COMBINING METHODS TO FORECAST THE 2004 PRESIDENTIAL ELECTION THE POLLYVOTE

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

We present an evaluation of a project to forecast the 2004 presidential election by applying the combination principle, a procedure which in other contexts has been shown to reduce error. This involved averaging within and across four categories of methods (polls, Iowa Electronic Markets 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 it the Pollyvote, signifying "many (methods)." Both approaches reduced error. With the Pollyvote, the mean absolute error was reduced by one third relative to the next most accurate method, the Iowa Electronic Markets, when tested across the 163 days preceding the election. Gains were achieved at all forecast horizons that we tested. On the morning of November 2, the Pollyvote had Bush winning 51.5 percent of the two-party vote, which came within 0.2 percent of the outcome (51.3%).

Paper prepared for presentation at the Southern Political Science Association,
New Orleans, January 6, 2005.

COMBINING METHODS TO FORECAST THE 2004 PRESIDENTIAL ELECTION
THE POLLYVOTE

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 a Political Forecasting Special Interest Group (SIG) was launched at the Forecasting Principles site of the Wharton School of the University of Pennsylvania. This having been a presidential election year, we set out to forecast the outcome of the contest. Specifically, the goal was 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 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. Forecasts from these four sources were then averaged (with equal weights), first once a week and later, as the campaign progressed, twice a week. We call these aggregate forecasts the Pollyvote – "pol" for political and "poly" for many methods. To lend a little levity to our project, with the same play on words in mind we adopted Polly the parrot as our mascot.

In producing the Pollyvote forecasts, we combined predictions *within* the component methods, averaging recent polls, averaging values of Bush futures contracts for the previous week, and averaging results of the quantitative models. This approach to election forecasting builds on research that has demonstrated in other contexts that combining forecasts reduces forecast error. That is, it improves forecast accuracy when compared with forecasts derived from one method alone. The reasons why combined forecasts produce greater accuracy of predictions are explained in Armstrong (2001), which cumulates findings from prior research on combining forecasts.

First, we discuss the combination principle in forecasting. This is followed by a discussion of the four components that went into the Pollyvote (the polls, quotes in the Iowa Electronic Markets, quantitative models, and the panel of experts predictions), and the proportional contribution of each to it. Next we compare the performance of the Pollyvote relative to its components, showing that the Pollyvote incurs a smaller error than any of them across the forecast horizon. In light of our experience, we conclude with some observations on the value of the combination principle in forecasting elections and, in particular, its application the 2004 race for the White House.

The Combination Principle in Forecasting

This project was designed to test the value of combining forecasts as a means of reducing error in election predictions. 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 the 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 is one of the most powerful techniques in a forecaster's toolkit (Armstrong 2001). On average, combining reduced forecast errors by about 12%. Often, though not always, the combined forecast was more accurate than the best method. Many of these studies were based on combining only two methods, and most of the combinations were based on similar methods (e.g., only 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. In addition, gains are expected to be higher with increased forecast uncertainty. Thus, combining is especially useful when the forecast horizon is longer. Prior studies also reveal that combining forecasts never harms forecast accuracy. In addition, combining substantially reduces the risk of large forecast errors.

To combine forecasts, one should

1. use different methods or data or both
2. include up to five methods when possible
3. combine forecasts mechanically, according to a pre-set procedure
4. apply equal weights to the various components going into the forecast, unless there is strong evidence of prior accuracy.

It having been established in previous research 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 a better "forecast" (after the fact) of the 2000 presidential election.

Combining Within Method

We 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, for the first three we demonstrate the value of combining within methods.

The polls. Trial heat polls, revealing public support among candidates for the election, are the traditional means of forecasting elections. 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 widely-publicized 1936 *Literary Digest* Poll failure, widely cited as a major turning point. The magazine had a perfect record in predicting the winners of 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 Roosevelt with 61% and Landon with 37%. Thus, there was a 20% error in predicting Roosevelt's vote. Squire (1988) provides the first empirical study of the failure. He did this by 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 upon 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 1948. Perry 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 (1979: 323).

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) between trial heats in June and the election result was greater than 7%. For July it 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, in the six weeks to two months prior to 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 showed 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 lack of generally accepted techniques for 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 later (Asher 2004; Jones 2004b).

One way to address this problem is to apply the combination principle, averaging poll results or taking the median score. (A notable success in predicting the 2004 presidential election, using the median of the most recent 30 days of polls, is Colley and Gott [(2004]), who correctly predicted the winner in each state, except Hawaii.) In our case, 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 2-3 days. On the assumption that the more recent polls contained more information, from the start we decided to average the three most recent polls to construct this component of the Pollyvote. It 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 them. However, on election eve we faced a problem. Fourteen polls were published between Sunday and Monday morning. There was no non-arbitrary way of deciding which were "the most recent polls." So, we averaged all 14 polls published in that period.

Table 1. Polls of Likely Voters to on the Eve of the 2004 Election
(Actual vote for Bush = 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 1 displays this component of the Pollyvote as it was shown in Polly's Table 1 on the morning of November 2nd. All poll results shown are based on interviews with likely voters. It is apparent that most polls were close to the election result. The mean absolute error (MAE) of the typical poll was just under 1.0% (which is in line with the typical errors reported by Perry). 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 error was reduced by 45%.

