Bayesian Inference in Binomial Logistic Regression: A Case Study of the 2002 Taipei Mayoral Election*

Chia-hung Tsai

Assistant Research Fellow,
Election Study Center, National Chengchi University

ABSTRACT

Bayesian statistics assumes that there are specific parameteric distributions for the unknown parameters. It fits the probability model of interest by incorporating prior information regarding the unknown parameters and the likelihood function of the observed data. Moreover, Bayesian statistics as well as non-Bayesian methods produce good asymptotic results. Using WinBUGS and R language, a binomial logistic regression model of voting choice in the 2002 Taipei mayoral election is developed. Adding the prior information from the first panel to the estimation of the second panel, the Bayesian model yields sharper estimates concerning the election outcome. Additionally, the replication of data provides a model check and the baseline of the new observations. The methodological contribution of this paper is the ability to fit a binary logistic regression model with the observed data using the Bayesian inference.

Key Words: Bayesian statistics, prior information, predictive posterior simulation, binomial logistic regression, voting choice

1. Election Forecast and Modeling

It is fair to argue that political scientists contribute a great deal of time to the prediction of election results. Election forecast has been done inductively with aggregate-level data and linear or non-linear model (e.g. Kramer, 1971; Tufte, 1978; Rosenstone, 1983; Abramowitz, 1988; Lewis-Beck, 1992; Tsai, 2000). This type of research views each election outcome as a case and assumes that a linear or non-linear model can generate the prediction while holding the values of the observed variables at their means. Although the prediction of the next election outcome is done in an *ad hoc* manner, there is no theory supporting such kind of prediction.

In Taiwan, individual-level and aggregate-level data are used to generate a linear or non-linear model, classifying cases and comparing the forecast results with the real ones (Liu, 2001). Hung (1994) classified the subunits of districts by aggregate-level data, which allows him to ascertain the intention of voters who answer "undecided" or refuse to reveal their intention. I-Chou Liu (1996a) proposed a prediction model based on Kelly and Mirer's (1974) "simple act" decision-making pattern and found that an individual's preferences over candidates are useful for making pre-election predictions. Nien-hsia Liu (1996) employed a CHAID model to classify respondents as twelve blocs, asserting that the scores of candidate images can successfully predict voting decisions. Shing-yuan Sheng (1998) argued that calculating the probability of each case to vote for a particular candidate is a method better suited to forecasting elections. She estimated the parameters of a voting model and used them to generate the predicted probability of voting for each particular candidate. By doing so, she obtained the aggregate-level vote share comparable to the actual results. Emile Sheng (2000) claimed that the measurement of "feeling thermometer" can tap into people's evaluation of candidates better than demographic and political attitude variables, thus he used it to classify the undecided voters. His model generates the prediction that closely approximates the actual results. Chung (2000) instead took advantage of Hung's (1994) model while considering the situation of strategic voting. However, his model appears to generate very different results from the actual election outcome.

Both approaches have come up with outstanding results, yet there are some inherent shortcomings of frequentist statistics. First, they merely consider the information given by the sample and generate the predictions that

fit their model or observed data the best. The association among variables in time t is used to derive the observations in time $t+1$, assuming that the association remains unchanged and no more variables needed to be included, i.e. there is no specification problem. In other words, they take the present observations into account, instead of using information from one sample as prior knowledge for a subsequent sample.

Secondly, it is a widely utilized assumption that there is an unknown parameter that can be estimated by a sample that represents the population. However, such an assumption fails to consider the uncertainty concerning the estimates. Instead, it merely considers the level of confidence for the estimates. The election forecast made by the ML estimators as matter of fact is related to a wide range of conditions (Zellner, 1971). Bayesian statistics can make probability statements about parameters' values, which lets us draw the predictive posterior distribution.

Lately, Bayesian analysis has been applied to political science. Western and Jackman (1994) pooled two data sets for a regression analysis to obtain posterior distributions for regression coefficients. They argued that a subjective probability of Bayesian inference is an excellent alternative to conventional inference that relies on a non-repeatable data collection. Quinn, Martin, and Whitford (1999) used the Bayesian setup of a multi-nomial probit model to estimate the factors of voter choice in the Netherlands. In addition to the field of voting behavior, Bayesian statistics also is prevalent in the study of congressional behavior. Martin (2001) presents how members of Congress react to the preference of the president and the Supreme Court as well as the other legislative body with a hierarchical probit model. He adopts the Bayesian approach because there is "behavioral heterogeneity" —the unit of analysis may behave differently with the context. That technique is useful in dealing with comparative research, which takes the heterogeneity of context into account (Western, 1998). Last, Clinton, Jackman and Rivers (2004) develop a Bayesian procedure and analyze roll calls from the 106th U.S. House of Representatives and Senate.

Essentially, every prediction involves the pattern derived from the data collected in the previous time point. The information obtained in a survey or election statistics allows us to project the possible characteristics of information of concern to the next time point. More formally, information from one sample is used as prior knowledge as another sample. For instance, we can predict the probability of an event under certain conditions by using the prior information regarding the event. One of the merits of Bayesian statis-

tics is that prior knowledge about parameters of interest is incorporated into the model with the appropriate likelihood function of observed data in a mathematical way. That not only makes the forecast more formal, but also generates better parameters of the model.

The 2002 Taipei mayoral election, in which the Kuomingtang (KMT) and Democratic Progressive Party (DPP) candidates competed against each other, provides us with a good opportunity to explain voter choice in a two-candidate campaign. To take advantage of prior information and Markov Chain Monte Carlo (MCMC)[1] estimation, this paper will use WinBUGS and R language to generate the estimates for parameter quantiles.

Section 2 is a brief introduction to the setup of Bayesian statistics. The posterior density distribution for binary logistic regression will be derived. Section 3 summarizes the data and the variables. Section 4 presents the logistic coefficients and the predictive posterior simulation. Section 5 is the conclusion.

