Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

A Bayesian allocation of undecided voters

by Balgobin Nandram and Jai Won Choi

June 2008

Statistics Canada Statistique Canada

Canada

Survey Methodology, June 2008
Vol. 34, No. 1, pp. 37-49
Statistics Canada, Catalogue No. 12-001-X

A Bayesian allocation of undecided voters

Balgobin Nandram and Jai Won Choi [1]

Abstract

Data from election polls in the US are typically presented in two-way categorical tables, and there are many polls before the actual election in November. For example, in the Buckeye State Poll in 1998 for governor there are three polls, January, April and October; the first category represents the candidates (*e.g.*, Fisher, Taft and other) and the second category represents the current status of the voters (likely to vote and not likely to vote for governor of Ohio). There is a substantial number of undecided voters for one or both categories in all three polls, and we use a Bayesian method to allocate the undecided voters to the three candidates. This method permits modeling different patterns of missingness under ignorable and nonignorable assumptions, and a multinomial-Dirichlet model is used to estimate the cell probabilities which can help to predict the winner. We propose a time-dependent nonignorable nonresponse model for the three tables. Here, a nonignorable nonresponse model is centered on an ignorable nonresponse model to induce some flexibility and uncertainty about ignorabilty or nonignorability. As competitors we also consider two other models, an ignorable and a nonignorable nonresponse model. These latter two models assume a common stochastic process to borrow strength over time. Markov chain Monte Carlo methods are used to fit the models. We also construct a parameter that can potentially be used to predict the winner among the candidates in the November election.

Key Words: Markov chain Monte Carlo; Metropolis sampler; Multinomial-Dirichlet model; Time-dependent model; Two-way categorical table.

1. Introduction

It is a common practice to use two-way categorical tables to present survey data. Our application is to predict the winner in an election using tables constructed from a short series of polls taken before the actual election. For many surveys, there are missing data and this gives rise to partial classification of the sampled individuals. Little and Rubin (2002, section 1.3) give definitions of the three missing data mechanism (missing completely at random - MCAR, missing at random - MAR, missing not at random - MNAR); ignorable models are used to analyze data from MAR and MCAR mechanisms and nonignorable models for data from MNAR mechanisms. Thus, for the two-way table there are both item nonresponse (one of the two categories is missing) and unit nonresponse (both categories are missing). One may not know how the data are missing, and a model that includes some difference between the observed data and missing data (*i.e.*, nonignorable missing data) may be preferred. For a general $r \times c$ categorical table we address the issue of estimation of the cell probabilities of the two-way table. This problem is important because, with a substantial number of undecided voters, an election prediction based on only the partially observed data may be misleading.

As in Nandram, Cox and Choi (2005) essentially there are four two-way tables, one table with all complete cases and three supplemental tables. Of the three supplemental tables, the first has only row classification (item nonresponse), the second has only column classification (item nonresponse), and the third does not have any classification (unit nonresponse). We have extended the ignorable and nonignorable nonresponse models for two-way categorical tables of Nandram, *et al.* (2005) to accommodate a third category (*i.e.*, time in a short sequence of election polls). We have extended these models even further to include a time-dependent nonignorable non-response structure. The inclusion of the time-dependent structure can provide a more efficient prediction. A Bayesian method permits modeling different patterns of missingness under the ignorability and nonignorability assumptions, and a time-dependent nonignorable non-response model is obtained.

Our application is in Ohio governor's election, and there are several related problems. The sampled persons are categorized by two types of attributes and the cells of such categorical tables are analyzed. However, only partial classification of the individuals is available because some individuals are classified by at most one attribute, and others are left unclassified. Specifically, we use tabular data from the Ohio polls to study the relation between a measure of voters' status (likely to vote and unlikely to vote) and candidate preference (Fisher, Taft and other) to illustrate our methodology. It is interesting that voters' status is related to candidate preference. Also, it is desirable to make an adjustment for undecided voters because the proportion of undecided voters is usually high, and they often decide the final outcome of an election.

