Some Probability-Based Approaches for Investigating the 2012 General Election Exit Poll

Clint W. Stevenson – Edison Research

Abstract
Election exit polling provides a unique opportunity to collect a variety of information on the voting population immediately after votes are cast on Election Day. Traditional election exit polling approaches normally use classical statistics. This paper provides an introduction and some examples of alternate ways to evaluate and analyze these exit poll data. These approaches include Dirichlet simulation, Dirichlet process clustering, multinomial electoral college simulation, and Bayesian regression. This exit poll framework provides an opportunity to apply these various distributions and simulation approaches as well as options to visualize the data in probabilistic ways. The primary focus in this summary will be to discuss the probability distributions associated with these data and the application of the distributions.

Keywords
Bayesian — Exit Poll — Posterior Distribution — Dirichlet — 2012 Election — Electoral College — Simulation

Author Contact: cstevenson@edisonresearch.com

Contents

Introduction

Probability and simulation methods are becoming increasingly common in many industries. Often the normal distribution is used and confidence intervals, standard errors, and the like are based on this distribution. The goal is to use approaches that best describe the data in meaningful ways. Options range from classical approaches to more modern probability approaches. Though using probability distributions and, in particular, Bayesian analysis are not new to election exit polling, it is often left as an academic exercise rather than applied in practice. In this summary, the goal will be to take several probability distributions and apply them to a large-scale exit poll framework. This paper will primarily discuss multinomial data and the probability distributions associated with these data. Ultimately, the goal of this paper is to apply techniques and approaches that have not traditionally been used in an exit poll framework.

Exit polling is fairly well established. Current exit polling approaches work well and provide very good insight into the voting populations. These approaches presented here should not be seen as a different or better way to analyze exit poll data but as complementary and as a way to provide further understanding of the data. These approaches will provide additional ways to understand election exit poll results as well as ways to visualize the probability distributions of the data. There are 50 states plus the District of Columbia where exit poll data are available. Consequently, results for all states cannot be provided in this paper. For this discussion Florida will be used as an example because, in addition to it being the closest race in the 2012 election, it provides a large amount of data that can be easily used in a number of settings.

1. Design of the 2012 Exit Poll

The national exit poll is a unique source of data because it does not rely on telephones to reach respondents and it only interviews those who actually voted. In the 2012 general election the National Election Pool (NEP) – ABC, The Associated Press, CBS, CNN, FOX, and NBC – commissioned a survey of voters in all fifty states and the District of Columbia from Edison Research. This consisted of a national and state-specific surveys. Of the surveys there were 19 states where the sample size was too small for individual state demographic or other breakouts. The majority of interviews are conducted in-person on Election Day in a probability sample that is stratified based on geography and past vote.

In recent years the number of voters who cast their ballot early or by absentee has increased. Consequently, the NEP has supplemented the in-person exit poll with a telephone survey conducted between October 26, 2013 and November 4, 2013. This early voter telephone survey includes a landline as well as a cell phone only component. Follow up questions identify the number of landline phones in the respondent's home and if the respondents also has cellphone. In this way the respondent can be classified as "landline only", "both landline and cell", and "cell phone only". In these state (and national) samples a pre-defined target of 30% of the respondents are called directly to cell phones.

Election Day respondents complete the exit poll by filling out a paper questionnaire. In 2012 there were five versions of the national exit poll questionnaire. In the state exit poll the number of questionnaires ranged from one to three.

2. Methods

The analyses presented here will examine techniques that can be applied to exit poll data in all 50 states and the District of Columbia. However, similar methods and approaches will perform equally well in other survey and polling research. When the exit poll sample is selected, polling locations from each state are randomly chosen with a known probability of selection based on previous election voting behavior. Voters are randomly selected within each polling place using a predefined sampling interval. Samples for each state are independently sampled from each other. The approaches used here serve to further research and to highlight the importance of probability when working with exit poll data.