2. Bayesian Analysis and the Bayesian Binomial Logistic Model

On of the easiest ways of stating Bayes' theorem is that the conditional probability of a sequence of events approximates the product of the probability of events and the probability of the given event under the condition of the events of interest. Suppose that there is a sequence of events A_m. The probability that the sequence of events A_m is true is denoted by $P(A_m)$. $P(A_m|H)$ means the probability that the statement A is true when the statement H is true, which is also called conditional probability. According to Bayes' theorem,

$$P(A_m|H) = [P(H|A_m) \cdot P(A_m)]/P(H) \tag{1}$$

The left-handed side of the equation is the conditional probability of event A given event H. The right-handed side of the equation is the probability of event H given event A divided by the probability of event H. In other words, the conditional probability of event A equals the fraction that the event H given event A is true. With this relationship between the two state-

---

1 MCMC simulation generates numerical representations of probability distribution functions for model parameters, calculating the mean, variance, and other moments for the parameters.

ments, we can obtain the conditional probability A given the hypothesis of H even if A is not observed as long as we have the information regarding A and H.

Regarding H, it can be conceived as the observed data or characteristics of a given sample. We replace H with y, which is observed data and dependent upon other observed variables as follows.

$$(\boldsymbol{y}|\boldsymbol{X})=\boldsymbol{X}\beta+\varepsilon \tag{2}$$

where y is a $m\times 1$ vector of observations and $y=0$ or 1. For the right-handed side of the equation, $\boldsymbol{X}$ is a $m\times k$ matrix of fixed elements, β is a $k\times 1$ vector of coefficients, and ε is a $m\times 1$ vector of random disturbances. In this study, the disturbances are assumed to have a standard logistic distribution with mean 0 and variance $\pi^2/3$. The dichotomous response 0 and 1 can be transformed to the probability $P(y=1)$ and $P(y=0)$ and the coefficient vector β represents the change of the probability when the fixed elements change. When the error terms have a logistic distribution,

$$P(y=1|\boldsymbol{X})=1/[1+exp(\boldsymbol{X}\beta)]. \tag{3}$$

Equation (3) represents the observed data and the coefficients to be estimated. For such a non-linear model, maximum likelihood estimation can be used to obtain the estimators that make the observed data the most likely. The estimates are unbiased, consistent and efficient in a large sample (Long, 1997). The likelihood function is

$$l(\beta, \sigma|\boldsymbol{X}, \boldsymbol{y})=\Pi_{i=1\sim n}\,[P_i]^{yi}\,[1-P_i]^{1-yi} \tag{4}$$

The most important properties of the logistic regression model are the parameters—β and σ. With the observed data $[\boldsymbol{X}, \boldsymbol{y}]$, the likelihood function can be maximized relative to the unknown parameter vector β and σ.

The whole estimation process of the parameter only employs the information of the given data, without using anything regarding the unknown parameter to be estimated, which is the gap between frequentist and Bayesian statistics.

According to Bayes' theorem, we can update our knowledge regarding the distribution of unknown parameter if its prior information is known. Suppose that we know that a football quarterback's prior interception rate over the past 10 seasons follows the normal distribution, for instance, then we can use this information to estimate his interception rate in this month given his current physical strength, teammates and playbook. In other

words, we can estimate the *posterior distribution* as follows.

$$p(\theta|X, y) = [l\,(\beta, \sigma|X, y) \cdot p(\theta)]/p(y) \tag{5}$$

Since $p(y)$ is a constant relative to θ, the posterior distribution is proportional to the product of the likelihood function and the prior distribution. Therefore, (5) can be rewritten as (6).

$$p(\theta|X, y) \propto p(\theta) \cdot l(\beta, \sigma|X, y) \tag{6}$$

The prior information concerning the unknown parameter, $p(\theta)$, is critical to the estimation procedure. It is the prior to observing X in the current time point and based on earlier experience or beliefs. We can assume it has a constant or a special distribution, such as normal, Beta, or gamma. The normal distribution will be used for the following analysis.

Bayesian statistics have been proved to generate asymptotically consistent and efficient estimates (Zellner, 1999). Due to the choice of the density distribution of the unknown parameter, however, the results may not be uniform. As the number of variables increases, we must assume more prior densities regarding those parameters. Jeffrey (1967) suggested that the prior distributions of β can be constant and the variance is $1/\sigma$. To incorporate more information regarding the β parameters, however, we can assume that the cluster of β follow multivariate distribution and the variance follows the inverse-Gamma distribution, which is called conjugate prior. A proper conjugate priors guarantee that a posterior distribution possesses the same property of the prior distribution.

Considering the normal distribution, Equation (6), the likelihood function can be written as follows.

$$
\begin{aligned}
l(\beta, \sigma|X, y) &= (2\pi\sigma^2)^{-n/2}\ exp\,[(-1/2\sigma^2)\,(y - X\beta)'\,(y - X\beta)] \\
&= (2\pi\sigma^2)^{-n/2}\ exp\,[(-1/2\sigma^2)\,\textstyle\sum(y_i - x_i\,\beta_i)^2] \\
&\propto (\sigma)^{-n}\ exp\,[(-1/2\sigma^2)\,\textstyle\sum(y_i - x_i\,\beta_i)^2]
\end{aligned}
\tag{7}
$$

It is assumed that $p(\theta)$ follows the normal distribution, $p(\theta) \sim N(\mu, \sigma^2)$ and it is independent and identically distributed (i.i.d.). It is uniform on $(\mu, \log\sigma)$ and proportional to $(\sigma^2)^{-1}$ (Gelman, Carlin, Stern, and Rubin, 2004). In this case, a posterior distribution with uninformative priors is specified as follows.

$$p(\theta|X, y) \propto (\sigma)^{-n-2}\ exp[(-1/2\sigma^2)\,\textstyle\sum(y_i - x_i\,\beta_i)^2] \tag{8}$$

Equation (8) incorporates the prior information, $p(\theta)$, and the likelihood

function, which can be evaluated by the method of maximum likelihood estimation. According to the computation formula of maximum likelihood estimation, the marginal posterior distribution of β is normal with mean B.hat and variance σ^2. Regarding the variance, the inverse gamma distribution can be used to obtain the posterior joint distribution, which is the conjugate prior (Gill, 2002).

Bayesian statistics seek the posterior probability about θ, and θ is conditional on the observed data. Therefore, the inference of θ should be characterized by the joint posterior density of model parameters. MCMC methods are useful at sampling values from the posterior distribution given the observed data. Jackman (2000a) has shown the applications of MCMC methods to binary probit model, linear regressions with AR (1) disturbances, and multinomial probit model. With millions of samples generated by MCMC methods, it seems that the assumption of normal density used by the maximum-likelihood-estimation (MLE) does not really hold. Instead, the distribution of the parameters is skewed to one side or the other.

