

**Forecasting Elections Using Expert Surveys:  
An Application to U. S. Presidential Elections**

Randall J. Jones, Jr., University of Central Oklahoma  
ranjones@ucok.edu

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

Alfred G. Cuzán, The University of West Florida

Updated on October 2, 2007

**Abstract**

Prior research offers a mixed view of the value of expert surveys for long-term election forecasts. On the positive side, experts have more information about the candidates and issues than voters do. On the negative side, experts all have access to the same information. Based on prior literature and on our experiences with the 2004 presidential election and the 2008 campaign so far, we have reason to believe that a simple expert survey (the Nominal Group Technique) is preferable to Delphi. Our survey of experts in American politics was quite accurate in the 2004 election. Following the same procedure, we have assembled a new panel of experts to forecast the 2008 presidential election. Here we report the results of the first survey, and compare our experts' forecasts with predictions by the Iowa Electronic Market .

Prepared for presentation at the annual meeting of the American Political Science Association,

Chicago, August 30 – September 2, 2007

In 2004 we used Delphi expert surveys to forecast the presidential election, apparently the first time that Delphi had been used for that purpose. The resulting forecasts were quite accurate. For the 2008 election we have assembled another panel of experts and begun to survey them more than a year ahead of Election Day. This paper presents a brief description of the Delphi technique, recounts our experience with Delphi in 2004, reports early results for 2008, and offers our reevaluation of Delphi for forecasting elections.

### **The Delphi Method**

The Delphi technique includes four features: anonymity, iteration, feedback, and statistical summaries of group responses. Forecasts can be obtained from a small sample of experts; recommendations range from 5 to 20. Ideally, the experts are a heterogeneous group, each contributing specialized knowledge and insights to the forecasting task. The experts remain anonymous, and are contacted at least twice. Having been provided with summary statistics for the group and reasons offered from individual responses in the first round, the participants are then 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, using either a mean or a median. Because panelists do not meet in person, the possibility of biased responses due to the influence of personalities or individual status is reduced. (Free software at [forecastingprinciples.com](http://forecastingprinciples.com) explains the process and may be used to aid in the construction and analysis of Delphi surveys.)

Delphi has existed since the 1950s, when it was developed by the Rand Corporation and proposed for use without any supporting evidence. Despite the lack of systematic evaluation, it was used extensively in the following years. This led to a highly critical report by Sackman (1975), who concluded that Delphi was of little value.

Some research did follow. Woudenberg (1991) summarized the results of 17 studies, some published in journals and others in theses or conference proceedings. Of particular interest for our project were eight comparisons between Delphi and one-round expert surveys – known as the Nominal Group Technique (NGT). The findings showed that Delphi was more accurate than NGT in two comparisons and less accurate in five, with one tie.

Rowe and Wright (2001) summarized the research from journal articles. They suggested that the differences between NGT and Delphi could be studied by comparing the accuracy of the first round (equivalent to NGT) with that of subsequent rounds. They found that Delphi improved accuracy in 12 of 14 studies. However, their analysis failed to correct for the fact that

by using repeated measures, the subsequent rounds were based on a much larger number of estimates.

Rowe and Wright's findings are strikingly different from Woudenberg (which they fail to cite). What might explain this? Rowe and Wright mention that one study appeared to have been published in two journals, which would only lower the score from 11 to 2. In addition, Rowe and Wright coded two papers as favoring Delphi, whereas Woudenberg said there was no evidence of comparisons with NGT in those studies. Woudenberg's finding thus reduced the number of studies favoring Delphi to nine, compared with two favoring NGT. The primary difference is that Rowe and Wright restricted their review to papers published in journal articles, a procedure that leads to the well-known file-drawer bias. That is, papers that support a researcher's position with statistically significant findings are much more likely to be published (Hubbard and Armstrong 1992). By including the unpublished papers, the net result across the two studies would still favor Delphi, but it leaves much uncertainty. In particular, what non-journal studies have been published since Woudenberg's study?

Perhaps the most important deficiency in the research is the failure to provide good evidence on the value of various Delphi features, such as the number of rounds to conduct. While there is obvious value in using great care in the formulation of questions, ensuring anonymity, and providing statistical summaries, there is much uncertainty on the value of multiple rounds and the types of feedback that might contribute to accuracy. On a practical level, additional rounds consume the time of experts and administrators, and increase the time required to obtain the forecasts. Furthermore, the situation may change between rounds.