We made comparable calculations for earlier periods during the campaign, during the last week of August and September. The former conforms to the long-term horizon and the latter to the mid-term horizon. In both the reduction in the MAE was 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 tool in forecasting. Actually, 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 of the College of Business faculty at the University of Iowa, has done 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 Zeizewitz, 2004: 112; see also Berg et. al., 2003).

We applied the combination principle to the IEM quotes by taking the average of the daily averages over the week leading to the most recent poll, approximately the same when voters were being interviewed. 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. IEM Quotes: Combining Across time vs. the Typical Daily Average, 2004
(Bush actual 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

In Table 2 we compare the daily averages during the last week of trading ending on November 1st with the composite quote. The error of the composite forecast computed over the week between October 26th and November 1st is only 2/3 as large that of the typical quote. Thus, in the week leading up to the election, when error would be expected to be at a minimum, it was

reduced by 35% when the combination principle was applied to the data. The error reductions were of 0% in the last week of August and 11% in the last week of September. Over the August 1-November 1 period, the error reduction is again 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 (for an earlier effort, see Rosenstone, [(1985)]. Most models consist of between two and seven variables and are estimated over anywhere between scarcely over a dozen elections to close to twice as many. (By contrast, historian Alan Lichtman uses 13 "keys" to predict whether the incumbents will be reelected.)

Table 3. Combining the Most Recent Quantitative Model Forecasts to the 2004 Election (Bush actual 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.

A common denominator across most quantitative models is at least one measure of economic conditions, although no two employ the same metrics. Also, most models include at least one public opinion variable, a trial heat poll or a presidential approval rating, although here again there is no unanimity on indicators. The individual track record of these models is mixed, although a weighted average "portfolio" of "credible models" performs according to expectations derived from "political scientists'understanding of presidential elections" (Bartels and Zaller, 2001: 14, 19). Among the best-known models are Abramowitz's, Campbell's, and Fair's (for a

comparison, see Cuzán and Bundrick, 2004b). Most models have undergone some revision since their first appearance, particularly after a forecast has gone wide of the mark.[1]

Up to ten quantitative model forecasts were included the Pollyvote. Most were not available until August and some of them, like Fair's and Lewis-Beck and Tien's, were revised in response to the latest economic estimates. The forecasts that were posted in Polly's Table on the eve of the election are displayed in Table 3, shown on the previous page. 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 attained by applying the combination principle to the quantitative models was 11%.

Panel of Experts. The Delphi technique, developed by the Rand Corporation in the 1950s, involves obtaining 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 between five and twenty experts. They 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 estimate. This process may be repeated for several rounds, after 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 the construction of surveys and their analysis.

Surprisingly, we found no published use of the Delphi technique for election forecasting. Yet, Delphi seems appropriate for election forecasting 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 impact when a candidate makes a gaff. Thus, we decided to conduct a Delphi forecast. Some two dozen American politics experts from the ranks of academia, Washington think tanks, the media, and former politicos were invited to participate in the project. 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.[2]

In each survey we asked panel members for their estimate of what Bush's share of the two-party vote would be on election day, along with an explanation of their prediction. After the

---

[1] For descriptions and comparisons of forecasting models, see Jones (2002, 2004a) and Cuzán and Bundrick (2004a, 2004b). 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* (XXXVII, 4: 733-767).

2 We thank Cati Verdi, who provided valuable assistance in gathering and analyzing the information used in this component of the Pollyvote.

first round, summary statistics for the group, along with reasons offered for the estimates, but without identifying their authors, were distributed to the panelists. They were then asked to offer another estimate or to confirm that they were sticking with their original one. Surprisingly, 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 from 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) are at most moderately correlated. Therefore, one would expect their combination to yield a reduction in error relative to the typical forecast of its component methods (see next section).

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

As shown in Table 5, during the August 1 – November 1 period, when all four of its components were in place, the value of 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 the experts or the quantitative models. This had a stabilizing effect on the Pollyvote.

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
Median	51.6	51.3	50.9	53.7	50.5
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.

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. Taking Polly for a ride will not be a roller coaster experience. Changing the metaphor, it will be a smooth flight.

The Pollyvote vs. Its Components

Tables 6 and 7 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 methods were available. The Pollyvote ranks second and first, respectively. That is, it incurs the second smallest mean error, after the polls, and the smallest mean absolute error. The polls rank first on the ME and third on the MAE, barely edging out the experts. The IEM ranks third on the ME and second on the MAE. In both tables the experts take fourth place and the models bring up the rear. The reduction in the MAE accomplished by using the Pollyvote against the second-place IEM is over 20%.