1. Balgobin Nandram, Department of Mathematical Sciences, Worcester Polytechnic Institute, 100 Institute Road, Worcester MA 01609. E-mail: balnan@wpi.edu.; Jai Won Choi, Department of Biostatistics, Medical College of Georgia, 1420 15th Street, Augusta, GA 30912. E-mail: jchoi@MCG.edu.

We do not know whether an ignorable nonresponse model or a nonignorable nonresponse model is appropriate, but one may have uncertainty about the ignorability of undecided voters in election polls. Referring to the Buckeye State Poll, Chen and Stasny (2003) stated that "The assumption of nonignorability of the nonresponse may be a reasonable assumption in this study because people might be reluctant to express their preference for an unpopular candidate, or if their current preferences are not firm or accurate enough for the standards of the interview." They also said that while Chang and Krosnick (2001) use ignorable models for their analyses, Chang and Krosnick (2001) suggested that nonresponse might be related to the unobserved data itself. Chen and Stasny (2003) fit three ignorable nonresponse models (A, B and C) and one nonignorable nonresponse model (D). We compare our results with theirs.

Nandram and Choi (2002 a, b) use an expansion model to study nonignorable nonresponse binary data. The expansion model, a nonignorable nonresponse model, degenerates into an ignorable nonresponse model (in the spirit of Draper 1995). This degeneracy occurs when a parameter in the nonignorable nonresponse model is set to a certain value; a good description of the centering idea is given in Nandram, *et al.* (2005, section 1.2). Because it is difficult to carry out this procedure as described, we use an alternative procedure as in Nandram, *et al.* (2005). This permits an expression of uncertainty about ignorability. This is the idea of centering a nonignorable nonresponse model on an ignorable nonresponse model, and we have used it in several of our papers to express uncertainty about ignorability or nonignorability. Here, for nonignorable nonresponse we attempt a related methodology, but the issues for a two-way categorical table are more complex, especially when a third category (*i.e.*, time) is included in these tables.

Using the approach of Chen and Fienberg (1974), Chen and Stasny (2003) describe the two issues we are discussing in this paper. For the two-way categorical tables they can handle item nonresponse only; unit nonresponse is excluded from their analysis. However, they assume that the data are missing at random and show how to obtain maximum likelihood estimators under their model. They also use a nonignorable nonresponse model (D), which they claim is their best model. It is noted in Little and Rubin (2002, chapter 15) that one issue of the nonignorable nonresponse model for this problem is that there are too many parameters, and many parameters are not identified, so they attempted a correction using hierarchical log-linear models. See Nandram, *et al.* (2005) for the case in which there are three supplemental tables.

Our methodology differs from those of Chen and Stasny (2003). The major difference is that we use a Bayesian approach. This permits us to use a method that does not rely on asymptotic theory, incorporate nonignorable missingness into the modeling and obtain time-dependent nonignorable model for estimating the proportion of voters for the three candidates. Looking to predict the winner more convincingly, we have also constructed a new parameter; it is relatively easy to analyze this parameter within the Bayesian paradigm. The Bayesian method permits modeling different patterns of missingness under two different assumptions (*i.e.*, ignorable and nonignorable missingness). Our idea is to start with an ignorable nonresponse model, which is then expanded into a nonignorable nonresponse model, and to the time-dependent nonignorable nonresponse model. It is worth noting that unit nonresponse is also included in our modeling which the other researchers consider as a separate problem using weighting adjustment (*e.g.*, see discussion in Kalton and Kasprzyk 1986). However, there can be nonignorability here as well, and one would need to include unit and item nonresponses simultaneously.

In this paper, our key contribution is to introduce a Bayesian method to analyze data from an $r \times c$ categorical table when there are both item and unit nonresponse, and the missing data mechanism can be nonignorable with a time-dependent structure. In Section 2, we describe the categorical data on voters' status and candidate preference with a time-dependent structure. In Section 3, we describe the methodology to obtain estimates of the cell probabilities incorporating the two types of missing data, and we show how to expand an ignorable nonresponse model into a nonignorable nonresponse model and time-dependent model. We also show how to use Markov chain Monte Carlo methods to fit the nonignorable nonresponse model. In Section 4, we analyze the Ohio election data to demonstrate the versatility of our methods. Finally, Section 5 has concluding remarks.

