

## ELECTION FORECASTING BY ABC NEWS

Richard C. Taeuber  
Leasco Systems & Research Corporation

In their December, 1964, critique of that Fall's election night programs on network television, TELEVISION Magazine commented that "The computers, standing shoulder to shoulder at the television networks' nerve centers on election eve were so packed with information it hardly seemed necessary for anyone actually to vote. The machines probably could have determined the whole course of the Presidential election from the raw data supplied by one man's ballot, preferably a high-income Negro Catholic living in a split-level house with two and three-tenths children in a Midwest Polish neighborhood that went for Alf Landon in 1936."

Since the first useage of computers in 1952, the election eve coverage has been possibly the greatest single public exposure of working computers. Their use in assisting in election night forecasting and analysis is an application unlike no other in the computer business, in that this application calls for a fully reliable operation, and no one or two week postponement of election day is possible if a program bug occurs. With the heavy involvement of computers in election coverage, it must be emphasized, indeed it cannot be stressed too strongly, that computers do not vote, they do not determine the outcome of any race, they do not merely total votes, nor do they guess at the projected results. Computers do assist the human analytical staff in comparing present happenings with past occurrences, and projecting the final outcome of this particular election. Technically speaking, if the coverage was of but a single race, there would be very little need for computers. However, with the network analytical staffs trying to keep track of what is happening in the Presidential contest in 50 states and the District of Columbia, in senatorial races in 34 states, gubernatorial races in 21 states, and in 435 congressional districts, the need for electronic assistance is evident.

In the projection of election results, the analytic staffs, served by computers which use procedures supplied from the minds of mortal man, face an interesting accuracy-tolerance problem in the calling of the winners of the individual races. In a race where one candidate receives 65% of the vote, then accuracy within 10% is more than sufficient to allow a correct call of the race. If that winning percentage slips to, say 50.5%, in a two-party situation, then accuracy within six-tenths of one percent may not be sufficient to prevent calling the wrong winner, although, from a statistical point of view, such accuracy (0.6%) would be really quite remarkable for the amount of information generally available. Although the emphasis in election eve forecasting in earlier years was on projecting winners, the coverage on all three networks now has progressed to the point where in 1968 not only are the winners forecast, but also the vote-

split, the plurality or margin of victory, and the total turnout. Further, the various why's and wherefore's of a particular victory will be analyzed and discussed on the air to a greater extent than ever before.

### Projection Model

With the threefold replication of this single application by the networks, one might raise the question of why the projections and analyses are not handled on a pooled basis by the three networks. Although in some sense the techniques employed by the three networks have similar ingredients, the merging together of these basic ingredients is handled quite differently by the three television networks. How, then, do we at ABC make the projections? We use a projection model which is based on a weighted combination of a baseline (or time zero) estimate, key precinct information, and raw vote information. Mathematically, the general form of this model can be expressed as

$$P = \alpha B + \beta K + \gamma R$$

where  $\alpha + \beta + \gamma = 1$

P = overall projection

B = baseline estimate

K = key precinct estimate

R = raw vote

The model is deliberately kept simple because we feel it important that the analytical personnel be able to understand not just the final numbers that come out but also the individual components of the model. The actual model is based on what we refer to as deviation analysis or swings, meaning that the model works with the difference between historical data and present day happenings, rather than with the absolute levels of vote being recorded in those reporting units incorporated into the model. The use of swings rather than absolute levels of vote may at times pose a problem for the analytical staff in years such as this, for there are no recent presidential races with three candidates and thus no three-party historic data is available at the precinct level. To cover a three or four-party race, then, a much greater burden is placed on the humans who are interpreting the computer output and thus, again, the need for a comprehensible model. It might be added that from the very beginning of the planning for the 1968 coverage, all systems planning has provided for handling as many as four candidates in any given race, and the move above two candidates really presents very little problem for the computer aspects of this forecasting exercise.

In the model described earlier, the weighting coefficients --  $\alpha$ ,  $\beta$ , and  $\gamma$  -- are computed to reward consistency in that they are inversely related to the variance of the information element with which they are associated.

They vary over time as the various forms of reports reach the ABC studios, but consistently total to one. At the beginning of the evening, the total weight is on the baseline for neither key precinct information nor raw vote information is available to us. As the evening wears on, the key precinct information is the first to come in and that weight begins to climb, by design never completely supplanting the baseline. Later in the evening, as the raw vote begins to come in, that weight climbs and the weights on the baseline and key precinct information begin to decline. Finally, in the model that we use, at the end of the "evening", the coefficient for the raw vote information goes to one. Thus, at the end of the evening, if the raw vote tabulation has also been completed, our model is in agreement with that raw vote tabulation. It might be added, in a parenthetical note, that such a system presented problems in 1960, for in that year in California the "final" tabulated vote showed Kennedy with a slight lead over Nixon in the Presidential contest. However, California, in that year, tabulated absentee ballots some two weeks after the election. The absentee ballots were so heavily for Nixon as to swing the victory in that state's Presidential contest from Kennedy to Nixon. It must be noted that California, in 1968, will tabulate its absentee ballots on election night, as will most other states.

Let us look into the three aspects of this model individually.

#### Baseline Estimates

The baselines are really a time zero projection based on any and all prior information available to the ABC staff. They are, in a sense, the best subjective forecast by the projection staff based on polls, on educated opinions, on informed judgments, and any other sources of information available to the network staff prior to the receipt of actual vote information on election day. The estimates are numeric in nature and give the exact projected vote split, not merely an estimate as to who will be the winner.

