EXPERT JUDGMENT IN FORECASTING PRESIDENTIAL ELECTIONS:

A PRELIMINARY EVALUATION

Alfred G. Cuzán and Randall J. Jones, Jr.

About the authors

Alfred G. Cuzán (acuzan@uwf.edu) is Professor of Political Science at The University of West Florida, Pensacola, Florida. With Richard J. Heggen and Charles M. Bundrick, he developed the fiscal model of presidential elections, and with J. Scott Armstrong and Randall J. Jones, Jr., he designed the Pollyvote.

Randall J. Jones, Jr. (ranjones@uco.edu) is Professor of Political Science at The University of Central Oklahoma, Edmond, Oklahoma. He was a collaborator in the Pollyvote project during the 2004 and 2008 presidential election campaigns, working primarily with expert forecasting surveys.

EXPERT JUDGMENT IN FORECASTING PRESIDENTIAL ELECTIONS:

A PRELIMINARY EVALUATION

Alfred G. Cuzán and Randall J. Jones, Jr.

This is a preliminary report on expert judgment in forecasting American presidential elections. In 2004 and 2008, the Pollyvote team[1] invited some leading scholars in American politics, none the author of a formal forecasting model, to estimate the share of the two-party vote that the incumbent party candidate (George W. Bush and John McCain, respectively) would take in the November general election. In this paper we examine the accuracy of the experts' forecasts over time and compare them to the forecast accuracy of campaign polls, the Iowa Electronic Market, and one econometric model, Ray Fair's "presidential vote equation" (Fair 2008). We choose Fair's model, in particular, because of its very long lead time. Fair posts his first forecast almost two years ahead of presidential elections and revises his estimates quarterly through the third quarter of the election year, with his final forecast coming less than a week before Election Day.

The 2004 expert panel included 17 members, while 15 experts comprised the 2008 panel. In 2004, the Pollyvote team administered three Delphi surveys, each concluded near the end of the month in July, September, and October. Each survey consisted of two rounds. Panelists were asked to estimate the share of the two-party vote that the incumbent party candidate would receive, rate their confidence in their forecast, and provide reasons for their predictions. After the first round, this information was collected and distributed to all panelists, and in a way that preserved anonymity. The panelists then were asked to offer another estimate or to confirm their original one. There was hardly any change in estimates between one round and the next.

In 2008, the panel was assembled earlier. The first survey was completed in August 2007, more than a year before the election. Six more surveys were conducted, one each in January, March, June, September, October, and November. This time, however, the Delphi approach was dropped after the first survey because the experts continued the pattern begun in 2004 of rarely changing their initial estimates made in the first round. Thus in 2008 all surveys but the first consisted of only one round, with no feedback.

To evaluate the performance of the experts against the other forecasting methods, we proceed as follows. We average trial heat polls[2] and the Iowa Electronic Market

---

[1] The original Pollyvote team consisted of J. Scott Armstrong, Alfred G. Cuzán, and Randall J. Jones, Jr. In 2008, Andreas Graefe became the fourth member of the group.

[2] A "trial heat poll" asks respondents some variation of this question: "If the election were held today, whom would you vote for," followed by the candidates' names.

(IEM) contract prices for candidate vote shares over the three weeks leading up to the day that each survey was completed. We then compare the mean of the experts' predictions against the polls, the IEM, and the value of Fair's forecast posted on the date closest to the time that the survey was completed. The results for 2008 and 2004 are reported visually in Figures 1 and 2, respectively.

Expert Surveys for 2008 Election

We first analyze expert surveys for 2008 because more surveys were conducted and across year longer time horizon than for 2004, so we have more information to report. The main conclusions that we glean from Figure 1 for the 2008 surveys are the following.

Comparing Experts' Accuracy Across Time. Across the 15-plus months covered by the seven surveys, our panel of experts was quite accurate in its estimates. In all but one of the surveys, the panel's error averaged less than 1.5%. In fact, across all surveys, the experts tied with the IEM on the mean absolute error (MAE=1.3%).

Surprisingly, the experts were most accurate in the very first survey, which was quite early in the campaign process, about 15 months before the election. For this survey, which had the longest horizon of any in our study, the average error of the experts was only 0.8%. The next most accurate surveys were, as one might expect, the two closest to the election, in October and very early November 2008. For each the error averaged about 1%.