There is also a lack of information on the conditions under which Delphi would be superior to NGT. Intuitively, one would expect that Delphi would be most useful in situations in which some of the experts have information that is not available to the others. Through the Delphi feedback process, beginning after the first round, this information would be shared with other experts in successive rounds. However, this possibility has not been verified by research.

### **Application of Delphi to Election Forecasts**

We found no previous published use of expert surveys for election forecasting. Yet they seem 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. Also, they may be able to adjust for excluded variables, such as a candidate's history and personality. Given that experts are knowledgeable of the issues and candidates well before the election, they are

especially suited for long-range forecasts of elections. On a practical level, expert surveys are inexpensive, particularly in contrast to polls.

In implementing the Delphi technique to predict the 2004 election, we contacted two dozen experts in American politics. Most were from the ranks of academia, though some were in research organizations, in the media, or were former politicians. We deliberately excluded analysts who forecast elections using regression models, because that method is represented as a separate component in a multi-method election forecasting technique that we also used: the PollyVote. In the end, 17 experts participated in at least one of three surveys, each consisting of two rounds.

In each survey we asked panel members to estimate Bush's share of the two-party vote on Election Day and to explain their predictions. After the first round, the median forecast and other summary statistics for the group, along with reasons offered for the estimates, were distributed to the panelists, without identifying the individual experts. The experts were then asked to offer another estimate or to confirm their original one. We found that in the second round experts seldom changed their original estimates or provided additional reasons. In the first survey, three experts adjusted their initial estimates in the second round. In the second and third surveys none did. This experience has led us to question the value of using more than one round for each survey.

Survey results were obtained and posted at [politicalforecasting.com](http://politicalforecasting.com) in August after the Democratic National Convention, in September after the Republican National Convention, and in October after the debates. The median predictions of Bush's share of the two-party vote were: 49.5% in the first survey, and 50.5% in each of the two that followed. The actual vote was 51.2% for Bush. As might be expected, the earliest forecast (August) had the largest error of the three (1.7%), and the predicted winner was wrong. The latter two surveys forecast the winner correctly with low error (0.7%). Judged against other methods, this was a credible performance. Our primary interest, however, was whether Delphi contributed to the accuracy of our combined forecast, the PollyVote. Delphi did increase accuracy: The PollyVote forecast was off by 0.3%. Had the Delphi forecast been omitted, the error would have increased slightly to 0.7% and the forecast still would have identified the winner.

### **Using Delphi for the 2008 Election**

For the 2008 presidential election, we extended Delphi to a longer-range forecast horizon by applying it to forecast party nominations in addition to the general election. The objective is to forecast the nominees of the major parties with a lead of seven to nine months, and the winner of the general election about 15 months prior. For the initial surveys, we used the same

procedure as in the 2004 election. We assumed that the experts would benefit from exchanging reasons for their estimates in two-round surveys, given the increased uncertainty due to the longer forecast horizon.

*Experts.* As in 2004, we conducted two-round surveys. The first survey, reported in this paper, began in late spring 2007 and concluded weeks later in mid-August. Between September of this year and late summer 2008, we intend to administer quarterly surveys, switching to monthly surveys in the last two months of the campaign.

In recruiting experts, we built on our experience using Delphi in 2004. We identified a group of experts from which a panel of 15 agreed to participate through the full 2008 campaign. These individuals are all associated with academic institutions and are knowledgeable of electoral politics through their scholarly interests, personal experience, or both. This time, in contrast to 2004, an effort was made to achieve ideological balance among the panelists, in addition to our earlier efforts to have representation from various geographic regions.

*Candidates.* Eight candidates for each major party's nomination were chosen for evaluation by the expert panel. Most of those selected had announced their candidacy, although some had not. Announced candidates who were believed to have virtually no chance of winning their party nominations were omitted. Unannounced potential candidates, considered to be viable, were included to increase the likelihood that the eventual party nominees were in the selected group. Democratic candidates included: Joe Biden, Hillary Clinton, Chris Dodd, John Edwards, Al Gore, John Kerry, Barack Obama, and Bill Richardson. Republican candidates were: Sam Brownback, Newt Gingrich, Rudy Giuliani, Chuck Hagel, Mike Huckabee, John McCain, Mitt Romney, and Fred Thompson.

*Questionnaire.* As shown in Appendix A, the expert questionnaire was comprised of three short sections. In the first section, each expert was asked to estimate, for each Democratic candidate, the probability of the candidate winning the Democratic nomination. This was the question:

What is the probability (between 0 and 1.0) that each of the following candidates will be the Democratic nominee for President?