The entire process of conducting an exit poll is complex with many moving parts. There are many statistical considerations that must be taken into account at the polling place, county, congressional district, state and national levels. These must also be incorporated into the operational and administrative components of an exit poll. Statistically, there are many probability distributions that are used as well as a mixture of those distributions. The distributions are both continuous and discrete. The goal in this paper is not to provide a comprehensive and exhaustive discussion of the intricacies of the operational and statistical aspects of an exit poll but to provide additional discussion on various ways to incorporate probability distributions into an exit poll framework. The core of this discussion is based on discrete data in the exit poll. The examples used in this paper will be based on the data obtained from the 2012 presidential election and will specifically address the use of the Dirichlet and Normal distributions.

3. Discussion

This discussion will specifically focus on four different probability techniques. First, the Dirichlet distribution and its relationship to candidates within a statewide election. Second, using the Dirichlet process to cluster similar precincts. Third, applying the Dirichlet/Beta distribution to obtain the probability of winning each state and then applying the binomial distribution to allocate each state's winning electoral votes. Fourth, using Bayesian principles to obtain a regression model.

3.1 Dirichlet Distributions for Statewide Estimation

An understanding of how exit poll data are distributed is critical to the results. The Dirichlet and Beta distributions are widely used as distributions for the multinomial and binomial distribution, respectively. The Dirichlet distribution is the multivariate generalization to the Beta distribution. Due to the nature of election exit polling, these distributions work well in this context. The Dirichlet distribution provides the flexibility to work with a wide range of exit poll data. Generally, data gathered during an exit poll (distinct candidates, Likert scale, polling locations, etc.) follow a multinomial distribution making the Dirichlet distribution an appropriate choice.

At the simplest of levels, an individual polling location can be estimated and probabilities obtained given the polling location data. For the 2012 presidential election generally only two candidates are considered – Barack Obama and Mitt Romney. The Dirichlet distribution provides the ability to obtain the posterior for this scenario and can be used as the prior using multiple parameters $Dir(\alpha)$ where $\alpha_1,...,\alpha_\kappa$ representing the vote for each candidate. In the case where there are only two candidates, a Beta distribution would work. However, the Dirichlet distribution easily makes the generalization for any number of candidates. Consequently, one can estimate the outcome, given these data, voting the same way in Equation 1.

Take for example an arbitrary state s with polling location i, containing m respondents in that polling location. We can simulate the probability of a win for candidate j in that polling location using the Dirichlet distribution in Equation 1.

$$Dir\left(\alpha_1+1,...,\alpha_\kappa+1\right) \tag{1}$$

Using this approach the probability of each candidate winning a particular polling location in each sample of size n can be calculated. From that point the total probability of a candidate can be simulated. Given a known sample weight and that each polling location has a known probability of selection the total probability that a candidate will win can be calculated.

Take for example Florida. In 2012 it was the closest presidential race in the country with the candidates differing by 0.88 (Obama: 50.01, Romney: 49.13) percentage points. In Florida there is a sample of n=49 exit poll locations. Suppose for location i we observe a sample of size 125 with 67 votes for Romney and 58 for Obama. This single polling location produces a distribution of $Dir\left(67+1,58+1\right)$. This can then be iterated for all polling locations in the sample. All the polling locations can then be combined (using each polling location as the primary sampling unit) to produce an overall probability that a candidate will win the state. In the case of Florida, as seen in Figure 1, we can see that, given the data, the probability that Obama will win is no better than flipping a coin.