Comparing Experts with Other Forecasting Methods. The accuracy of the experts' predictions, compared with the other three methods, appears to depend on the lead time of the forecasts. As evident in Figure 1, for the first four survey periods the performance of the experts was either better than or competitive with the performance of polls, the Iowa market, and Fair's model.

In the first survey period, ending mid-August 2007, the experts essentially tied the polls and were more accurate than either the IEM or Fair's model. The second survey, ending January 2008, was more accurate than any of the other methods' performance at that time. The third survey, ending late March 2008, was much more accurate than polls of that period, about a point more accurate than Fair's model, and about tied with the Iowa market. As late as June, the experts were still more accurate than two competing methods and tied with the third, Fair's model. In short, although the differences were not great, the experts were the most consistently accurate during all of the first four survey periods.

As the fall campaign got underway, the situation changed. By the time of Survey 5, which ended in mid-September, the experts' error jumped to 2.5%. That was the experts' worst performance during the 2008 cycle, although the polls were even less accurate during that time. In the last two surveys, covering October and very early November, the experts recovered. Their estimates were off by about 1%, which was competitive with the polls and the IEM.

2

Figure 1. Comparing the Performance of Three Forecasting Methods 2008 American Presidential Election
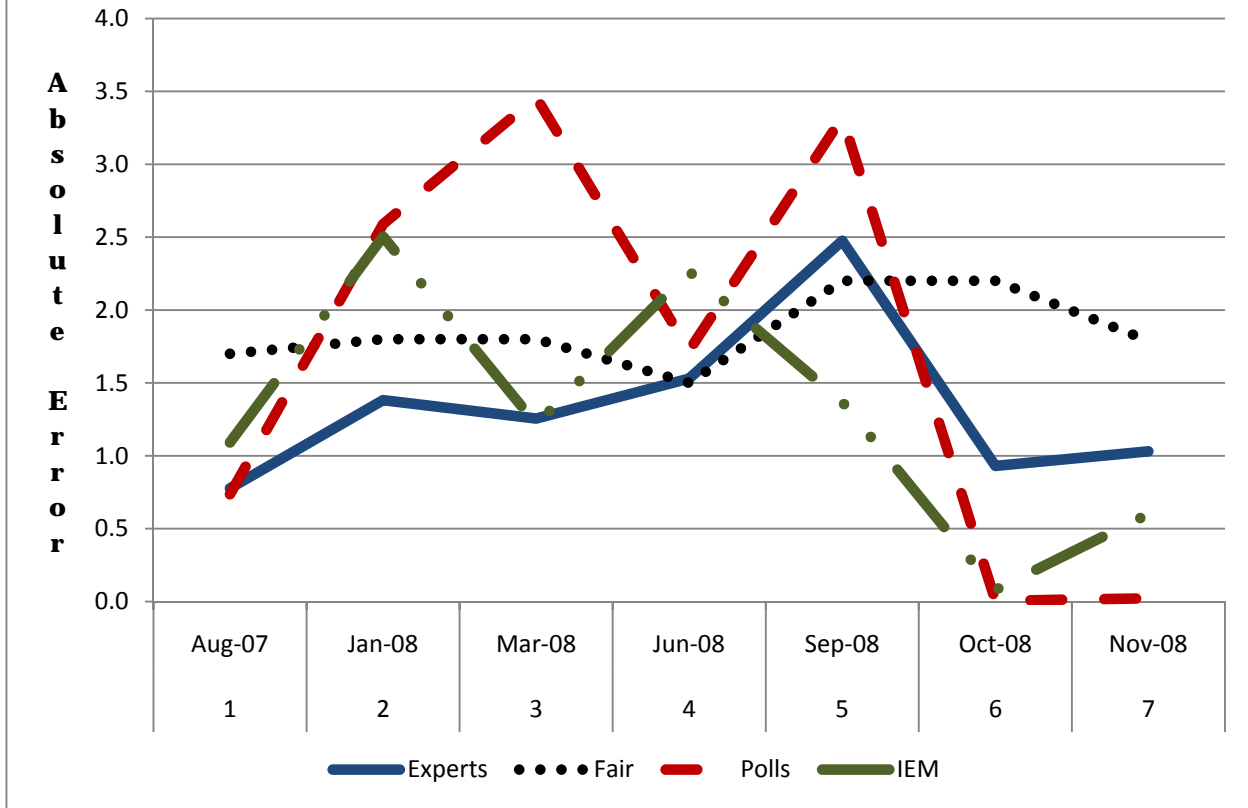
Inspection of the changing slopes in the plots reveals no apparent relationship between the experts and the polls or the IEM through the fourth survey. The experts' predictions basically are flat from the second to the fourth survey, while the polls and the IEM move in opposite directions. However, beginning with the September survey, it appears that the experts started following the polls, as the two plots shift in the same direction. It was at this time, as noted previously, that the experts incurred their largest error and the polls their second largest, contrasting with the IEM, which scored its best prediction at this time. Incidentally, across the entire period the polls scored both the highest and the lowest errors, matching the actual outcome perfectly in the last two periods.