Regarding the difference between Bayesian and classical statistics, Gill (2002), Lavine (1999), Jackman (2000a), and Gelman et al. (2004) have written at length. According to Jackman (2000a) and Lavine (1999), the fundamental distinction between the two approaches lies in the way θ is treated. The Bayesian statistics treats θ as though it were a random variable, and we make posterior probability statements about it with the sample information. However, frequentists treat θ as a fixed constant and assume one can estimate it by sampling repeatedly and constructing an appropriate confidence interval. To assess the plausibility of the estimated θ, one can also use MLE to find the local maximum of the likelihood function.

Nevertheless, only the Bayesian statistics emphasize the probability of θ. In doing so, all of the sample information can be included so as to make inferences about θ. With his example of the cancer rate, Lavine (1999) points out that once the number of events in which we are interested is observed, the frequentist approach utilizes only the likelihood function and the number of events to assess how well the parameter explains the data. The Bayesian statistics however take other possible numbers of events into account. By calculating the conditional probability, one can obtain the posterior distribution and a more constrained estimate of θ.

Thanks to the R project and WinBUGS,[2] the computational processing can be done with a few commands. In addition to the parameter estimates, standard deviation, and the credible interval, WinBUGS generates the posterior predictive simulation, which allows us to tell if the replicated data conditional on the parameters fits the original data. In doing so, we can claim that the parameters can be used to predict the future observations

3. Explaining the 2002 Taipei Mayoral Election

The 2002 Taipei mayoral election was held on Dec. 6, 2002. Nine weeks prior to Election Day, the Election Study Center of National Chengchi University began collecting rolling samples and called all of the respondents again twice.[3] In the two waves, re-interviewed respondents were asked their voting intentions, candidate evaluations, party evaluations, and partisanship. Due to the advantage of Bayes' theorem, prior information regarding the whole data is included in the estimate of the parameter. Therefore, each wave of cross-section data can be used to return the prior information for the next wave of data, if each wave is an independent sample and encodes enough information about the population. In this sense, this election data is chosen to test the forecasting ability of Bayesian approach.

To explain the probability of voting choice in a single-member district election, scholars investigated candidate characteristics. Fu (1998) suggested that one of the most important factors in the 1994 Taipei mayoral election was the candidate characteristics. Hawang (1996) also found that in the 1996 presidential election candidate images and ability is much more important than other political attitudes. Issue voting in the presidential election has been examined as well. Examining the distance between the respondents' issues position and political parties', Hsieh, Niou, and Lin (1995) found that the relative position of the respondents and political parties determines the evaluation of parties, which consequently influences party choice. Wang (2001) asserted that policy ideology does play a role in Taiwan. Jiang (2002) reported that in the 1996 and 2000 presidential elections partisanship as well as issue positions of the respondent had an impact on voting behavior. Chen (2004) used multinomial probit model to estimate the effect of issues as

---

2 For more information of the two software packages, see www.r-project.org for R language and http://www.mrc-bsu.cam.ac.uk/bugs for WinBUGS.

3 This project is sponsored by the National Science Council (NSC 92-2414-H-004-020).

choice-specific variables, and found that the issue of independence and the issue of reform affect the probability of voting choice. Last, Shyu (1998) constructed the "Lee Teng-hui complex," showing that it shaped people's voting pattern across the national elections in the 1990s.

Partisanship, which is measured by the question of which party the respondent likes, has a great impact on voting choices in the legislative elections (Liu, 1996b; Shyu, 1991, You, 1996). Chen (1998) found that partisanship and candidate evaluations determined voter choice in the 1996 presidential election. Party attachment to the KMT, DPP, People First Party (PFP) and Taiwanese Solidarity Union (TSU) is coded as four dummy variables, which taps into the multi-party system after the 2000 presidential election.

To assess the importance of candidate evaluations between Ma and Lee, a -10 to 10 scale is included. To measure candidate evaluations in this survey, the respondents were asked how much they like the two candidates on a scale from 0 to 10. The difference in scores between the two candidates is used as the indicator of candidate evaluations.

"Ethnicity"—Hakka, Minnan, mainlander, and aboriginal—is an important social cleavage in Taiwan.[4] It has been found that most party supporters of the DPP are Minnanese, and that mainlanders strongly support the KMT (Lin, 1989). The incumbent mayor's father was born in Mainland China, but the challenger is from a Minnan and Hakka family. The different ethnic background of the two candidates linked the campaign to this social cleavage. For the convenience of analysis, being a mainlander is coded one, and zero otherwise.

Although the issue of independence and other issues have been proved to have an impact on the voting decision, the respondents in this survey were not questioned about their perceptions of political parties. Moreover, the purpose of the following analysis is aimed to simulate the sample and deduce posterior distribution from it, thus only partisanship, ethnicity, and candidate assessment are included in the model.

Notice that the "non-response" answers are treated as missing values. Missing data is a serious problem for the survey data. One of many examples is to include abstention in the variable of voting choice (Lacy and Burden, 1999). In WinBUGS, missing data can be imputed as "NA" in the dataset. Jackman (2000b) has shown that we can treat the dependent vari-

---

4 The ancestors of Minnanese and Hakka emigrated from mainland China in the 15th century. Mainlanders instead came to Taiwan after 1949.

able in the probit and logit model as missing or partially observed, thus it can be analyzed as an ordinary regression model. He also shows we can deal with item-nonresponse with the Bayesain approach. By imputing the values of the missing data, however, the conditional distribution of the missing data has to be considered (Gelman et al., 2004).[5] In other words, it is necessary to assume the distribution of the missing data, which may complicate my investigation of voting behavior. Therefore, I reserve the inclusion of missing data for another paper.

A set of covariates is used to estimate the Bayesian binary logistic regression model, and the result is summarized in Table 1. The mean and standard deviation of each parameter is estimated with the 2.5% and 97.5% credible intervals, which show the region in which 95% of all parameter estimates will be drawn. In other words, we can view it as a region of acceptance of the null hypothesis that the parameter is equal to zero. If the region does not contain zero, we are confident about the effect of the variable.