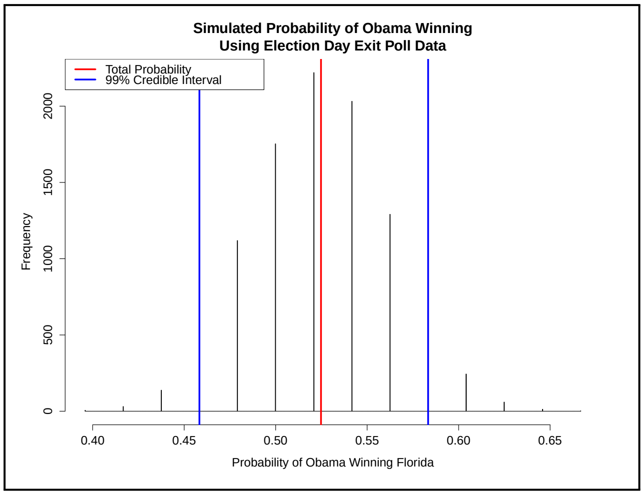

Figure 1. Probability That Obama Wins Florida

3.2 Dirichlet Process Clustering

Though the Dirichlet distribution itself is a continuous distribution it is often used for the parameter vectors in a multinomial (discrete) distribution. In this way it is considered a distribution over distributions. This concept of having a distribution over distributions can be used in classification approaches to identify and place observations into groups. The following three analogs show ways that observations can be clustered and assigned to groups.

3.2.1 Chinese Restaurant Model (CRM)

Suppose in the Chinese Restaurant Model (CRM) there exists a restaurant with an infinite number of tables. At each of these tables there are seats for an infinite number of people. The first customer walks into the restaurant and takes a seat at one of the tables. The next customer enters the restaurant and takes a seat at a new table with probability $\frac{\alpha}{1+\alpha}$, or selects the same table as the first customer with probability $\frac{1}{1+\alpha}$. Where α is a user-defined dispersion parameter. This continues until the $(n+1)^{st}$ customer sits down at a new table with probability $\frac{\alpha}{n+\alpha}$ or table k with probability $\frac{n_k}{n+\alpha}$, where n_k is the number of people already sitting at table k.

3.2.2 Pólya's Urn Model (PUM)

Similar to the Chinese Restaurant Model is the Polya Urn Model (PUM). In this way we can take an urn that contains $\alpha \cdot G_0$ balls of a specified color. We randomly select a ball from the urn, return it to the urn, and add an additional ball of the same color to the urn. This process continues through an infinite number of balls. As this process continues there begins to be a clustering of the same ball colors.

3.2.3 Stick Breaking Model (SBM)

A third approach is known as the Stick Breaking Model (SBM). Here it can be envisioned that there is a stick which is broken at point B_1 where $\beta_1 \sim Beta(1,\alpha)$[1]. The first (left) part of the stick is v_1 and the second (right) part of the stick is v_2. The right side of the stick is then broken again at $\beta_2 \sim Beta(1,\alpha)$ and the stick to the right of that is now $v_2 = (1-\beta_1)\cdot\beta_2$. This is similar to the Chinese Restaurant Model but in this case a person is assigned to the first group with probability v_1.

3.2.4 Infinite Mixture Models and Clustering

The connection between the CRM, PUM, and SBM models can be seen as the table, color, and stick are each a distinct partition where α is used as a dispersion parameter. These analogies can be extended into survey work, and in particular, exit polling. Where the table/color/stick can be compared to groups or partitions of polling places and, for example, the customer at the table can be considered the polling place itself. As new data from polling places are added new partitions can be considered or the polling place can be placed in an already existing partition. This could, in theory, continue indefinitely. However, in reality there are a finite number of polling places and a finite number of partitions that can be created. However, this provides a flexible approach to classify polling locations with other, similar polling locations.

A common approach to classify and group observations is *k-means clustering*. However, using k-means requires a fixed number of partitions to be defined in advance. Once the number of partitions is defined, the observations are then assigned. Sometimes this works well but other times identifying the exact number of partitions in advance is not feasible. Clustering the number of partitions using the Dirichlet process is based on the polling locations and the dispersion parameter (prior). Therefore, the number of partitions does not need to be determined in advance.

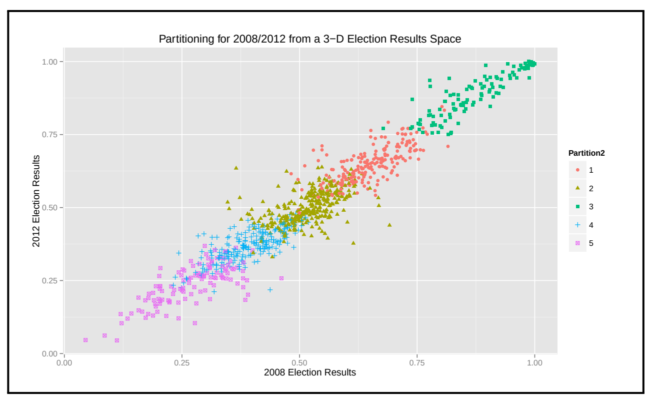

Figure 2. 2008 and 2012 Reported Vote and Partitioning Using a Multi-Dimensional Dirichlet Process Clustering

