingly, the median prediction of Bush's share in the two-party vote did not change much from one survey to the next: 49.5% in the first, and 50.5% in the following two surveys.

Combining Across Methods to Construct the Pollyvote

Combining is ideal where forecast errors from different methods are negatively correlated or uncorrelated with each other. If forecast errors are positively correlated, combining is still useful the more the correlation coefficients fall short of +1.0. As shown in Table 4, the errors are at most only moderately correlated for the four methods incorporated into the Pollyvote: the average of the three most recent polls, the weekly average of the daily IEM quotes in the week leading up to the most recent poll, the average of the quantitative model forecasts, and the median prediction of the panel of experts. Therefore, one would expect that combining forecasts from all of these methods would produce less forecast error than would typical forecasts of the component methods. This phenomenon is explained further in the next section.

. . . . . .

Table 4 about here

. . . . . .

3 We thank Cati Verdi, who provided valuable assistance in gathering and analyzing the

As shown in Table 5, during the August 1 – November 1 period, when all four of its components were in place, the Pollyvote oscillated narrowly around a mean of 51.5. As it happens, this was its final forecast for Bush's share of the two-party vote. Most of the little variation in the Pollyvote is accounted for by the polls, whose coefficient of variation is three times that of the Pollyvote. In fact, the rolling average of the three most recent polls accounted for 78% of the variation in the Pollyvote. By contrast, there was minimal variation in the results from both the experts and the quantitative models. This had a stabilizing effect on the Pollyvote.

. . . . . .

Table 5 about here

. . . . . .

In a sense, then, Polly is not a flighty bird. Those following the election during the year might view Polly as rather dull, preferring the excitement, uncertainty, and anxiety that can be produced by wild swings in the polls. Polly, however, prefers accuracy to excitement.

The Pollyvote vs. Its Components

In Table 6 we compare the Pollyvote and its components on the mean error (ME) and the mean absolute error (MAE) over the August 1-November 1 period, when all four forecasting methods were available. The Pollyvote has the lowest MAE. At 0.47, it reduces the error of the typical component (1.36) by nearly 2/3.

. . . . . .

Table 6 about here

. . . . . .

---

information used in this component of the Pollyvote.

Our primary interest is to determine the extent that would-be users of the typical method can benefit from using the Pollyvote. The three major competitors to the Pollyvote, all with their advocates, are polls, prediction markets, and quantitative models. On average across the time period, the typical (average) absolute error (MAE) for these three components was 1.36, as shown in Table 6. By contrast, the Pollyvote error was 0.47. This constitutes an error reduction of 65%, which is more than twice the largest error reduction observed in prior studies on combining—24% (Armstrong, 2001: 428; Makridakis and Wrinkler, 1983). Relative to the IEM, which in Table 6 ranks second after the Pollyvote, the error is reduced by 23%.

The Pollyvote vs. the Typical (Uncombined) Poll and IEM Quote

In this section we evaluate the performance of the Pollyvote relative to the typical uncombined forecast made with polls and Bush/Kerry quotes on the Iowa Electronic Market. These are the two principal rivals of the Pollyvote and the next most accurate. We compare the errors resulting from each of the methods, first over the entire August 1-November 1 period, when forecasts from all four prediction methods were available, and then across three forecast horizons.

Table 7 displays how the Pollyvote, the polls, and the IEM quotes rank on the mean error (ME) and the mean absolute error (MAE) over the entire three-month period. There were 93 daily averages of IEM quotes. Over the same time frame there was a total of 163 polls and an equal number of observations for the Pollyvote, which was recalculated with every new poll.

. . . . . .

Table 7 about here

. . . . . .

As noted previously, the MAE for the Pollyvote was 0.47 for this period. As shown in Table 7, MAE for the polls was 1.56%. Thus, the Pollyvote reduces this error by about 70%.

Compared to the IEM, the Pollyvote reduced error by 36%. On average then, the Pollyvote reduced the forecast error by more than half when compared with polls and the IEM over the three months prior to the election.