Table 1 demonstrates that partisan attachment to the KMT, DPP, PFP, TSU differs significantly from zero. Being a KMT or PFP supporter would increase the probability of voting for Ma. On the contrary, being a DPP partisan would decrease this probability. The TSU partisanship also has a negative impact on voting for Ma. Being a Mainland Chinese does not influence the probability of voting, which is partly because of the intervening effect of partisanship. The relative difference between the two candidates has an impact on voter choice. The higher the relative difference, the more likely that voting for Ma is. Overall, these results show that the Taipei voters use candidate evaluations and partisanship as their voting guides, and that the ethnicity cleavage did not weigh in people's choices in this election. Appendix A shows the procedure of iteration of two chains, each of which contains 2000 times and the first 1000 times are discarded. Given that the initial values drew from the normal distribution, the parameters β and τ will converge after thousand times of iterations.

By adding the information gained from wave 1, it is expected that the logistic estimates will be shaper. The mean and standard deviation of the logistic coefficients in the first model are set up as the parameters of the normal distributions that the coefficients of the second model are expected

---

5 In Jackman's (2000b) analysis, however, he did not actually impute values for the dependent variable; instead, he used Bayesian simulation to generate samples of the dependent variable from the posterior density of parameters.

Table 1 Posterior Distribution for Binary Logistic Regression Model, Wave 1 (Voting for Ma=1, Lee=0)

Parameters	Mean	S.D.	C.I. 2.5%	Median	C.I. 97.5%
Intercept	−6.965	0.137	−7.218	−6.964	−6.707
KMTPID	0.412	0.1552	0.1034	0.4156	0.7278
DPPPID	−1.16	0.2619	−1.691	−1.16	−0.6545
PFPPID	0.4032	0.1621	0.08768	0.4081	0.7321
TSUPID	−1.542	0.8241	−3.49	−1.44	−0.1869
Mainlander	0.05492	0.02098	0.01443	0.05503	0.09735
Candidate Evaluation	0.08961	0.1258	−0.1657	0.09269	0.3427
deviance	777.4	3.955	771.5	776.8	786.2

Note: The dependent variable is voter choice between Ma and Lee. Voting for Ma is coded as 1, 0 otherwise. "C.I." means credible level. The variance is equal to $(tau)^{-1/2}$.

to follow. In other words, the original uninformative distributions $N\sim(0, 1)$ are replaced by $N\sim(\mu, \sigma)$, where μ and σ are obtained from the estimates of the first model.

Table 2 presents the Bayesian output of the second model using the uninformative priors, and Table 3 shows the estimates with the informative priors. The credible interval shows that being a mainlander has an insignificant impact on voter choice. The effects of KMT, DPP, PFP, TSU partisan-

Table 2 Posterior Distribution for Binary Logistic Regression Model with Uninformative Priors, Wave 2 (Voting for Ma=1, Lee=0)

Parameters	Mean	S.D.	C.I. 2.5%	Median	C.I. 97.5%
Intercept	−7.463	.108	−7.669	−7.462	−7.25
KMTPID	.499	.122	.264	.501	.749
DPPPID	−1.043	.2167	−1.5	−1.041	−.634
PFPPID	.4387	.1319	.178	.441	.687
TSUPID	−2.871	1.249	−5.87	−2.637	−1.036
Mainlander	.074	.099	−.129	.074	.261
Candidate Evaluation	.051	.015	.020	.051	.083
variance (σ^2)	.004	.005	.016	.005	.001

Note: The dependent variable is voter choice between Ma and Lee. Voting for Ma is coded as 1, 0 otherwise. "C.I." means credible level. The variance is equal to $(tau)^{-1/2}$.

ship remain significant, and the relative evaluation is still a good predictor.

In Table 3, the mean and standard deviation of each parameter with informative priors are presented. They seem to be very similar to those in Table 2, but Figure 1 demonstrates that the TSU partisanship has a different distribution with the informative prior. Not only the standard deviation becomes smaller, but also the mean shifts to -2.152 from -2.871. For the rest of the parameters, the change of priors has very little influence on their distributions.

Table 3 Posterior Distribution for Binary Logistic Regression Model with Informative Priors, Wave 2 (Voting for Ma = 1, Lee = 0)

Parameters	Mean	S.D.	C.I. 2.5%	Median	C.I. 97.5%
Intercept	-7.459	.104	-7.666	-7.458	-7.24
KMTID	.499	.121	.260	.499	.738
DPPID	-1.058	.217	-1.494	-1.05	$-.665$
PFPID	.440	.127	.198	.437	.705
TSUID	-2.152	.675	-3.543	-2.128	$-.902$
Mainlander	.067	.099	$-.128$	.067	.266
Candidate Evaluation	.051	.015	.022	.051	.081
variance (σ^2)	.006	.012	.021	.007	.002

Note: The dependent variable is voter choice between Ma and Lee. Voting for Ma is coded as 1, 0 otherwise. "C.I." means credible level. The variance is equal to $(tau)^{-1/2}$.

The different results shown in Table 2 and Table 3 suggest that the prior information regarding the mean and standard deviation of the parameter indeed has an impact on the estimates. In the conventional logistic regression model, the variance of the likelihood function depends on the mean (Gill, 2002). In the Bayesian setup, however, the distribution of the prior information determines the variance function. The assumption of the variance influences the form of the estimates. Zellner and Rossi (1984) have shown the different estimates of β by using no prior information and informative prior information. The estimated β of the posterior density function is the function of the maximum likelihood estimates and the prior information. The findings above conforms to Zellner and Rossi's theory.