Though there are many exit poll variables and varying metrics that can be used to assign group membership, the general approach remains the same. The partitions and their group membership can be assigned new elements while allowing new partitions to be formed as new data is obtained. Figures 2 and 3 provide a political partition based on 2008, 2012, and 2012 exit poll vote. This process allows for potentially

[1] Note that the expected value of a Beta(1,α) is $\frac{1}{1+\alpha}$

an infinite number of partitions but uses a "rich-get-richer" approach meaning that already existing partitions are more likely to gather additional elements.

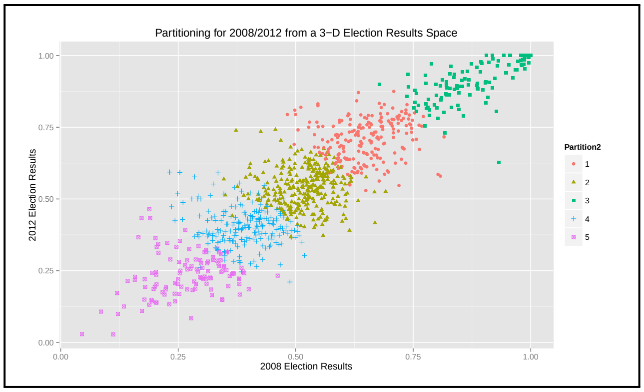

Figure 3. 2008 Reported Vote and 2012 Exit Poll and Partitioning Using a Multi-Dimensional Dirichlet Process Clustering

3.3 Multinomial Electoral Vote Simulation

A useful characteristic relating to probability distributions is the ability to use known data and then simulate from the posterior distribution. Using the exit poll framework, the statewide candidate estimates can be used and applied using the Dirichlet distribution approach. This means that the estimates from each state[2] can be used to determine the probability that a given candidate will win each state. The vote from the electoral college is, for the most part, awarded using an all-or-nothing approach[3]. Consequently, the Binomial distribution fits this situation very well.

Using the statewide exit poll estimates from the 2012 Presidential election we can sample from the electoral vote posterior distribution. To focus on the concept of using probability approaches on exit poll data, the completely random sample of respondents approach is applied to the data as shown in Figure 4. Additionally, by comparing the completely random sample in Figure 4 to Figure 5 one can easily observe the importance of correctly incorporating the primary sampling units and the design of the sample. This figure identifies the distribution of electoral vote outcomes as well as the 0.005 and 0.995 quantile points. Further, for comparison, the 270 threshold to win the presidential election is identified with a dashed green line. The actual outcome of 332 electoral votes based on the 2012 electoral college is shown with a solid red line. Here the probability that Obama will win over Romney is calculated for each state using the Dirichlet distribution. With the probability of success established for each state we can incorporate these probabilities into a winner-take-all Binomial distribution for all 50 states and the District of Columbia. This is equivalent to flipping a weighted coin for each state and allocating the electoral votes for each state based on the outcome of the weighted "coin flip". This procedure is then replicated N (in this example, ten million) times to create the posterior distribution as shown in Figure 4. For the 2012 posterior distribution the median number of electoral votes is 338 and the mean number is 336.