The probabilities for all eight candidates equaled one. Each expert then was asked to provide the reasons for the probability estimates that were made. The same procedure was followed in the second section for the eight Republican candidates.

In the third section, each expert was asked to estimate the percent of the two-party vote that the candidate of the incumbent Republican Party would likely receive in the general election, by answering the following question:

Considering only the major party candidates, what share of the 2-party vote do you expect the nominee of the incumbent Republican Party to receive in the 2008 general election? \_\_\_\_\_%

Of course, the answer also was an unwritten estimate of the Democratic candidate's vote, given that experts' estimates were of the Republican share of the two-party vote.

In addition, the experts were asked to provide a prediction interval for their estimated Republican vote share, by answering this question:

What is the probability (between 0 and 1.0) that the actual vote will fall within +/- 5% of the forecast that you have given? \_\_\_\_\_

The wide interval (10%) was chosen because of the uncertainty associated with the forecasts' long lead time, more than a year before the election. As with the party nomination forecasts, experts were given an opportunity to provide reasons for the election outcome they predicted.

*Rounds.* The survey was comprised of two rounds. In the first round, the experts completed the questionnaire estimating the Democratic and Republican candidates' chances of winning their respective party nominations, and forecasting the general election result. The estimates made by all of the experts, along with their stated reasons, were compiled and then returned to each panelist for the second round. This gave the experts an opportunity to revise their estimates after reviewing the others' responses. The first survey was considered completed after the second round.

### **Survey Results – Party Nominations**

The results showed that the mean absolute difference in the median probability estimate between rounds was minimal, 0.01 for Democrats and 0.03 for Republicans. In most cases, the difference was zero. Only in the case of McCain was the absolute change substantially larger (0.10). Accordingly, in Table 1 we report only the second-round probabilities for each candidate. The median, rather than the mean, is used as the forecast in order to provide stability against extreme estimates. For comparison, we report the Iowa Electronic Market estimates for July 27 through August 10, during which time most of the median experts were completing the second round survey. The Iowa prediction market makes a useful comparison because, like expert surveys, it is also a method for collecting all information available to participants (in this case, traders). However, the comparison is not exact, since the Iowa prediction market offers trading contracts only for the top three Democrats and top four Republicans. Other candidates are consolidated into "rest of the field" contracts for each party. The rank order of the estimates is similar, with only one discrepancy. In the race for the Republican nomination, Thompson ranks

third in Delphi and second in the Iowa market.

*Democratic Nomination.* Three candidates for the Democratic nomination had median probabilities of 0.10 or higher: 0.45 for Hillary Clinton, 0.30 for Barack Obama, and 0.18 for John Edwards.

*Republican Nomination.* The median probabilities of four Republicans were 0.10 or higher. The top three were clustered rather closely, with Giuliani leading the group at 0.30, Romney at 0.25, and Thompson at 0.20. McCain trailed at 0.10.

Table 1 Probabilities of Candidates Winning their Party Nominations		
<i>Democrats</i>	<i>Delphi survey</i>	<i>Iowa Electronic Market</i>
Clinton	.45	.53
Obama	.30	.32
Edwards	.18	.08
Richardson	.05	---
Biden	.00	---
Dodd	.00	---
Gore	.00	---
Kerry	.00	---
<i>Republicans</i>		
Giuliani	.30	.34
Romney	.25	.24
Thompson	.20	.28
McCain	.10	.06
Huckabee	.01	---
Brownback	.00	---
Gingrich	.00	---
Hagel	.00	---
* IEM data are July 27 -- August 10 averages.		

### **Survey Results – Outcome of General Election**

The median prediction of the two-party vote for the Republican was 48%. By comparison, the IEM closing price average over the July 27-August 10 period was 47.4.

None of the experts forecasted that Republicans will retain the White House in 2008. The median probability was 0.90 that the Republican share of the two-party vote would fall between +/- 43% and 53%. The experts thus have expressed some uncertainty over the outcome of the election. As with our 2004 Delphi study, the results differed little between rounds. In this case, the medians were the same. This suggests that, when forecasting elections, the feedback of group responses and reasons provides no new information to the experts.