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

	Pollyvote	Polls	IEM	Experts	Models
Mean	-0.17	0.08	0.32	1.27	-2.34
S.D.	0.51	1.6	0.63	0.5	0.16
Range	2.5	8	2.8	1.0	0.5
Min.	-1.33	-4.03	-1.2	0.8	-2.5
Max.	1.17	3.97	1.6	1.8	-2.0

Table 7. Pollyvote and Its Components Compared on Mean Absolute Error (MAE)
August 1 – November 1
Error = 51.3 – (forecast)
(N=163)

	Pollyvote	Polls	IEM	Experts	Models
Mean	0.47	1.20	0.61	1.27	2.34
S.D.	0.27	0.99	0.36	0.50	0.16
Max.	1.33	4.03	1.60	1.80	2.50

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

In this section we evaluate the performance of the Pollyvote relative to the typical, i.e., uncombined forecast of the incumbent share of the two-party vote in the 2004 presidential election made with the polls and the IEM quotes. These are the two principal "rivals" of the Pollyvote and the next most accurate. We compare the errors incurred with each of the methods first over the entire August 1-November 1 period and then across three forecast horizons.

Table 8 displays how the Pollyvote, the polls, and the IEM quotes rank on the mean error (ME) and the mean absolute error (MAE) over the August 1 – November 1 period, when all methods were available. There are 93 observations for the IEM quotes, a daily average for as many days in the three-month period. Over the same time frame, though, there were a total of 163 polls and, hence, 163 observations for that variable, as well as for the Pollyvote, which was recalculated with every new poll.

On the MAE the Pollyvote again ranks first. In fact, using the Pollyvote instead of the second-ranked IEM slashes the MAE by more than 1/3. Moreover, it does so with the smallest variation (its standard deviation is 1/2 that of the IEM and 1/4 that of the polls). As in the previous section, the polls rank first on the ME but last on the MAE. Compared to the Pollyvote and the IEM, the polls incur large errors, but these are distributed almost evenly above and below the mean.

Table 8. The Pollyvote vs. the Polls and the IEM on the
Mean Error (ME) and Mean Absolute Error (MAE)
August 1 – November 1
Error = 51.3 – forecast

	Pollyvote	Polls	IEM
ME	-0.17	0.06	0.36
MAE	0.47	1.56	0.73
N	163	163	93

Table 9 compares the typical forecast of the three methods (the Pollyvote, the polls, and the 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 would open as soon as the opposition candidate was known, which in 2004 happened to be in March, the mid-term horizon to end before the debates, and the short-term horizon to consist in the week before the election. We designated the last week of September as the closing of the mid-term horizon (the first debate took place on the evening of September 30th) and the last week of October as that of the short-term horizon (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 contrast, by the end of August the first Delphi survey of the panel of experts had been completed and eleven forecasts obtained with quantitative models were available. Thus, in Table 9 we take the last week of August to represent the long-term horizon. This was some time after the Democratic convention but before that of the Republicans.

Table 9. The Pollyvote, the IEM, and the Polls Compared on the
Mean Absolute Error (MAE) Across the Forecast Horizon

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

As is evident in Table 9, the MAE is lower with the Pollyvote across the forecast horizon. Relative to the IEM, its closest competitor, the Pollyvote reduces error by about half in the long term, and by 8% and 15% in mid- and the short-term, respectively. The reductions are even greater with the polls.

Taking the comparison between the Pollyvote and the IEM further, on six days in August the IEM dipped below 50 percent. By contrast, the Pollyvote never forecast a loss in the popular vote for Bush. This is illustrated in Figure 1, which compares the Pollyvote against its IEM component.

. . . . . .
Figure 1 about here
. . . . . .

Conclusion

In this paper we have presented the results of combining forecasts from four different methods for the purpose of forecasting the outcome of the 2004 presidential election. Combining forecasts within the three methods appropriate for this purpose (the polls, the IEM, and the quantitative models) resulted in smaller forecast errors than for the typical forecast produced by each method. Also, combining across methods reduced error substantially, considerably more than what has been achieved in other fields. Finally, on election eve the combined forecast, the Pollyvote, put Bush's share of the two-party vote at 51.5%, coming within 0.2% of the actual result.

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. 2004a. Deconstructing the 2004 Presidential Election Forecasts: The Fiscal Model and the Campbell Collection Compared. Presented at the Northeastern Political Science Association, Boston, November 13, 2004. Available: http://www.uwf.edu/govt/facultyforums/documents/2004NEPSAdeconstructingtheforecasts_001.pdf

_____. 2004b. Fiscal Effects on Presidential Elections: A Forecast for 2004. Paper presented at the September meeting of the American Political Science Association. Available: http://www.uwf.edu/govt/facultyforums/documents/fiscaleffectsprselect2004.pdf.

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

_____. 1998. The Effect of Economic Events on Votes for President: 1996 Update. November 6. Available: 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. 2004a. You Can Predict the 2004 Presidential Election (Maybe!). Available: http://morris.wharton.upenn.edu/forecast/Political/Jones_comments.html#You_Can_Predict

_____. 2004b. Variation in Poll Results. Available: http://morris.wharton.upenn.edu/forecast/Political/Jones_comments.html#Variation in Poll Results

_____. 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 Voldia 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.

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 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

Figure 1. Pollyvote and IEM Quotes, March 8 - November 1, 2004