2. Data on 1998 Ohio Polls

The Center for Survey Research (CSR) at the Ohio State University conducted the Buckeye State Poll (BSP) during the 1998 election for Senator, Governor, Attorney General, State Secretary, Treasurer and Columbus Mayor. In certain months before the election, CSR conducted pre-election surveys as part of the BSP and included additional questions to collect information related to the respondent's likelihood of voting and candidate preference. In the BSP, households are sampled using the Random Digit Dialing (RDD) method, and one adult per household is selected to be interviewed using the last birthday method (Lavrakas 1993).

It is pertinent to briefly describe the RDD method. Polling firms make extensive use RDD, and the main goal of RDD is to develop a representative sample of the overall voter population. RDD sampling assumes that a representative sample cannot be obtained using listed telephone numbers in the directory. Each telephone number has 10 digits, the first three form the area code, the next three form the prefix (colloquially called the exchange), and the last four (suffix) identify a particular subscriber or a household (one household can have more than one phone number). The area codes are geographically based and typically identify localities in a state, and the exchanges can be geographically oriented. There are ten million numbers to dial but roughly less than 25% of these are real telephone numbers. Thus time and money are wasted in dialing unused numbers. We discuss this further in Section 3.

Chen and Stasny (2003) and Chang and Krosnick (2001) analyzed data from three BSP pre-election forecasting polls. Details of each of these three BSP pre-election surveys can be found in Table 1. These BSP pre-election surveys measured respondents' candidate preferences three times (January, April and October) for the November 1998 Ohio Governor race. In addition, respondents were asked for their self-reported likelihood of voting in the upcoming election using two questions. Chang and Krosnick (2001) also used filter variables (such as registered to vote, self-reported likelihood of voting, and voted in the last major election, *etc*.) to obtain those most likely to vote. Thus prediction is based only on the respondents likely to vote. Those registered to vote are classified into likely to vote, unlikely to vote and undecided. Chang and Krosnick (2001) showed that deterministic allocation of undecided respondents provide improvement in forecasting voters' candidate preferences, as compared to exclusion of all undecided respondents. Chen and Stasny (2003) used probability models to allocate the undecided voters and compared their forecasting with that of Chang and Krosnick (2001).

The data set in Chen and Stasny (2003) is slightly different because we use the undecided counts (unit nonresponse) on both variables. A voter can be undecided on at least one of the two categorical variables at each of the three polls. Chen and Stasny (2003) only study the data with undecided in exactly one variable, not both. In Table 1 for the undecided voters in both variables the counts for the January, April and October polls are respectively 5, 3 and 4; these numbers are bolded. In fact, the inclusion of these counts into our model, is an extension of the models in Chen and Stasny, and generalizes our methodology considerably.

We briefly describe the 2×3 categorical table of Ohio election data by voters' status (VS) and candidate preference (CAN). Here VS is a binary variable, and there are two levels: likely to vote and not likely to vote; CAN has three levels: Fisher, Taft, others. There are also undecided voters in VS and CAN. The bulk of the undecided voters come from voters who are "likely to vote" and "unlikely to vote" and the numbers are 173, 142 and 138 for January, April and October respectively; the undecided voters for Fisher, Taft and others are much smaller.

Table 1
Classification of October 1998 Buckeye State Poll by voting status and candidate

	Candidate				
Status	**Fisher**	**Taft**	**Other**	**Undecided**	**Total**
a. January, 1998					
Likely to vote	127	183	8	109	427
Not likely to vote	57	94	4	59	214
Undecided	0	2	0	**5**	7
Total	184	279	12	173	648
b. April, 1998					
Likely to vote	114	135	1	61	311
Not likely to vote	104	149	3	78	334
Undecided	2	6	0	**3**	11
Total	220	290	4	142	656
c. October, 1998					
Likely to vote	112	140	23	61	336
Not likely to vote	96	108	21	73	298
Undecided	7	11	1	**4**	23
Total	215	259	45	138	657

NOTE: These data are taken from Chang and Krosnick (2001); Chen and Stasny (2003) used a very similar data set; they did not use 5, 3, 4, the number of undecided voters in both variables.