Table 8 compares the typical forecast of the three methods (Pollyvote, polls, and IEM) on the mean absolute error across the forecast horizon. Although there is no hard and fast rule for dividing the forecast horizon of an election campaign, we consider that the long-term horizon begins when the opposition candidate is known, which for 2004 was in March. The mid-term horizon ends before the debates, and the short-term horizon consists of the week before the election. We designated the last week of September as the closing of the mid-term horizon because the first debate took place on the evening of September 30th. The last week of October was the end of the short-term horizon, since the election was held on November 2nd. As for the long-term horizon, unfortunately, in March the panel of experts was only in the planning stage and there were only three quantitative model forecasts available. By the end of August, however, the first Delphi survey of the panel of experts had been completed, and 11 forecasts had been reported from the quantitative models then available. Thus, in Table 8 we take the last week of August to represent the long-term horizon. This was some time after the Democratic convention, but before the Republican convention.

. . . . . .

Table 8 about here

. . . . . .

As is evident in Table 8, across the three forecast horizons the MAE is lower for the Pollyvote than for either the IEM or polls. Relative to the IEM, which is the most accurate individual method, the Pollyvote reduces error by about half in the long term, and by 8% and 15%

in mid-term and the short-term, respectively. Pollyvote error reductions are even greater when compared with the polls. Comparing the Pollyvote and IEM further, on six days in August the IEM dipped below 50 percent. By contrast, the Pollyvote never forecast a loss for Bush in the popular vote. This is illustrated in Figure 1, which compares the Pollyvote with its IEM component.

. . . . . .

Figure 1 about here

. . . . . .

Conclusion

In this study we combined forecasts within the three prediction methods (polls, IEM, and quantitative models) and obtained substantially smaller forecast errors than for the typical forecast produced by each method. Combining across methods also reduced error substantially, considerably more than what has been achieved in previous studies on combining. Over the three months prior to the election, on average the Pollyvote reduced forecast error by more than half when compared with polls or the IEM. Finally, on election eve the combined forecast, the Pollyvote, predicted Bush's share of the two-party vote to be 51.5%, coming within 0.26% of the actual result (the very latest estimate, which puts Bush's share at 51.24%).

When we began this project we anticipated that after the election we would need to recalibrate the Pollyvote, adjusting the weights of its components and the averaging formula within them. Recall that we assigned equal weights to each of the four components. Also, into the Pollyvote we entered the weekly average of the daily average of the Bush|Kerry quotes and, except on election eve, the average of the three most recent polls. Finally, we weighted all the

forecasting models equally. We picked these formulas for pragmatic reasons. Yet, in retrospect, at least in this project they yielded results that would be hard to beat.

REFERENCES

Abramowitz, Alan I. 2004. When Good Forecasts Go Bad: The Time-for-Change Model and the 2004 Presidential Election. *PS: Political Science and Politics*, 37 (4): 745-746.

Armstrong, J. Scott. 2001. Combining Forecasts. In J. Scott Armstrong, (ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Kluwer Academic Publishers, 417-439.

Asher, Herbert. 2004. *Polling and the Public: What Every Citizen Should Know*. 6th ed. Washington: CQ Press.

Bartels, Larry M. and John Zaller. 2001. Presidential Vote Models: A Recount. *PS: Political Science and Politics*, 34: 9-20.

Berg, Joyce, Robert Forsythe, Forrest Nelson, and Thomas Rietz. Forthcoming. Results from a Dozen Years of Election Futures Markets Research. In Charles R. Plott and Vernon L. Smith (eds.), *Handbook of Experimental Economics Results*. Amsterdam: North Holland. Available: http://www.biz.uiowa.edu/faculty/trietz/papers/iemresults.pdf

Bikhchandani, Sushil, David Hirshleifer and Ivo Welch. 1998. Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades. *Journal of Economic Perspectives*, 12: 151-170.

Campbell, James E. 2004a. Forecasting the Presidential Vote in 2004: Placing Preference Polls in Context. *PS: Political Science and Politics*, 37 (4): 763-768.

_____. 2004b. Introduction—The 2004 Presidential Elections Forecasts. *PS: Political Science and Politics*, 37 (4): 733-735.

_____. 2000. Polls and Votes: The Trial-Heat Presidential Election Forecasting Model, Certainty, and Political Campaigns. In James E. Campbell and James C. Garand (eds.),

Before the Vote: Forecasting American National Elections. Thousand Oaks, CA: Sage Publications.

_____. 1996. Polls and Votes: The Trial-Heat Presidential Election Forecasting Model, Certainty, and Political Campaigns. *American Politics Quarterly*, 24: 408-433; forecast addendum, 532-533.

Campbell, James E. and Kenneth A. Wink. 1990. Trial Heat Forecasts of the Presidential Vote. *American Politics Quarterly*, 18: 251-269.