From the Bayesian point of view, one of the methods to check the posterior distribution is replicating the data from it (Gelman et al., 2004). Basically, simulated values from the posterior predictive distribution can be

Figure 1: Posterior Density of Coefficients with Different Priors, Wave 2
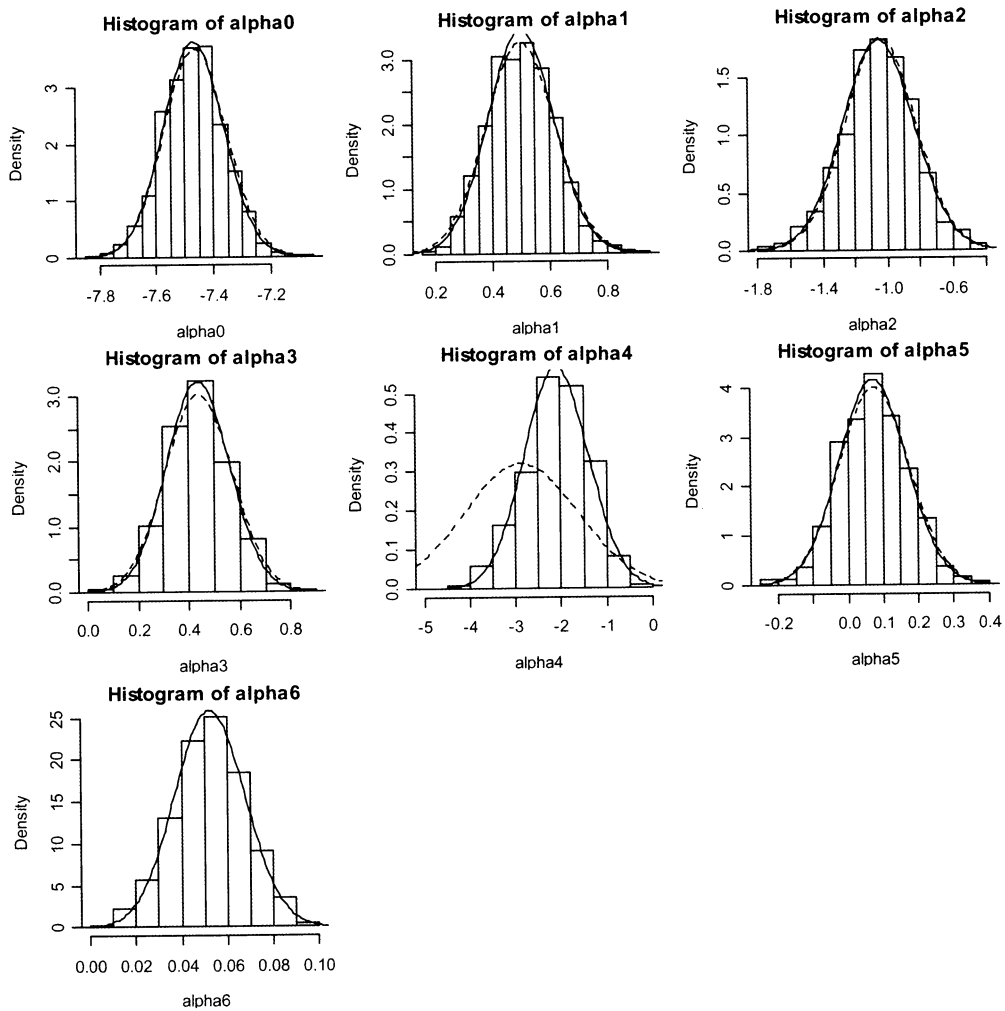
Note: In each histogram, the dotted line represents the normal curve of the distribution with the uninformative prior.

graphed on a histogram. In doing so, we are able to assess our model to see how well the posterior distribution matches the original data. Figure 2 and Figure 3 display the results of nineteen replications of data from the original data and parameters. All of them are transformed from a bimodal distribution to a normal distribution with 1000 draws from the posterior distribution. They demonstrate that posterior distribution of Y approaches the original data in the shape of normal distribution after certain simulations. Fig-
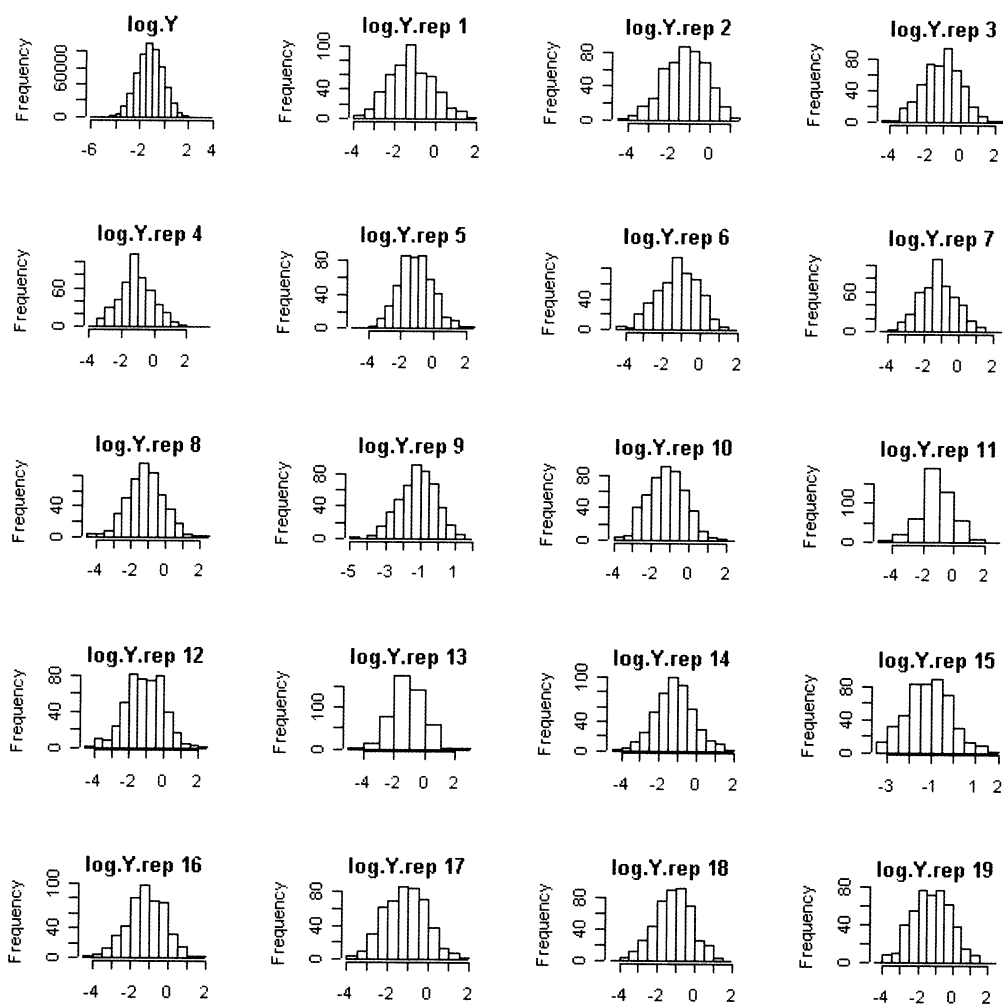
Figure 2: Nineteen Replications of Data, Wave 1

ure 2 displays nineteen simulations. Among them, the fourteenth sub-figure actually has a distribution very similar to the original data. In other words, we can replicate our data drawn from our model and the result approximates the original data. Regarding wave 2, the distribution in the nineteenth sub-figure is very similar to that of the original data (see Figure 3). Using posterior predictive simulations, it is shown that the posterior distribution, which takes prior information and the likelihood function into account, is very similar to the original data.