In the January 1998 poll, about 73% of the voters are completely classified, 27% have no decision about candidate preference, only 1% did not know whether they would vote or not, and only five persons were completely unclassified among the 648 participants. The data set, used in our study, is presented in Table 1 as a 2×3 categorical table of voters' status and candidate preference. Our problem is to predict the winning candidate by estimating the proportion of final votes for each candidate.

The samples obtained in January, April and October are independent. There is no oversampling for a particular subpopulation or weighting of the original sample. Like many telephone surveys, RDD frame suffers from the common problem of undercoverage. As telephone coverage is not uniform over age, race, sex, income and geography, there is a need to poststratify the original sample to reduce the coverage bias by properly weighting the original data.

We perform a preliminary test of heterogeneity of the cell proportions across the three polls. Assuming a missing at random mechanism, we fill in the undecided votes. We assume that for each row (column) the undecided voters are filled in proportionally to the cell counts. Let n_{tjk} denote the adjusted cell counts with $n_t = \sum_{j=1}^{r} \sum_{k=1}^{c} n_{tjk}$, and let p_{tjk} denote the cell proportions. For a model of heterogeneous proportions, we assume that

$$\boldsymbol{n}_t \mid \boldsymbol{p}_t \overset{\text{ind}}{\sim} \text{Multinomial}(n_t, \boldsymbol{p}_t) \text{ and } \boldsymbol{p}_t \overset{\text{iid}}{\sim} \text{Dirichlet}(\mathbf{1}), t = 1, \ldots, T,$$

where $\mathbf{1}$ is a rc-vector of ones.

For a model of homogeneous proportions, we assume that

$$\boldsymbol{n}_t \mid \boldsymbol{p} \overset{\text{ind}}{\sim} \text{Multinomial}(n_t, \boldsymbol{p}),\, t=1, \ldots, T, \text{ and } \boldsymbol{p} \sim \text{Dirichlet}(\mathbf{1}).$$

Then, the Bayes factor of heterogeneity versus homogeneity is

$$\text{BF} = \frac{1}{\{(rc-1)!\}^2}\left[\prod_{j=1}^{r}\prod_{k=1}^{c}\left\{\frac{\prod_{t=1}^{T} n_{tjk}!}{(\sum_{t=1}^{T} n_{tjk})!}\right\}\right]\frac{\{\sum_{t=1}^{T} n_t + rc - 1\}!}{\prod_{t=1}^{T}(n_t + rc - 1)!}.$$

Thus, using the adjusted cell counts, the logarithm of the Bayes factor (LBF) is approximately 12.4, showing very strong evidence for heterogeneity, and supporting our time-dependent model.

In a similar manner, we have computed the Bayes factors of $\boldsymbol{p}_1 = \boldsymbol{p}_2 \neq \boldsymbol{p}_3$ or $\boldsymbol{p}_1 \neq \boldsymbol{p}_2 = \boldsymbol{p}_3$ versus homogeneity; the LBFs are 7.6 and 4.4 respectively. Thus, the time-dependence occurs for both periods, January-April and April-October.

3. Methodology

We have constructed a time-dependent nonignorable nonresponse model for the 1998 Ohio Poll data. For comparison we have also considered two other models, an ignorable and a nonignorable nonresponse model. These latter two models are not time-dependent because we assume that the three time points come from the same stochastic process (*i.e.*, no correlation across time). Our main contribution is the time-dependent model. We have used the ignorable and nonignorable nonresponse models for a single time point in Nandram, *et al.* (2005). Although these two models are not appropriate in the present context, they are natural to motivate our time dependent non-ignorable nonresponse model. Essentially we start with the ignorable nonresponse model which is expanded into a nonignorable nonresponse model, and we extend the non-ignorable nonresponse model to a time-dependent model.