It is also instructive to compare the performance of individual experts with each other and with the polls and IEM. Using the mean absolute error (MAE) across the seven surveys, nine of the 15 experts were more accurate than the polls and seven beat the IEM.

In sum, except for the September survey, the experts' accuracy was competitive with the best-performing methods, as evident from the experts' absolute error in six surveys being less than 1.5%. As noted, the experts were the most consistently accurate through during the first four of the seven surveys administered, and nearly matched the

polls and the IEM during the last two. Their error rate was lowest both in the surveys that were most distant from the election and in the surveys closest to the election. In other words, the panel was most accurate during the longest and the shortest time horizons. It thus appears that forecasts by experts in presidential elections are likely to be most accurate six to twelve months prior to the election and in the weeks immediately preceding the election.

Expert Surveys for 2004 Election

Turning to the 2004 presidential election, forecast results for that election are shown in Figure 2. First, recall that in that year, when the Pollyvote was first introduced, the expert surveys were conducted later than in 2008. As shown in Figure 2, the earliest survey was concluded in July; along the election year timeline, this took place between the surveys 4 and 5 in Figure 1, which ended in June and September of 2008, respectively.

Figure 2. Comparing the Performance of Three Forecasting Methods 2004 American Presidential Election
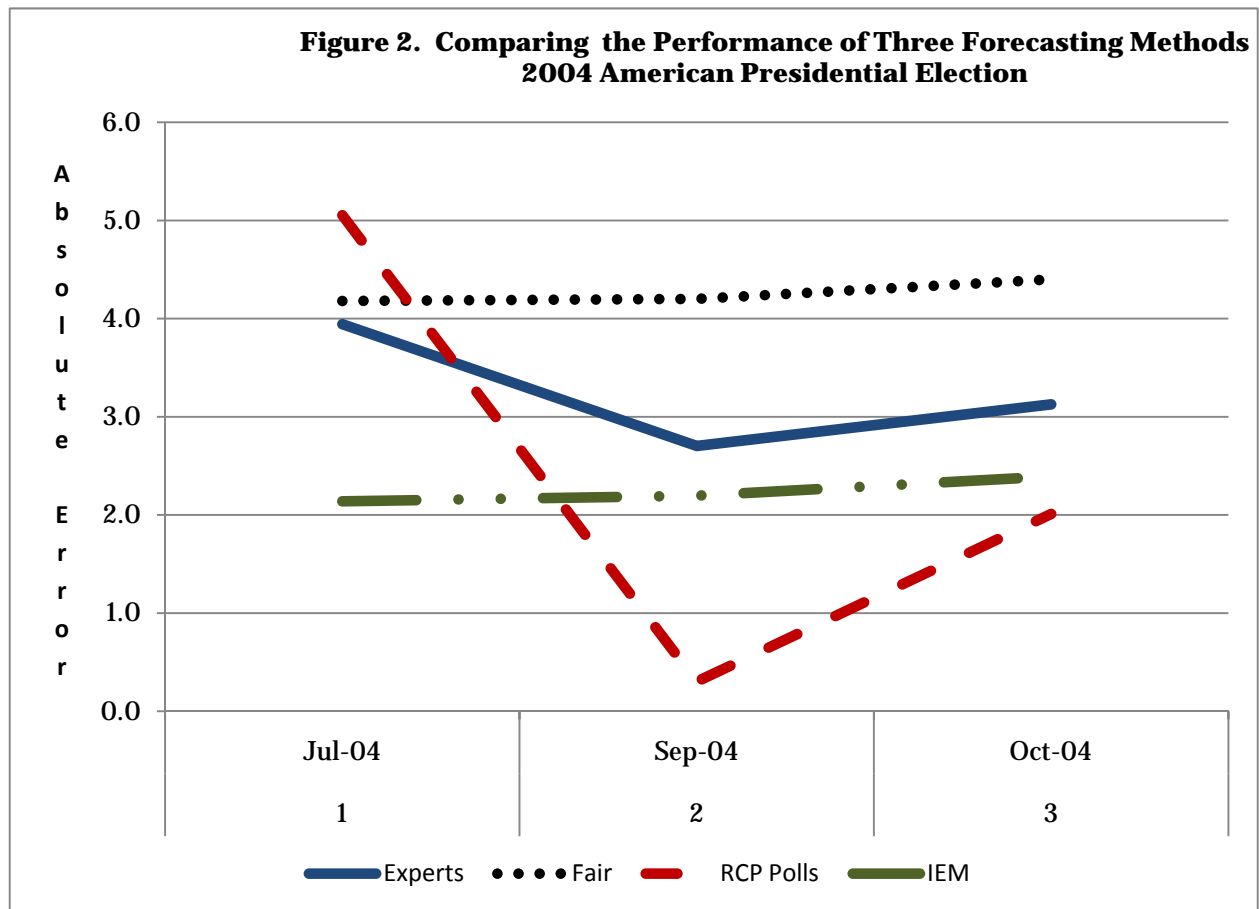
In 2004, all the methods performed less well than in 2008. On average, the MAE for each method was two to three times greater than in 2004. As in 2008, both the