Figure 3: Nineteen Replications of Data with Informative Priors, Wave 2
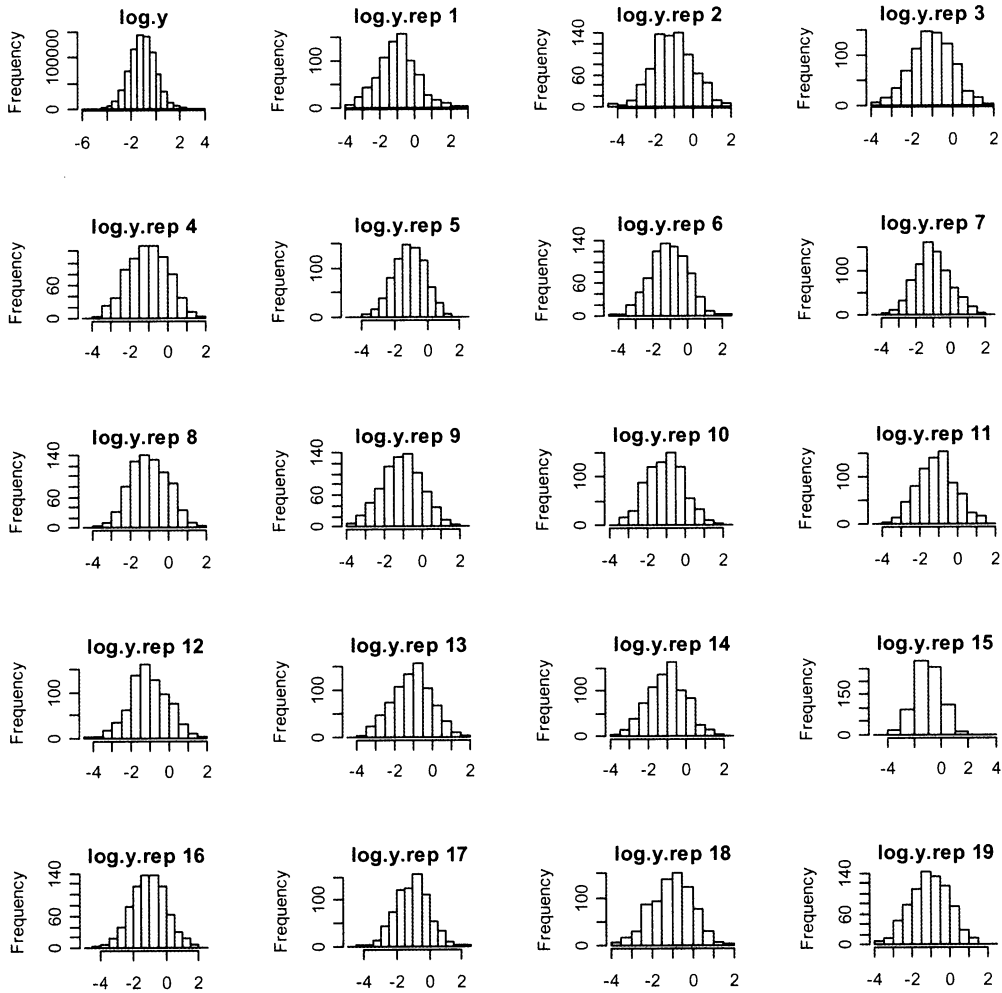
4. Election Forecast

Bayesian statistics allows us to incorporate the prior information and to make inferences about the future observations. With uninformative and informative priors, the predicted value of the probability of voting choice can be obtained from the linear model of the parameters. In Figure 4, a histogram with the normal distribution with two sets of priors is graphed. It is shown that most of the predicted Y centered around 0.1, which represents

0. 525 as the probability of voting for Ma.[6] The distribution of Y with the informative prior has a slightly higher proportion of the total value around the mean 0.1, which implies that the prior information increases the centrality of the predicted values. With more information regarding the parameters specified in the model, the simulated Y should approach the future observations, if everything else is held constant.

Although the election forecast generated by the mean of the independent variables and the parameters is fewer than the actual result of the election, which is 64.10%, it demonstrates that more prior information helps the estimation of the dependent variable. As matter of fact, the actual probability, 0.64, falls in the region of 95%. With the distribution that contains most of the predicted values conditional on the parameters, election forecast turns out to be the mean of a group of point estimates plus a given level of standard deviation. Adding the latest data to the model and setting up the level of accepted standard deviation, we should be more confident about the "prior" estimate of the future observation.

5. Conclusions

With only a little previous research, this paper aims to apply Bayes' theorem to the field of political behavior, which is preoccupied with the frequentist statistics. The essential property of Bayesian statistics can be summarized as two points. First, a probability model includes prior knowledge about the parameters if available. Secondly, the value of the unknown parameters is conditioned on the probability model and observed data.

The Bayesian inferences for parameters of the binary logistic model are presented in this paper. The substantial finding of the model itself is that the KMT, DPP, PFP and TSU partisanship has a significant impact on the probability of voting choice. Candidate evaluations also affect whom the Taipei voters cast their ballot for. Ethnic background however is not a good predictor.

The Bayesian model sets up the distribution of the variance and other parameters as the prior information. The informative prior information from the first wave of the survey affects one of the estimates of the second

---

6 The predicted value of the probability is calculated from the formula $p = exp(z)/1 + exp(z)$, where z denotes the sum of the multiplication of the estimates with the mean of the independent variables.