In RDD stratification and clustering are used to reduce the excess artificial numbers. Stratification by area code and some exchanges is used; geographic ordering (state or region) with systematic selection provides implicit stratification of exchanges. If an exchange is used to form a stratum, there are still ten thousand numbers to dial, still a large waste with numerous redundant numbers. The Mitofsky-Waksberg (see Waksberg 1978) procedure is a stratified two-stage cluster sampling design used to reduce the artificial numbers. Exchange areas are divided into equal size, and a random sample of exchanges is taken with replacement from those eligible (according to the measure of size of each exchange area). Within selected exchange area, a fixed number of telephone numbers is generated at random, without replacement and dialed. Thus, there is also differential probabilities of selection (*i.e.*, unequal cluster sizes) that must be considered in a comprehensive analysis. There are other variants of this procedure. RDD was adequate in 1998 Ohio election, but because of new technological innovations (*e.g.*, cellular phone, email, internet, *etc.*), the usefulness of RDD may be diminished. In this paper, our method and models do not include stratification, clustering or differential probabilities of selection.

Our models are used to estimate the proportions of voters voting for Fisher, Taft and other in the October poll. Then, assuming no catastrophic change in the November election, we predict the proportion of voters voting for Fisher, Taft and other. In this way we can predict the winner in the November election. We are excited by a referee's suggestion that one can use a mixture model to cover the possibility of a catastrophe.

In Sections 3.1 and 3.2 we describe the notations and the three models. In Section 3.3 we show how to fit the time-dependent nonignorable nonresponse model. The ignorable and nonignorable nonresponse models can be fit in a similar manner (see Nandram, *et al.* 2005 for details). In Section 3.4 we show how to specify the two parameters (μ_0 and c_0^2), and in Section 3.5 we show how to do estimation in the October poll and prediction in the November election.

3.1 Notation

Let $I_{tjk\ell} = 1$ if the ℓ^{th} voter belongs to the j^{th} row and k^{th} column of the two-way table at time t and $I_{tjk\ell} = 0$ otherwise, $t = 1, \ldots, T,\ j = 1, \ldots, r,\ k = 1, \ldots, c,\ \ell = 1, \ldots, L.$ That is, $I_{tjk\ell} = 1$ denotes the cell of the $r \times c$ table that a voter belongs to. In our application $T = 3$, $r = 2$ and $c = 3$. Let $J_{ts\ell} = 1$ if the ℓ^{th} voter falls in table s ($s = 1, 2, 3, 4$) and $J_{ts\ell} = 0$ otherwise, $s = 1, \ldots, 4,\ \sum_{s=1}^{4} J_{ts\ell} = 1$; $J_{ts\ell}$ indicates which table an individual belongs to and $\boldsymbol{J}_{t\ell} = (J_{t1\ell}, J_{t2\ell}, J_{t3\ell}, J_{t4\ell})$.

Let the cell counts be $y_{tsjk} = \sum_{\ell=1}^{n} I_{tjk\ell} J_{ts\ell},\ s = 1, 2, 3, 4$ for the four tables at each poll. Here y_{t1jk} are observed and $y_{tsjk},\ s = 2, 3, 4,\quad t = 1, \ldots, T$ are missing (*i.e.*, latent variables). For y_{t1jk} we know that $\sum_{j=1}^{r}\sum_{k=1}^{c} y_{t1jk} = n_{t0}$, the number of individuals with complete data. For y_{t2jk} we know that $\sum_{k=1}^{c} y_{t2jk} = u_{tj}$, where the row margins $u_{tj},\ j = 1, \ldots, r$ are observed. For y_{t3jk} we know that $\sum_{j=1}^{r} y_{t3jk} = v_{tk}$, where the column margins $v_{tk},\ k = 1, \ldots, c$ are observed. For y_{t4jk} we know that $\sum_{j=1}^{r}\sum_{k=1}^{c} y_{t4jk} = w_t$ (unit nonresponse). In this analysis n_{t0}, $\boldsymbol{u}_t$, $\boldsymbol{v}_t$ and w_t are held fixed (*i.e.*, fixed margin analysis) and known.