Colley, Wesley N. and J. Richard Gott III. 2004. Electoral Scoreboard. Available: http://colleyrankings.com/election2004/release.html

Cuzán, Alfred G. and Charles M. Bundrick. 2005. Deconstructing the 2004 Presidential Election Forecasts: The Fiscal Model and the Campbell Collection Compared. *PS: Political Science and Politics*, 38, 2 (April), 255-262,

Fair, Ray. 2002. The Effect of Economic Events on Votes for President: 2000 Update. November 1. http://fairmodel.econ.yale.edu/RAYFAIR/PDF/2002DHTM.HTM

_____. 1998. The Effect of Economic Events on Votes for President: 1996 Update. November 6. http://fairmodel.econ.yale.edu/RAYFAIR/PDF/1998A.HTM

_____. 1996. The Effect of Economic Events on Votes for President: 1992 Update. *Political Behavior,* 18: 119-139.

Gelman, Andrew and Gary King. 1993. Why Are American Presidential Election Campaign Polls So Variable When Votes Are So Predictable? *British Journal of Political Science*, 23: 409-451.

Jones, Randall J., Jr. 2002. *Who Will Be in the White House? Predicting Presidential Elections*. New York: Longman Publishers.

Lemert, James B. 1986. Picking the Winners: Politician vs. Voter Predictions of Two Controversial Ballot Measures. *Public Opinion Quarterly,* 50: 208-221.

Lichtman, Allan J. 2003. The Keys to Election 2004. *The Gazette*. April 25, 2003. Available: http://www.gazette.net/200317/weekend/issues_ideas/155679-1.html

_____ . 1996. *The Keys to the White House, 1996: A Surefire Guide to Predicting the Next President,* Lanham, MD: Madison Books.

_____ and Ken DeCell. 1990. *The Thirteen Keys to the Presidency*. Lanham, MD: Madison Books.

_____ and Volodia I. Keilis-Borok. 1981. Pattern Recognition Applied to Presidential Elections in the United States, 1860-1980: Role of Integral Social, Economic, and Political Traits. *Proceedings of the National Academy of Sciences*, 78 (11), 7230-7234.

Makridakis, Spyros and Robert L. Winkler. 1983. Averages of Forecasts: Some Empirical Results. *Management Science*, 29: 987-996.

Perry, Paul. 1979. Certain Problems with Election Survey Methodology. *Public Opinion Quarterly*, 43: 312-325.

Rhode, Paul W. and Koleman S. Strumpf. 2004. Historical Presidential Betting Markets. *Journal of Economic Perspectives*, 18 (2): 127-142.

Rosenstone, Steven J. 1983. Forecasting Presidential Elections. New Haven, CT: Yale University Press.

Rowe, Gene and George Wright. 1999. The Delphi Technique as a Forecasting Tool: Issues and Analysis. *International Journal of Forecasting*, 15: 353-375.

Squire, Peverill. 1988. Why the 1936 *Literary Digest* Poll Failed. *Public Opinion Quarterly,* 52: 125-133.

Wolfers, Justin and Andrew Leigh. 2002. Three Tools for Forecasting Federal Elections: Lessons from 2001. *Australian Journal of Political Science,* 37 (2): 223-240.

Wolfers, Justin and Eric Zitzewitz. 2004. Prediction Markets. *Journal of Economic Perspectives*, 18 (2): 107-126.

APPENDIX

Participants in the Panel of Experts Delphi Survey Project

1. Randall Adkins – Associate Professor of Political Science, University of Nebraska at Omaha
2. Michael Barone - Senior Writer, U. S. News and World Report
3. Karlyn Bowman - Resident Fellow, American Enterprise Institute
4. George Edwards – Distinguished Professor of Political Science and Jordan Chair in Presidential Studies, Bush School of Government and Public Service, Texas A & M University.
5. Ada Finifter - Professor of Political Science, Michigan State University
6. Chris Garcia – Professor of Political Science and former President of the University, University of New Mexico
7. Karen Hult – Professor of Political Science, Virginia Polytechnic Institute and State University (Virginia Tech)
8. Gary Jacobson – Professor of Political Science, University of California, San Diego
9. Charles O. Jones – Nonresident Senior Fellow, The Brookings Institution and Hawkins Professor of Political Science Emeritus, University of Wisconsin
10. Kenneth Mayer – Professor of Political Science, University of Wisconsin
11. Leon Panetta – Director of the Panetta Institute of Public Policy, California State University, Monterey Bay
12. Thomas Patterson – Bradlee Professor of Government and the Press, Kennedy School of Government, Harvard University
13. Larry Sabato – Gooch Professor of Politics and Director of the Center for Politics, University of Virginia
14. Harold Stanley – Geurin-Pettus Distinguished Chair in American Politics and Political Economy, Southern Methodist University
15. Charles Walcott – Professor of Political Science, Virginia Polytechnic Institute and State University (Virginia Tech)
16. Martin Wattenberg – Professor of Political Science, University of California, Irvine
17. Herbert Weisberg – Professor of Political Science, Ohio State University