worst and the best performance belonged to the polls, while on average the IEM scored the smallest error across the entire period, and Fair's model the largest.

The experts fell in the middle of the group. In the earliest Delphi survey, the panel's forecast error was nearly 4%. Although large, it was 1% smaller than the polls, and slightly lower than Fair's model. It declined to less than 3% in the next survey, and ended somewhat above 3% in the final one. Compared to the other methods, the expert panel ranked third in accuracy in the two surveys closest to the election. As in 2008, in the polls ranked first in accuracy nearest the election.

Comparing the performance of individual experts with each other and with the polls and IEM during the three surveys periods in 2004, we find that only four experts, ¼ of the total, were more accurate than the polls and that the same four also were more accurate than the IEM. This contrasts with 2008, when about half of the panelists beat the polls, the IEM, or both.

As noted in the previous discussion, in 2008 the experts' forecast error was in the 1%-2% range or less for all but one survey period. At no point was the experts' error this low in 2004. In 2008, the experts beat at least two of the competing indicators in a majority of the surveys—but not in 2004.

Thus in absolute terms the experts performed better in 2008, but so did all the other methods. However, relative to the polls and the IEM, the experts taken as a group were a little more accurate in 2004 than in 2008. This becomes apparent when comparing the three surveys of 2008 administered closest to the election with the surveys of 2004, given that both sets occurred at about the same time during the election year. In 2008, the ratio of the experts' MAE to the polls and to the IEM during the three survey periods closest to the election is 1.5:1.1 and 1.5: 0.68, respectively, but in 2004 it is 3.3:2.6 and 3.3:2.2, respectively).[3]

Interestingly, as Figure 3 shows, the absolute error of expert forecasts and of polls during comparable periods of both the 2004 and 2008 election years followed parallel paths. In both years, the error was highest in the summer (July and early September, respectively). In fact, in 2008, the experts' error peaked in the fifth survey, and it was at this time, too, that the error of the polls peaked for the second time. We interpret these findings to mean that starting in the summer the experts become increasingly influenced by polls, so that the accuracy of the latter impacts that of the former.

---

[3] We believe that the reason why experts were more accurate than polls and the IEM for 2008, compared to 2004 when the reverse was true, can in part be attributed to the earlier beginning date of forecasts for 2008. This provided a year of longer range forecasts, which normally give experts a competitive advantage over other methods.