Whenever it is convenient, we will use notations such as

$$\sum_{s,\,j,\,k} y_{tsjk} \equiv \sum_{s=1}^{4}\sum_{j=1}^{r}\sum_{k=1}^{c} y_{tsjk}, \quad \prod_{s,\,j,\,k} \pi_{tsjk} \equiv \prod_{s=1}^{4}\prod_{j=1}^{r}\prod_{k=1}^{c} \pi_{tsjk}$$

and $\boldsymbol{y}_{t(1)} = (\boldsymbol{y}_{t2}, \boldsymbol{y}_{t3}, \boldsymbol{y}_{t4})$, $\boldsymbol{y}_{t(2)} = (\boldsymbol{y}_{t1}, \boldsymbol{y}_{t3}, \boldsymbol{y}_{t4})$, *etc*., where $\boldsymbol{y}_{ts} = (y_{tsjk}, j = 1, \ldots, r, k = 1, \ldots, c, t = 1, \ldots, T, s = 1, 2, 3, 4)$. Also, we let $\boldsymbol{y}_1 = (\boldsymbol{y}_{11}, \ldots, \boldsymbol{y}_{T1})$ and $\boldsymbol{y}_{(1)} = (\boldsymbol{y}_{1(1)}, \ldots, \boldsymbol{y}_{T(1)})$ with $\boldsymbol{y}_{(1)} = (\boldsymbol{y}_{t(1)}, \ldots, \boldsymbol{y}_{t(4)})$. Also, $\sum_{s,\,j,\,k}^{4,\,r,\,c} y_{tsjk} = n_t$. We will also use $y_{ts\cdot\cdot} = \sum_{j,k} y_{tsjk}$, $y_{t\cdot jk} = \sum_s y_{tsjk}$, *etc*., $\boldsymbol{y}_t = (\boldsymbol{y}_{t1}, \boldsymbol{y}_{t2}, \boldsymbol{y}_{t3}, \boldsymbol{y}_{t4})$ and $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T)$.

3.2 Nonresponse models

Letting $\boldsymbol{I}_{t\ell} = (I_{tjk\ell}, t = 1,\ldots,T, j = 1, \ldots, r, k = 1, \ldots, c, \ell = 1, \ldots, L)$, for all models, we take

$$\boldsymbol{I}_{t\ell} \mid \boldsymbol{p}_t \overset{\text{iid}}{\sim} \text{Multinomial}\{1, \boldsymbol{p}_t\}, \tag{1}$$

where

$$\sum_{j=1}^{r}\sum_{j=1}^{c} p_{tjk} = 1,\ p_{tjk} \geq 0, t = 1, \ldots, T, j = 1, \ldots, r, k = 1, \ldots, c.$$

For the ignorable nonresponse model we take

$$\boldsymbol{J}_{t\ell} \mid \boldsymbol{\pi}_t \overset{\text{iid}}{\sim} \text{Multinomial}\{1, \boldsymbol{\pi}_t\}. \tag{2}$$

That is, there is no dependence on the cell status of an individual. For the nonignorable nonresponse models we take

$$\boldsymbol{J}_{t\ell} \mid \{I_{tjk\ell} = 1, I_{tj'k'\ell} = 0, j \neq j', k \neq k', \boldsymbol{\pi}_{tjk}\} \overset{\text{iid}}{\sim} \text{Multinomial}\{1, \boldsymbol{\pi}_{tjk}\}. \tag{3}$$

Assumption (3) specifies that the probabilities an individual belongs to one of the four tables depend on the two characteristics (*i.e.*, row and column classifications) of the individual. In this manner we incorporate the assumption that the missing data is nonignorable. Note that conditional on the specified parameters in (1)-(3), one voter's behavior is correlated with another at the same time t, but there is independence over time. It is worth noting here that while the parameters in (2) are identifiable, those in (3) are not identifiable. This is where the difficulty in the nonignorable nonresponse model arises, and special attention is needed.