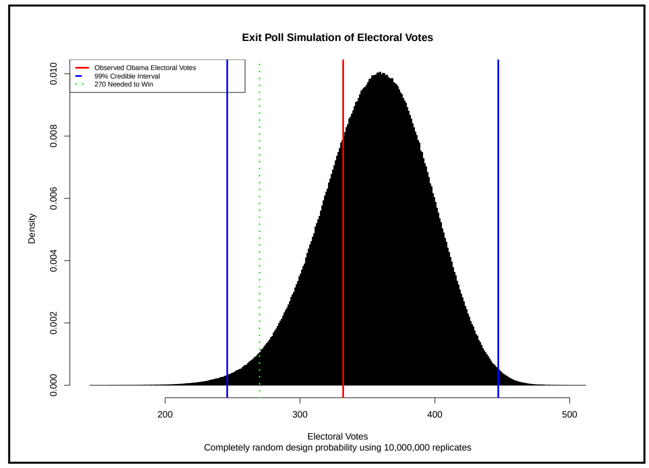

Figure 4. Posterior for Obama Electoral College Without Using Precinct Weights

This same procedure can be applied to better correspond with the primary sampling units within each state. As noted in Section 3.1 the probability that a candidate will win the state is described using the responses from the precincts and then calculating the total probability of success in a hierarchical way. These statewide probabilities are then used to sample from the posterior distribution. The differences in Figure 4 and Figure 5 highlight how the probability of a candidate winning varies as a result of measuring *primary sampling units* compared to only *secondary sampling units*.

Clearly, 'calling' a national election based purely on sam-

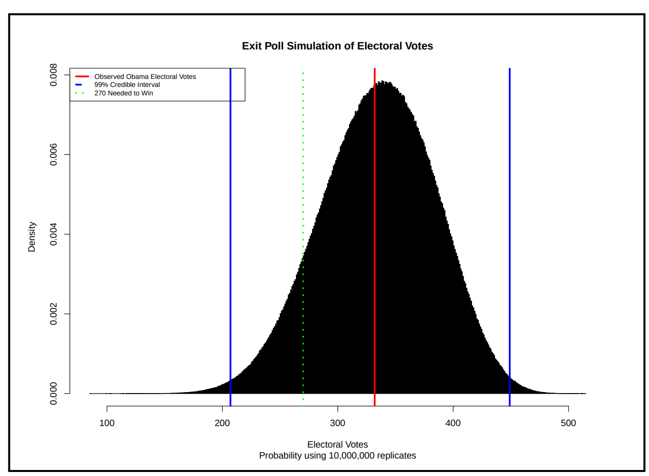

Figure 5. Posterior for Obama Electoral College Vote

[2] Estimates for Colorado, Oregon, and Washington are conducted using only absentee/early voter Random Digit Dial (RDD) phone surveys.

[3] Nebraska and Maine awards the winner two votes and then one additional vote for each congressional district.

ple data is not the most favorable strategy due to sampling variability. However, updating the probability that a candidate will win with additional known data in each of the given states will decrease the variability in the posterior distribution. This can be accomplished by using additional known prior data or, as is often the case in elections, by adding the final precinct election results provided shortly after the polling places close. Due to the nature of elections, informed priors are often available and can be incorporated into the estimates to improve the probability distribution. In this way, specific models can be developed to handle states with more or less available prior data and improve the overall model.

3.4 Statewide Bayesian Regression

The idea of taking a distribution and producing an estimate can be taken one step further. Here the data will be illustrated using a common regression model from a Bayesian perspective. In this way we can address the question on how the quantity, y, varies as a function of a vector of quantities, x. We can take the currently collected data and model the results using other quantities that are available. In some ways, due to the nature of linear regression, prior information is already implicitly included in exit poll regression models.

Table 1. Coefficients for Traditional Linear Regression

Name			
Coefficients	Est.	Std. Error	$\Pr(>\|t\|)$
2008 Dem %	0.7308	0.06337	≈ 0
2012 Dem Exit Poll %	0.3145	0.06342	≈ 0

It is quite clear that the past Democrat vote from 2008 and the current exit poll vote from 2012 are very good predictors of the 2012 final precinct reported vote. Furthermore, using the classical linear regression, the R^2 value is 0.95 indicating that a significant amount of variation in vote is explained by these two predictor variables. Using the Florida data, a Bayesian regression model can be fit to the data using the classical regression model from Table 1.

There are two primary goals that are addressed by regression models in this paper: 1) general understanding of the data within a given state. In other words identifying variables that aid in a linear prediction of the candidate's vote; and 2) predicting y, given x, for future observations.