Table 1. Polls of Likely Voters on the Eve of the 2004 Election
(Bush two-party vote = 51.3)

Date of poll	Source of poll	Bush's share of two-party vote in the poll	Absolute error
10/31/2004	George Washington U. Battleground Poll	52	0.7
10/31/2004	Fox News	50	1.3
10/31/2004	*Washington Post* three-day tracking poll	50	1.3
10/31/2003	Pew Research Center (	52	0.7
11/1/2004	CBS News/ *New York Times*	52	0.7
11/1/2004	NBC News/*Wall Street Journal*	51	0.3
11/1/2004	Gallup/CNN/USA Today	50	1.3
11/1/2004	Marist College	49	2.3
11/1/2004	TIPP four-day tracking poll	51	0.3
11/1/2004	Zogby three-day tracking poll	51	0.3
11/1/2004	Fox News	49	2.3
11/1/2004	Harris Poll	52	0.7
11/1/2004	Rasmussen Reports three-day tracking poll	51	0.3
11/2/2004	CBS News (polled election eve)	51	0.3
	Mean		0.91
	Combined forecast	50.8	0.50

Table 2. IEM Quotes: Combining Across Time vs. the Typical Daily Average, 2004
(Bush two-party vote = 51.3)

Date	Quote	Absolute Error
10/26	51.0	0.3
10/27	50.7	0.6
10/28	50.7	0.6
10/29	50.8	0.5
10/30	51.6	0.3
10/31	51.7	0.4
11/01	50.8	0.5
Mean		0.46
Combined	51.0	0.30

Table 3. Combining Quantitative Model Forecasts of the 2004 Election
(Bush two-party vote = 51.3)

Issued or posted in Polly's Table on or about	Author	Forecast	Absolute Error
October 29 (update)	Fair	57.7	6.4
September 2	Lockerbie	57.6	6.3
January 29	Norpoth*	54.7	3.4
September 2	Holbrook	54.5	3.2
September 2	Campbell	53.8	2.5
August 12	Abramowitz	53.7	2.4
July 26	Hibbs	53.0	1.7
September 2	Lewis-Beck and Tien	49.9	1.4
August 26	Wlezien & Erickson	51.7	0.4
October 29 (update)	Cuzán and Bundrick	51.2	0.1
	Mean Absolute Error		2.8
	Composite	53.8	2.5

* Norpoth's forecast was calculated on the date shown but not posted in Polly's Table until April.

Table 4. Correlations of Errors: Polls, Experts, IEM and Models
August 1 – November 1, 2004 (N= 163)

	Experts	IEM	Models
Polls	0.03	0.23	0.59
Experts	-	0.26	0.56
IEM	-	-	0.37

Table 5. Descriptive Statistics, the Pollyvote and Its Components
August 1 – November 1, 2004 (N = 163)

	Pollyvote	Polls	IEM	Models	Experts
Mean	51.5	51.2	50.9	53.6	50.0
s.d.	0.51	1.56	0.63	0.16	0.50
Min	50.1	47.3	49.7	53.3	49.5
Max	52.6	55.3	52.5	53.8	50.5

Notation: s.d.= standard deviation.

Table 6. Mean Error (ME) and Mean Absolute Error (MAE)
for Pollyvote and Its Components
August 1 – November 1
Error = 51.3 – forecast
(N=163)

	Pollyvote	Polls	IEM	Experts	Models	Average(1)
ME	-0.17	0.08	0.32	1.27	-2.34	
MAE	0.47	1.2	0.61	1.27	2.34	1.36

(1) Excluding the Pollyvote.

