# Underlying Motivational Indicators behind Voting Behavior in Major Elections: A Statistical Methodology

Dhiren Ghosh
Synectics for Management Decisions, Inc. Arlington, VA, 22209, USA
Dhiren@smdi.com

## Abstract

The fundamental assumption in voting behavior is that it is not random, at least for (presumably) intelligent voters who are not committed to one candidate as a hardcore Democrat or a Republican in the US or Conservative or Labour in the UK. If this fundamental assumption is correct, there are factors underlying motivational indicators behind the voting behavior. The following statistical methodology is used to identify them. The methodology starts with exhaustive pretests and methodological studies to develop a set of questions which are pairwise uncorrelated or orthogonal for the voters well before the election to determine the relative importance of underlying issues. These include the political issues such as taxation, fiscal policy, job creation strategy as well as demographic factors of the candidate such as age, race, gender, education and extend of political experience and also demographic factors of the voters. The initial steps of the statistical technique of factor analysis usually plays an important role in developing a set of a few questions whose answers are approximately pairwise orthogonal. This is completed before the election. After the election we find out the actual voting behavior and perform logistic regression with the dichotomous voting behavior and the results of the motivational components. The coefficients of multiple logistic regressions indicate the motivational factors and their importance (after standardization by dividing by their standard errors).

Key Words: Factor analysis, logistic regression; orthogonal; Path analysis

It is not uncommon in sociological enquiry to study the attitude and values of a particular experimental population buy collecting data through a well developed questionnaire developed for the specific purpose. By detecting correlations that are statistically significant, sociologists and statisticians can use attitudinal data to predict the future behavior of various subgroups of a population of interest. The situation can also be reversed, however, using the product use data from a subpopulation to identify the probable motives behind their actions. Particularly in fields such as marketing and politics, product usage data may be used to identify issues of importance for customers and voters in major elections. In this paper we develop a statistical methodology that identifies the motivation behind the voting behavior in a major election.

We assume at the outset that a (presumably) intelligent voter takes voting in a major election to be a serious matter. We envisage a mathematical model underlying the voting behavior of voters in elections. The voters assign weights to each of the major issues and we assume the weights add up to 100. If the voter agrees with a candidate's position on a particular issue, the total weights for that issue goes in favor of the candidate. If the voter partially agrees with the candidate's position on an issue, only a fraction of the total weight is assigns to the candidate. If the voter completely disagrees with the candidate's position on that issue, the voter assigns a zero weight to the candidate. The voter goes through the list of issue for each candidate and the candidate who gets the most weight is the voter's choice. Hence he/she weighs the issues carefully and they arrive at a

conclusion or choice of the candidate. The usual procedure for determining the rationale for the voting behavior is to interview a random sample of actual voters after the election and ask him/her why he/she voted in the manner he/she behaved, i.e. why he/she voted for the candidate of his/her choice. This is usually done in two different ways. The voter is asked to list the reasons in a decreasing order of importance for the decision, alternatively the voter is given a list of reasons prepared by the surveyor beforehand and the voter is asked to number in order of importance perceived by him/her, starting with 1, 2, and 3 in decreasing order of importance.

However, our experience shows any such survey conducted after the election when the results are known, the voters who voted for the losing candidate tend to justify their action rather than give the true reasons for their choice. On the basis of our experience we reverse the procedure. We select a random sample of (highly) probable voters as yet undecided and then ask them to weigh the factors that motivate them to vote for a candidate of their choice. The following list is an example of the possible motivational areas in a federal election in the U.S.:

1. Taxation policy
2. Foreign policy
3. Defense spending
4. Attitudes toward Social Security and Medicare
5. Action to reduce unemployment
6. Political party affiliation
7. Gender of the candidate
8. Political experience of the candidate
9. Age of the candidate
10. Immigration reform
11. Gun control

Such a list is developed after extensive pretesting and methodological research and may involve statistical technique of Factor Analysis to ensure orthogonality of the issues, i.e. that low or absence of correlation among the issues. The candidates are asked to assign a weight to each issue that reflects their magnitude of the motivational factors behavior their voting behavior. The weights should add up to 100 so that the motivational factors of different voters are in a comparable framework.

This subpopulation is then interviewed again after the election. The data obtained can then be analyzed in two different ways.

In the simple method, the voters are divided in two groups on the basis of their voting record. For each group the weights are arranged (mean or median) for each issue and a rough index developed and the result may indicate the magnitude of motivation factors that determines the voting behavior of voters.

However, a more sophisticated method would be the application of the logistic regression model otherwise known as the logit model. The dependant variables of a dichotomous logistic equation 1 or 0. Transforming the probability to odds reduces the upper bound. We then take the logarithm of the odds to remove the lower bound. If $x_1, x_2, \ldots x_n$ are the explanatory variables. The model takes the form:

$$\ln\left[\frac{p}{1-p}\right] = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_n x_n$$

Where $p$ is the probability that $y$ = 1 (corresponding to one candidate). The term on the left is referred to as logit or log-odd. The $x$s are the weights ascribed to by the respondents.

The logistic regression model does not include a random error term as in regular regression.

The logit equation may be solved for $p$ to obtain:

$$p = \frac{exp[\alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_n x_n]}{1 + exp[\alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_n x_n]}$$

There are other methods such as the probit and the log-log model, but the logit model is preferred because:

1. The regression coefficients are related to the odds
2. It can be extended for three or more categorical variables, i.e., three or more candidates.

Further the $\beta$s can be compared after dividing each $\beta$ by its standard error. The standardized $\beta$s provide the relative magnitude of the motivations behind the voting pattern. We then reverse the process by ascribing 1 to the other candidate and 0 to the candidate considered above. This, in turn, will provide the motivations of the voters voting for the other candidates.

The use of Path analysis to examine causal structure among continuous variables was pioneered by Wright and had been popular in providing a graphical representation of a set of algebraic relations among variables. It shows the direct effects, the causal chain, direct plus indirect effects, and also points out spurious effects. Path analysis also helps the decomposition of variance and zero-order correlation.

Logistic regression can also be applied where the number of dependant variables is more than two. Polytomous logistic regression for unordered categorical variables can be applied for political party identification by determining the effects of individual predictors.

**References**

Allison, Paul D. (2012) *Regression Using SAS. Theory and Applications (2nd ed.).* SAS Institute, Inc., Cary, NC.

Harman, Harry. (1970) *Modern Factor Analysis*. University of Chicago Press, Chicago, IL.

Hosmer D.W. and Lemeshow. (2012) *Applied Logistic Regressions (2nd ed.)* John Wiley and Sons, New York.

Menard, Scott. (2010) *Logistic Regression. From Introductory to Advanced Concepts and Applications*. Sage Publications Inc., California.

Wright, Sewell. (1960) *Path Coefficients and Path Regressions: Alternative to Complementary Concepts?* Biometrics, vol. 16, No. 2, 189-202.