For the purposes of this paper the sample of polling locations using the final end of night results are used as the response variable. Generally for all states past data tends to be a very good predictor of current results. In some states there are other predictors (e.g. precinct boundary changes, current voter registration, weather, etc.) that work well while in other states those same predictors provide no additional information and make the model unnecessarily complex. Therefore models for each state should be addressed separately.

This concept can be extended upon and hierarchical linear models can be fit to the data. This is an area for further research with respect to exit polling. Bayesian hierarchical models are beyond the scope of this dicussion and will not be covered in this paper.

The example in Table 1 uses both 2008 and 2012 precinct data from the Florida presidential race. For the purpose of this paper a non-informative prior is applied to the model. The focus is on the techniques and the idea of Bayesian regression rather than the choice of prior. The coefficients, the standard error, and the p-value from the classical regression model are shown in Table 1. To take a Bayesian approach a sample is simulated from the joint posterior distribution of the coefficients β and the error standard deviation σ. The regression vector β is simulated from the multivariate normal density with mean $\hat{\beta}$ and covariance $V_\beta \sigma^2$. Continuing to use Florida as an example we can produce the coefficient distribution along with the 99% *credible interval* seen in Figure 6.

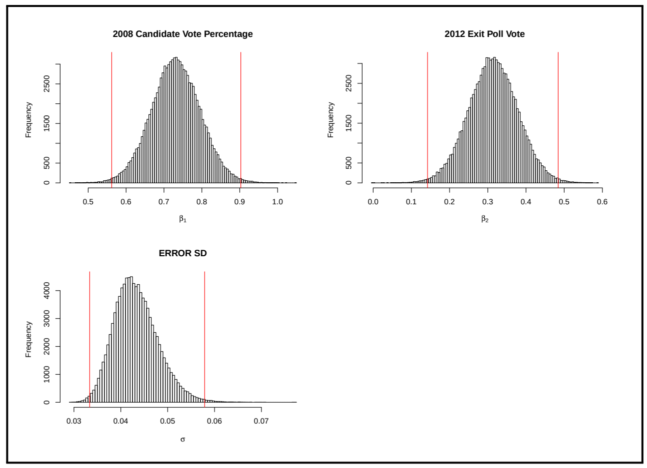

Figure 6. Bayesian Regression Coefficients – Florida

The error standard deviation is simulated in the same way. To simulate a draw from (σ^2, β), σ^2 is drawn from the inverse $Inv-Gamma\left((n-k)/2, S/2\right)$ density. The inverse-gamma distribution is used as the marginal posterior distribution for the unknown variance, σ^2, using a non-informative prior.

Again, the regression model presented here is an example model used for demonstration purposes (i.e. no formal model selection procedure was used). Furthermore, for this same purpose the non-informative prior is used. It's clear from the output of the regression summary that there is a strong effect for 2008 candidate vote percentage, precincts with high Democrat vote in 2008 tend to have a very predictable Democrat vote in 2012. As one would expect the 2012 exit poll results have a strong effect when predicting the final polling location results. This example regression model for Florida is provided in Equation 2.

$$E\left(CAND_j|x,\theta\right) = \beta_0 + \beta_1 \cdot CANDEP2012_j + \beta_2 \cdot CAND2008_j \tag{2}$$

We can check to see if the observed data from the polling places are consistent with the fitted model. In Figure 7 the approach is based on the use of the posterior predictive distribution. This figure shows the interval bands for the polling places as well as the precinct's reported vote. With this approach one can look to see if the observed response values are consistent with the corresponding predictive distributions. Based on the model and the predictive distribution, the model fits quite well without outliers in any of the precincts.

4. Conclusions

Several important conclusions about the analysis of exit poll data can be drawn from this review of approaches using probability distributions. First, it is clear that there are many probability distribution components to an exit poll. Classical statistical approaches using the normal distribution are often used but other distributions, both categorical and continuous, can be used as well. These approaches provide additional insight into the data as well as providing a way to include prior information. This research on exit polling serves as an exploration of ways to investigate and analyze data and to provide alternate, complementary approaches that may be more fully integrated into standard election (and non-election) exit polling.