Figure 3. Comparing the Experts and the Polls 2004 and 2008
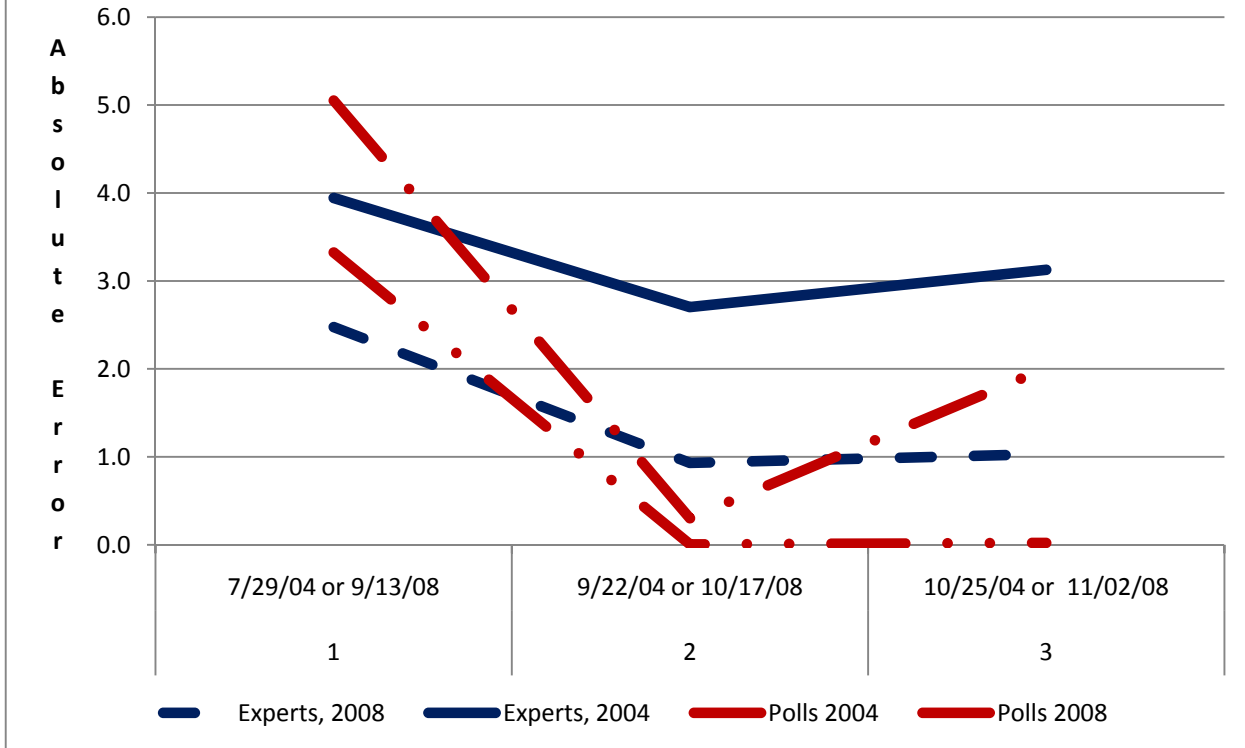
Comparing the performance of the experts who served in both panels, the correlation of their errors is -0.23 if all are included, and 0.01 if the one outlier expert who incurred the largest error of this subgroup is omitted. This finding coincides with that of Song et al. (2007), who found no consistency of success rates in predicting the outcome of American football games from one year to the next. They explained this as an instance of "regression to the mean" (Song et al., 2007, 410).

Discussion.

As we have seen, both the experts and the polls incur their highest errors during the summer. This is the time when the political parties are nominating their candidates and the election campaign is about to begin in earnest. The commotion and excitement associated with these events is registered in the polls, which vary considerably during this period. This may very well have an impact on the experts. As shown in Figure 1, beginning with survey #3 the experts' judgment appears to be guided by the polls, something that is not observed earlier in the election year.

As we noted in Figure 1, relative to the other three methods, the experts performed best in the longer time horizon. This is not surprising, for the value of expertise relative to polls and betting markets should be greatest in the long-term, when the public and bettors have not yet focused on the election in a meaningful way. Experts, who are familiar with the history of presidential elections, its patterns and

6

cycles, and influences associated with outcomes under varying conditions, are prepared to take the long view, better to anticipate what the eventual result will be. It is only as the election gets closer, when the public begins to pay attention to the campaign and their opinions begin to crystallize that the polls and the IEM outperform the experts as forecasting methods. Of course, in time the polls and markets become more meaningful predictors, and it appears that the experts begin to follow them more closely. However, given the performance of polls during the pre-election summer, especially between July and early September, during that time of the election year the experts might want to borrow a page from Ulysses and safeguard their judgment from being led astray by the siren song of polls.

REFERENCES

Fair, Ray C. 2008. Presidential Vote Equation. http://fairmodel.econ.yale.edu/

Song, Chi Ung, Bryan L. Boulier, and Herman O. Stekler. 2007. The Comparative Accuracy of Judgmental and Model Forecasts of American Football Games. *International Journal of Forecasting,* 23, 405-413.