#### Key Precinct Estimates

Key precincts are the primary information source on which all three networks base both their projections and their analyses. The term "key precinct" refers to precincts which are selected by one of a variety of means, and then staffed by a network representative who phones the results directly to the computer center. This provides information that can be received and analyzed as a sample of the total vote prior to its inclusion in the normal collection process. The key precincts are selected on a probabilistic basis, either completely randomly, or they may be constrained in their selection to represent various strata or ethnic groups. For the 1968 coverage, ABC selected its precincts in a two-stage sampling process. In the first stage,

communities or sub-elements of the state were selected with probability proportional to the voting age population of that sub-element of the state. Within the communities, precincts were selected randomly but were oversampled. The oversample was then weeded down based on considerations of historic performance, on the availability of historic data, and on the potential availability of the results election night. If the vote from an individual precinct would not be available within a reasonable time frame on election night, if at all, that precinct was discarded from the list of key precincts. It does a forecasting model little good to receive the report of the key precinct after 80 to 90% of the raw votes of the state have been tabulated. The precincts, after being selected, are researched to give both their historic performance in several immediate prior elections and also their ethnic composition. This latter information will be used in the portions of the computer program which project and help analyze the whys of the election and how the various ethnic blocks are voting or not voting. The need to obtain historic information also provides certain problems with precincts whose boundaries or whose composition, in any one of a number of senses, have changed since the preceding election. For example, the precinct may have new boundaries, it may have a new high rise apartment, or it may have been partially cleared for redevelopment within the last two years. Because of the mobility and dynamic changes which are so widespread in this country, one cannot simply rule out those precincts which have undergone such changes, for to do so would eliminate a very major portion of the electorate from objective consideration in the projections.

But the statistical selection of the sample precincts is only part of the battle. These precincts must be staffed to report quickly and directly to the studio. Here there are problems! ABC uses a staff of one or two on-site persons to handle each key precinct. This staff is generally supplied by the League of Women Voters in each state. They visit the precinct several weeks in advance to determine, among other things, the availability of a phone for quick reporting to the ABC studios. If there is no phone available for use in the immediate vicinity, then ABC has a private telephone installed for use of the reporting team. In one instance, in 1966, even this presented a problem, for the precinct was in a rural farm house and the farmer would not permit a phone to be installed in his house. The resourceful team members looked around outside, but they could not find a suitable building in the vicinity. They did notice a very large oak tree, and, following a request to the telephone company, a phone was installed in the oak tree. There are also additional problems that the field reporters face, such as a family of skunks under the floor of a precinct polling place; precincts in which only Spanish was spoken; fuses which blew; fire in telephone company offices; and, in more than one instance,

teenagers using the public booth telephone who may still have been on the line the next morning. The reporters also performed a variety of chores, including providing dinner or snacks to permit election officials to get on with the ballot counting without taking a dinner break. And the reports from 1966 also include comments on the problem of how to prevent absentee ballot-box stuffing and observed instances of unlocked ballot boxes.

#### Raw Vote

The final component of the model is the raw vote as it is reported by the News Election Service. The News Election Service is a step-child of the three networks and the two main wire services. It was started in 1964 when it became very obvious that a five-fold tabulation of the raw vote made little practical sense and could be done much faster, more completely, and much more economically on a pooled basis. This pool in 1968 will receive reports from about 80% of the precincts in the United States as well as from the county seats of all the counties in the United States. The tabulations will be performed by computer for all 50 states, and the source which provides the most vote, at that moment, that is, either the NES tabulation of precinct returns or the county reports, will be forwarded to the members for broadcast or relay to the viewing, listening, or reading public.

The incorporation of the raw vote into the projection model calls for an awareness of the patterns of reporting within many of the states. For instance, in New York State, it is well known that New York City is reported before the rest of the state, and that it casts a higher proportion of Democratic votes than the upstate portions of the state. There is a similar well known pattern in Illinois, with Chicago or Cook County reporting earlier and being more preponderantly Democratic than downstate Illinois. Similar patterns, moreover, hold in varying degrees in a good many states where the early vote may be more Republican or more Democratic and the discrepancy may disappear monotonically, or may, in fact, swing the other way during the middle of the tabulation process, and then later disappear. It might be noted

parenthetically that the faster tabulations which will be produced in '68, with computer assistance, add an additional element of uncertainty to the accuracy of these historic reporting patterns. One way to minimize the effect of this uncertainty in reporting patterns would be to use finer geographic breakdown than the state level to permit the model to compensate for any such pattern. Such segmenting is being done, but the timeliness of the availability of aggregated data may force whole-state useage at various times throughout the evening. When this occurs, a correction factor may be entered to compensate for the historic patterns. Such a correction factor would normally be expected to go to zero at the end of the evening. However, one might refer back to the earlier comments about California in 1960 to note that perhaps this correction factor should not go to zero in certain states. Although the computerization of the election tabulations and the movement throughout the country towards nonpaper ballots (and thus faster counting) may disrupt the patterns of statewide reporting, the computerization does provide timelier availability of the data which permits the use of finer geographic breakdowns, even to the county level.

In closing, one might draw an analogy with weather forecasting. Both weather forecasting and election forecasting are done by technically competent staffs which have to make allowance for many factors outside of the control of the projectionist. One case leads to ruined picnics; the other case may salvage a supposedly "ruined" evening by providing the hope that the forecasters on television have "goofed" in some manner. The weather forecasters receive a bit of tolerance for their prognostic inaccuracies and are judged by their peers for their interpretation of the available evidence. Unfortunately, those who make judgments as to the competence of the election forecasters do not have knowledge of the totality of the available evidence, or even an indication as to what that available evidence is. Election projectionists are allowed little leeway for error in their interpretations. The volatility of the American voter continues to present a problem to those who are involved in election forecasting, but the credibility of the projections has run high in the past and is increasing.