These procedures are only a few of the many ways that can be used to analyze exit poll data. These approaches provide an alternate way to summarize and report on these data. It also provides additional visualization and ways to view the data and how the data are distributed. Probability distributions can be graphed in this way and, given the data, one can easily observe the probability of many outcomes.

This paper discusses several probabilistic concepts in an election exit poll framework: models using the Dirichlet/Beta distributions, Dirichlet process to facilitate clustering of polling locations, multinomial models to establish a posterior probability distribution of electoral votes, and statewide regression modeling using the normal distribution. There are many other Bayesian approaches that exist that should be further explored. Further topics include small sample sizes, missing data, censored data, and a deeper investigation into absentee/early voting. Additionally, these approaches can be used to investigate various complex sample design techniques (e.g. stratified, cluster, multi-phase, etc.) and evaluate how the designs interact with probabilistic approaches in an exit polling context. Further hierarchical modeling may provide additional insight into the complexities of the exit poll data.

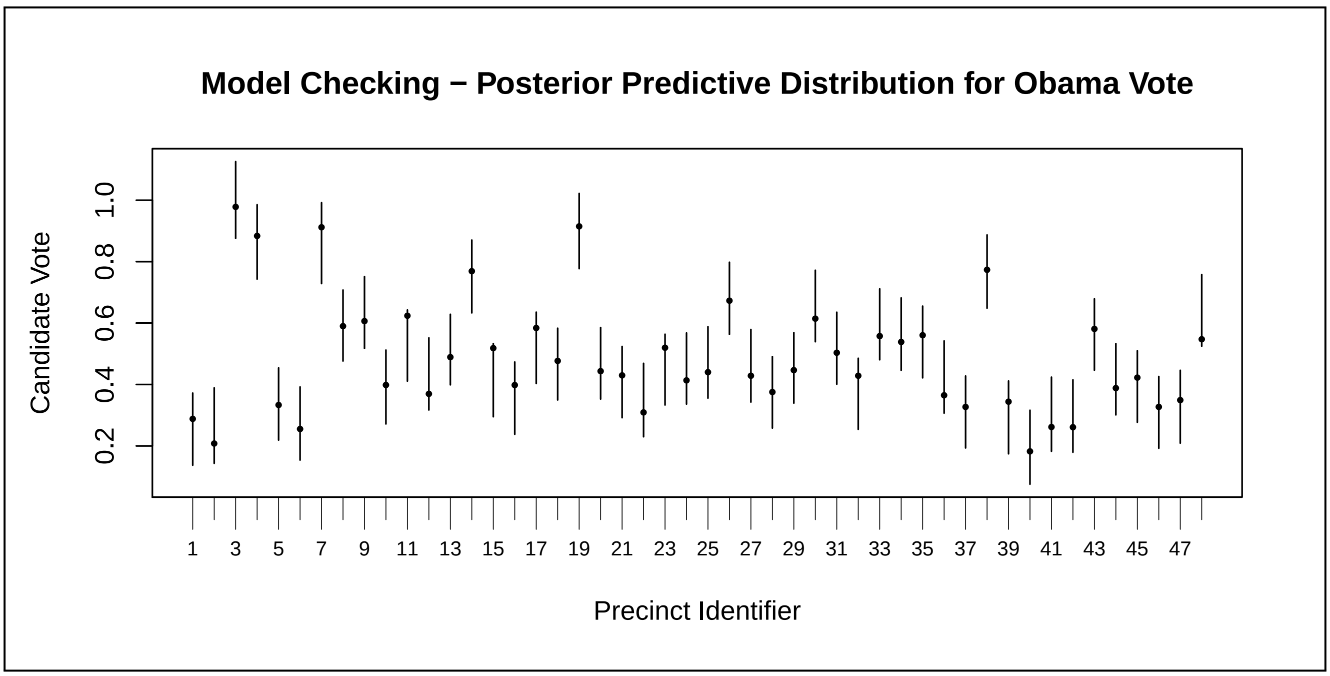

Figure 7. Outlier Detection Using the Posterior Predictive Distribution