Figure 4: Predictive Density of Data with Uninformative and Informative Priors, Wave 2
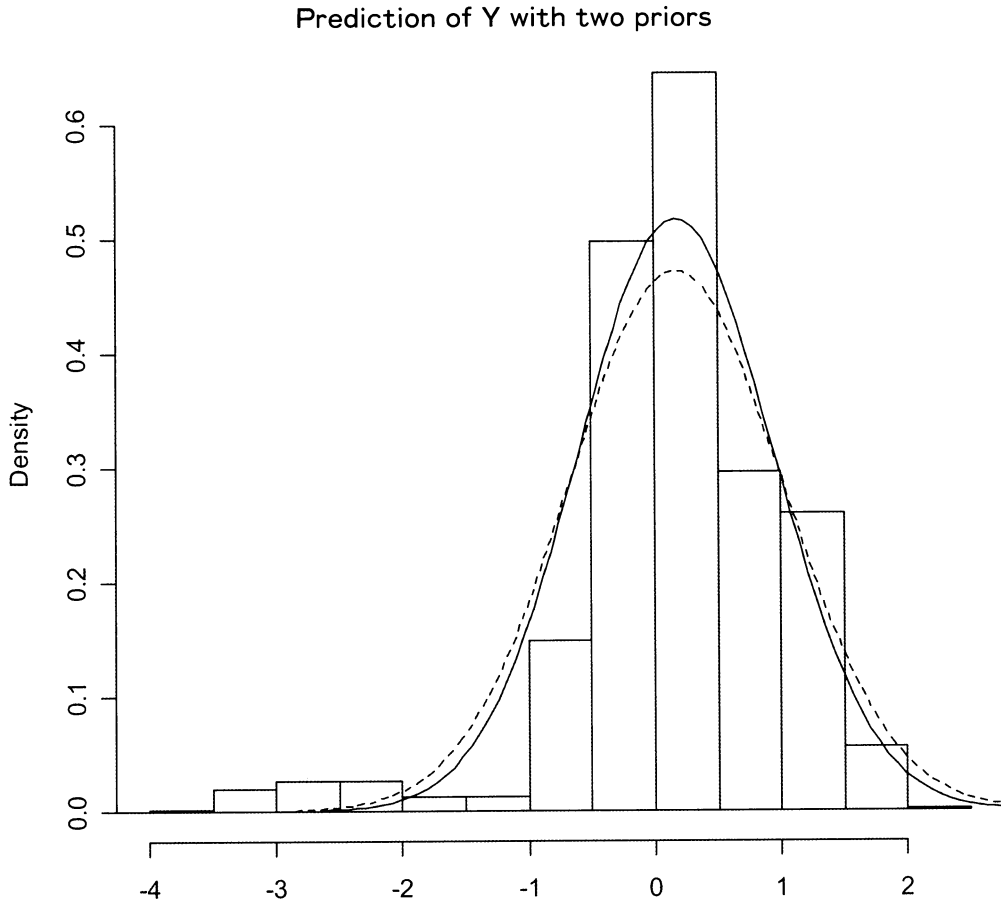
Prediction of Y with two priors

Note: In each histogram, the dotted line represents the normal curve of the distribution with the uninformative prior. The x-axis is the value of Y.

wave. Replication of observations drawn from the posterior distribution shows a distribution of data similar to the original one in both waves. This research also simulates the predicted probability of voting for Ma. The forecast generated by the mean of a group of point estimates plus a given level of standard deviation is not similar to the election outcome, but it is expected that the computation of new data with the existing parameters and information can come up with a better election forecast.

As for election forecast, the Bayesian analysis is useful in terms of the prior information regarding the distribution of political attitude and other

characteristics of population. With more computational knowledge, the idea of posterior distribution can be helpful investigating the unobserved data and missing values, which is often seen in the study of political behavior. The idea of the subjective probability embedded in the prior information certainly deserves more attention.

References

Abramowitz, Alan I.
　　1988　"An Improved Model for Predicting Presidential Election Outcomes," *PS: Political Science & Politics* 21:843-847.
Chen, Lu-huei
　　1998　"Presidential Voting of 1996 in Taiwan," *Journal of Electoral Studies* 5(2):161-184. (in Chinese)
Chen, Tse-hsin
　　2004　"Vote Choice in the Multiparty Election: An Example of Taiwan 2000 Presidential Election," manuscript.
Chung, Wen-chung
　　2000　"Election Forecast and Strategic Voting: A Case Study of 2000 Presidential Election," *Theory and Policy* 14(2):55-92. (in Chinese)
Clinton, Joshua, Simon Jackman and Douglas Rivers
　　2004　"The Statistical Analysis of Roll Call Data," *American Political Science Review* 98(2):355-370.
Fu, Ming-ying
　　1998　"Candidate Evaluation and Voting Decision: The Case of Taipei Mayoral Election in 1994," *Taiwan Political Science Review*, 3:195-243. (in Chinese)
Gelman, Andrew, John B. Carlin, Hal S. Stern and Donald B. Rubin
　　2004　*Bayesian Data Analysis*. London: Chapman & Hall.
Gill, Jeff
　　2002　*Bayesian Methods: A Social and Behavioral Sciences Approach*. Boca Raton, Fla.: Chapman & Hall/CRC.
Hawang, Shiow-duan
　　1996　"The Importance of Candidate Images and Capabilities in Presidential Elections," *Journal of Electoral Studies* 3(1):103-135. (in Chinese)
Hsieh, John fuh-sheng, Emerson M. S. Niou and Huei-ping Lin
　　1995　"Issue Voting in the 1994 Gubernatorial and Mayoral Elections: An Application of Rational Choice Approach," *Journal of Electoral Studies* 2(1):77-94.
Hung, Yung-tai
　　1994　"Election Prediction: An Aggregate Data Assisted Model," *Journal of Electoral Studies* 1(1):93-110. (in Chinese)
Jackman, Simon
　　2000a "Estimation and Inference via Bayesian Simulation: An Introduction to Mar-

kov Chain Monte Carlo," *American Journal of Political Science* 44(2):369-398.