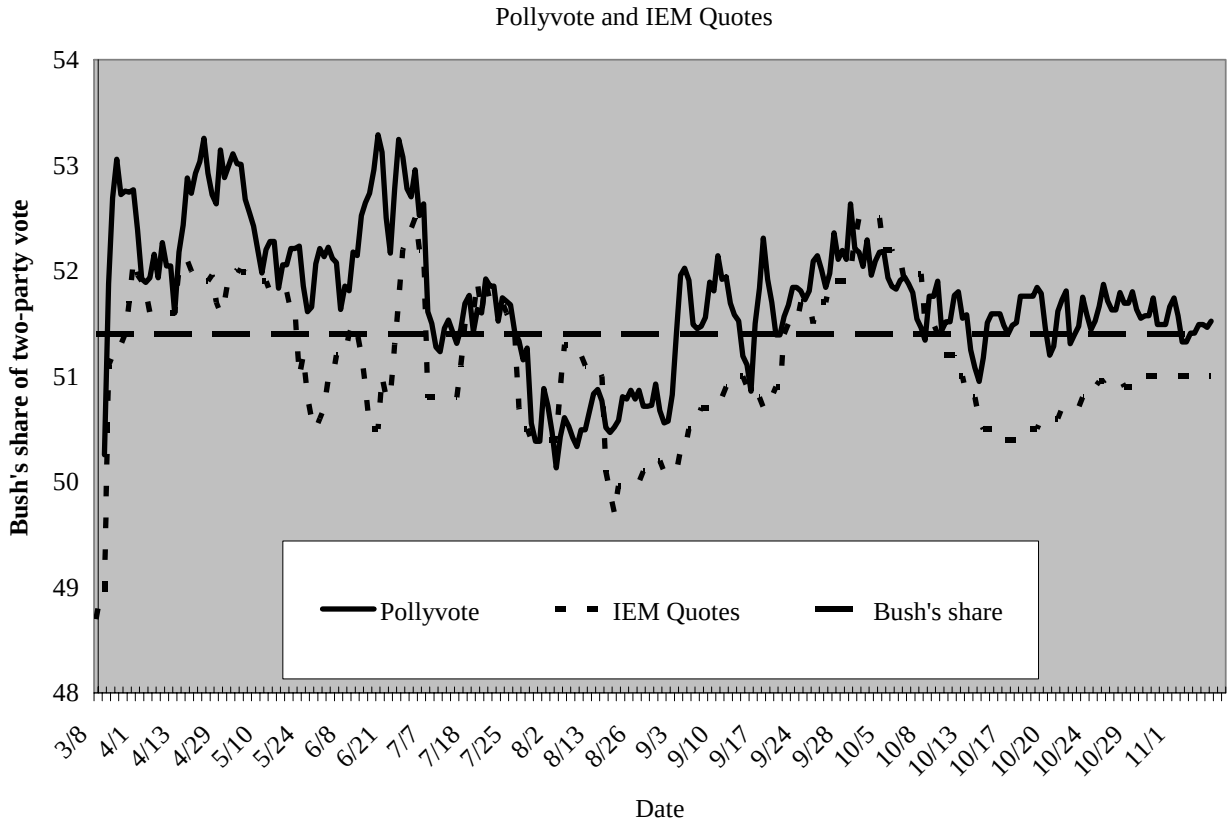
Table 7. The Pollyvote vs. Polls and IEM on the Mean Error (ME) and Mean Absolute Error (MAE) for August 1 – November 1
Error = 51.3 – forecast

	Pollyvote	Polls	IEM	Average(1)
ME	-0.17	0.06	0.36	
MAE	0.47	1.56	0.73	1.15
N	163	163	93	

(1) Excluding the Pollyvote.

Table 8. The Pollyvote, the IEM, and the Polls Compared on the Mean Absolute Error (MAE) Across the Forecast Horizons

Forecast horizon	Pollyvote	IEM	Polls
Long-term	0.53	1.09	1.25
Mid-term	0.76	0.83	2.21
Short-term	0.35	0.41	1.14


Pollyvote and IEM Quotes
Bush's share of two-party vote
54
53
52
51
50
49
48
Pollyvote
IEM Quotes
Bush's share
3/8
4/1
4/13
4/29
5/10
5/24
6/8
6/21
7/7
7/18
7/25
8/2
8/13
8/26
9/3
9/10
9/17
9/24
9/28
10/5
10/8
10/13
10/17
10/20
10/24
10/29
11/1
Date