### **Conclusion**

We examined the use of expert surveys for election forecasting. Our forecasts are long-range, extending more than a year before the election. This forecast horizon, with limited available information, is well-suited to Delphi because of the early insights that knowledgeable experts may provide. In 2004, we found that Delphi contributed to the accuracy of our multi-method combined forecast, but that is only anecdotal evidence from one forecast. We shall see whether survey results from the 2008 election are consistent with those from 2004.

At this early stage in the 2008 campaign, each of our expert panelists believes that the next president will be a Democrat. The reasons (not reported) vary, but the conclusions are similar. The median prediction favors the Democratic candidate over the Republican by a 52% to 48% margin, on a two-party basis.

The median probability of Hillary Clinton becoming the Democratic nominee is .45. The Republican nomination contest is much closer, with three candidates clustered at .30, .25, and .20, and Giuliani enjoying the slim lead over Romney and Thompson.

As the campaign progresses, we will continue to conduct expert surveys, generally once a quarter and more frequently during the fall campaign. In the future, however, we will use one-round surveys (NGT). This change is justified by the lack of support for multi-round surveys in prior research, as well as our findings that the results of second-round surveys have been little different than initial surveys. Furthermore, one-round surveys have the added practical advantage of saving time and expense.

*Acknowledgments:* Andreas Graefe provided useful comments.

## References

- Cuzán, Alfred G., J. Scott Armstrong, and Randall J. Jones, Jr. (2005), "Combining Methods to Forecast the 2004 Presidential Election: The Pollyvote." Paper presented at the annual meeting of the Southern Political Science Association. New Orleans. January.
- Hubbard, Ray and J. S. Armstrong (1992), "Are Null Results Becoming an Endangered Species in Marketing?" *Marketing Letters*, 3, 127-136.
- Rowe, Gene and George Wright, (2001), Expert Opinions in Forecasting: Role of the Delphi Technique. In *Principles of Forecasting: A Handbook for Researchers and Practitioners*, edited by J. S. Armstrong. Boston: Kluwer Academic Publishers.
- Sackman, Harold (1975), *Delphi Critique: Expert Opinion, forecasting, and Group Process*. Lexington, Mass: D. C. Heath.
- Woudenberg, Fred (1991), "An evaluation of Delphi," *Technological Forecasting and Social Change*, 40, 131-150.

Appendix A  
**Delphi Expert Survey Questionnaire – Round 2**

*Please type your answers in the space provided, save the completed questionnaire file, and return it by email attachment to: [armstrong@wharton.upenn.edu](mailto:armstrong@wharton.upenn.edu).*

**Part I – Democratic nomination**

What is the probability (between 0 and 1.0) that each of the following candidates will be the Democratic nominee for President? (The total for the Democratic candidates should equal 1.0.)

- \_\_\_\_\_ Joe Biden
  - \_\_\_\_\_ Hillary Clinton
  - \_\_\_\_\_ Chris Dodd
  - \_\_\_\_\_ John Edwards
  - \_\_\_\_\_ Al Gore
  - \_\_\_\_\_ John Kerry
  - \_\_\_\_\_ Barack Obama
  - \_\_\_\_\_ Bill Richardson
  - \_\_\_\_\_ Other
- 
- 1.0

Please comment as to the reasons for your above forecasts, including your assessment of issues that are important in the race for the Democratic nomination:

**Part II – Republican nomination**

What is the probability (between 0 and 1.0) that each of the following candidates will be the Republican nominee for President? (The total for the Republican candidates should equal 1.0.)

- \_\_\_\_\_ Sam Brownback
  - \_\_\_\_\_ Newt Gingrich
  - \_\_\_\_\_ Rudy Giuliani
  - \_\_\_\_\_ Chuck Hagel
  - \_\_\_\_\_ Mike Huckabee
  - \_\_\_\_\_ John McCain
  - \_\_\_\_\_ Mitt Romney
  - \_\_\_\_\_ Fred Thompson
  - \_\_\_\_\_ Other
- 
- 1.0

Please comment as to the reasons for your above forecasts, including your assessment of issues that are important in the race for the Republican nomination:

**Part III – General Election Opponents**

Considering only the major party candidates, what share of the 2-party vote do you expect the nominee of the incumbent Republican party to receive in the 2008 general election? \_\_\_\_\_%

What is the probability (between 0 and 1.0) that the actual vote will fall within +/- 5% of the forecast that you have given? \_\_\_\_\_

Please comment as to the reasons for your above forecasts, including your assessment of issues that will influence the general election outcome:

We invite your comments and suggestions about the survey process (including the spreadsheet layout).  
Please type them here.