2000b "Estimation and Inference are 'Missing Data' Problems: Unifying Social Science Statistics via Bayesian Simulation. This paper appeared in Political Analysis," *Political Analysis* 8(4):307-322.

Jeffery, Harold

1967   *Theory of Probability*, 3rd rev. ed. London: Oxford University Press.

Jiang, Chao-kuo

2002   "Taiwan ming chung yen chung cheng tang li ch'ang chih pien ch'ien," (Change of Party Positions from the View of Taiwanese). 2001 TEDS Conference, Taipei.

Kelley, Stanley, Jr. and Thad W. Mirer

1974   "The Simple Act of Voting," *American Political Science Review* 68:572-591.

Kramer, Gerald

1971   "Short-Term Fluctuations in U.S. Voting Behavior, 1896-1964," *American Political Science Review* 65:131-143.

Lacy, Dean and Barry Burden

1999   "The Vote-Stealing and Turnout Effects of Ross Perot in the 1992 U.S. Presidential Election," *American Journal of Political Science* 43(1):233-255.

Lavine, Michael

1999   "What Is Bayesian Statistics and Why Everything Else Is Wrong," *The Journal of Undergraduate Mathematics and Its Applications* 20:165-174.

Lewis-Beck, Michael S. and Tom W. Rice

1992   *Forecasting Elections*. Washington, D.C.: CQ Press.

Lin, Chia-lung

1989   "Opposition Movement in Taiwan under an Authoritarian-Clientelist Regime: A Political Explanation for the Social Base of the DPP," *Taiwan: A Radical Quarterly in Social Studies*: 117-143. (in Chinese)

Liu, I-chou

1996a "Forecasting Elections: Tests of Some 'Simple' Models," *Journal of Electoral Studies* 3(2):107-130. (in Chinese)

1996b "The Behavior of Taiwanese Voters in 1992: Consolidation of Partisan Ties," In *Taiwan's Electoral Politics and Democratic Transition: Riding the Third Wave*. Hung-mao Tien (ed.) N.Y.: M.E. Sharpe.

Liu, Nien-hsia

1996   "The 1996 Presidential Election Prediction: A Test of the Voting Behavior Model of No-Response Voters in the Poll," *Journal of Electoral Studies* 3(2): 131-156. (in Chinese)

2001   "The Knowledge Context of the Election Forecast Study in Taiwan: A Paradigm of Discipline Production," *NPF Research Report*, October issue. (in Chinese)

Liu, W. C.

1995   "The Genetic Forecasting Model for the Mayor Election of Taipei City," *Journal of Electoral Studies* 2(1):1-16.

Long, J. Scott

1997   *Regression Models for Categorical and Limited Dependent Variables*. Sage Publications.

Martin, Andrew
    2001   "Congressional Decision Making and the Separation of Powers," *American Political Science Review* 95(2):361-378.
Quinn, Kevin M., Andrew D. Martin, Andrew B. Whitford
    1999   "Voter Choice in Multi-Party Democracies: A Test of Competing Theories and Models," *American Political Science Review* 43:1231-1247.
Rosenstone, Steven J.
    1983   *Forecasting Presidential Elections*. New Haven: Yale University Press.
Sheng, Emile C. J.
    2000   "President Elections Forecast—Using Feeling Thermometer to Predict Undecided Voters," *Journal of Electoral Studies* 7(2):75-107. (in Chinese)
Sheng, Shing-yuan
    1998   "Voting Decision and Election Prediction," *Journal of Electoral Studies* 5(1):37-76. (in Chinese)
Shyu, Huo-yan
    1991   "Party Identification and the Vote Choice: An Analysis of Electorate's Party Images, Partisan Preferences and Voting Behavior in Taiwan," *Journal of Social Science and Philosophy* 4(1):1-57. (in Chinese)
    1998   "The Political Psychology of Lee Teng-hui Complex and Its Effects on Voting Behavior," *Journal of Electoral Studies* 5(2):35-72. (in Chinese)
Wang, Ding-ming
    2001   "The Impacts of Policy Issues on Voting Behavior in Taiwan: A Mixed Logit Approach," *Journal of Electoral Studies* 8(2):65-94. (in Chinese)
Western, Bruce
    1998   "Causal Heterogeneity in Comparative Research: A Bayesian Hierarchical Modeling Approach," *American Journal of Political Science* 42:1253-1259.
Western, Bruce and Simon Jackman
    1994   "Bayesian Inference for Comparative Research," *American Political Science Review* 88:412-423.
Tsai, Chia-hung
    2000   "American Voter Responses to International Political Events and Economic Conditions: 1920-1996," *Euramerica* 30(3):143-191
Tufte, Edward R.
    1978   *Political Control of the Economy*. Princeton, N.J.: Princeton University Press.
You, Ying-lung
    1996   *Ming-yi yu Taiwan cheng-chi bian-chien* (Public Opinion and Changing Taiwan Politics) Taipei: Yue-Tan Press. (in Chinese)
Zellner, Arnold
    1971   *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley.
    1999   "Bayesian and Non-Bayesian Approaches to Scientific Modeling and Inference in Economics and Econometrics," manuscript (http://gsbwww.uchicago.edu/fac/arnold.zellner/index.html)
Zellner, Arnold and Peter E. Rossi
    1984   "Bayesian Analysis of Dichotomous Quantal Response Models," *Journal of Econometric* 25:365-393.

貝式估計在二元勝算模型的應用：
以 2002 年台北市長選舉爲例

蔡佳泓

政治大學選舉研究中心助理研究員

摘　　要

　　貝式統計假設未知的參數有特殊的參數分布，在估計模型時包含未知參數的事先資訊以及最大概似函數。而且，貝式統計以及傳統的統計方法的估計皆是無偏估計。本研究使用 WinBUGS 及 R 語言對 2002 年台北市長的選舉資料估計一個二元勝算的投票模型，得出事後的分布以產生參數估計及可信區間。在加入第一次調查所得出的事先資訊到第二次調查的模型之後，發現貝式模型提供一個更準確的依變項預測值的估計。本研究在方法上的貢獻是提供一個應用貝式統計在二元勝算迴歸模型的例子，並且嘗試納入事前資訊以增強選舉的預測。

關鍵詞：貝式定理，事前資訊，預測事後模擬，二元勝算迴歸，投
　　　　票選擇