It follows from (1) and (2) that for the ignorable model

$$g(\boldsymbol{p}, \boldsymbol{\pi} \mid \boldsymbol{y}) \propto \prod_{t=1}^{T}\left[\left[\prod_{s=1}^{4} \pi_{ts}^{y_{ts\cdot\cdot}}\right]\left[\prod_{s=1}^{4}\prod_{j=1}^{r}\prod_{k=1}^{c} \frac{p_{tjk}^{y_{tsjk}}}{y_{tsjk}!}\right]\right] \tag{4}$$

subject to $\sum_{k=1}^{c} y_{t2jk} = u_{tj}$, $j = 1, \ldots, r$, $\sum_{j=1}^{r} y_{t3jk} = v_{tk}$, $k = 1, \ldots, c$, and $\sum_{j=1}^{r}\sum_{k=1}^{c} y_{t4jk} = w_t$. Note that under ignorability the likelihood function in (4) separates into two pieces, one that contains the π_{ts} only and the other the p_{tjk}, and inference about these two parameters are unrelated; see Section 3.2 of Nandram, *et al.* (2005) for the original discussion of this model. Also, it follows from (1) and (3) that for the nonignorable nonresponse models the augmented likelihood function for $\boldsymbol{p}, \boldsymbol{\pi}, \boldsymbol{y}_{(1)} \mid \boldsymbol{y}_1$ is

$$g(\boldsymbol{p}, \boldsymbol{\pi}, \boldsymbol{y}_{(1)} \mid \boldsymbol{y}_1) \propto \prod_{t=1}^{T}\left[\prod_{s,\,j,\,k}^{4,\,r,\,c} \frac{\pi_{sjk}^{y_{tsjk}}}{y_{tsjk}!}\right]\left[\prod_{j,\,k}^{r,\,c} p_{tjk}^{y_{t\cdot jk}}\right] \tag{5}$$

subject to $\sum_{k=1}^{c} y_{t2jk} = u_{tj}$, $j = 1, \ldots, r$, $\sum_{j=1}^{r} y_{t3jk} = v_{tk}$, $k = 1, \ldots, c$, and $\sum_{j=1}^{r}\sum_{k=1}^{c} y_{t4jk} = w_t$; see Nandram, *et al.* (2005) for a description of identifiability in a similar situation.

For the ignorable and nonignorable nonresponse models, we take

$$\boldsymbol{p}_t \mid \boldsymbol{\mu}_2, \tau_2 \overset{\text{iid}}{\sim} \text{Dirichlet}\ (\boldsymbol{\mu}_2 \tau_2), t = 1, \ldots, T+1, \tag{6}$$

where we consider prediction at $T+1$, one step ahead (November). The probabilistic structure in (6) permits a "borrowing of strength" across time. Note that the k-dimensional vector $\boldsymbol{x}$ has a Dirichlet distribution if $p(\boldsymbol{x} \mid \boldsymbol{\alpha}) = \prod_{j=1}^{k} x_j^{\alpha_j - 1}/D(\boldsymbol{\alpha})$, $x_j \geq 0$, $j = 1, \ldots, k$, $\sum_{j=1}^{k} x_j = 1$, where $D(\boldsymbol{\alpha})$ is the Dirichlet function and $\alpha_j > 0$, $j = 1, \ldots, k$. For a quick reference see Ghosh and Meeden (1997, pages 42, 50, 127) in connection with the Polya urn distribution, and more appropriately its use as a conjugate prior in multinomial sampling; starting with our first paper (*i.e.*, Nandram 1998) we have been using the Dirichlet-multinomial extensively in our research.

We next describe the stochastic models for the $\boldsymbol{\pi}_{tjk}$. For the ignorable nonresponse model, we take

$$\boldsymbol{\pi}_t \overset{\text{iid}}{\sim} \text{Dirichlet}\ (\mathbf{1}),\ t = 1, \ldots, T, \tag{7}$$

where $\mathbf{1}$ is a four-dimensional vector of ones. We need (7) because T is small (*i.e.*, $T = 3$ in our application). Thus, we use the uniform prior in R^4 (essentially noniformative); otherwise we will have to specify the unknown parameters of the Dirichlet distribution with virtually no data. For the nonignorable nonresponse models we take

$$\boldsymbol{\pi}_{tjk} \mid \boldsymbol{\mu}_1, \tau_1 \overset{\text{iid}}{\sim} \text{Dirichlet}\ (\boldsymbol{\mu}_1 \tau_1), t = 1, \ldots, T, j = 1, \ldots, r, k = 1, \ldots, c. \tag{8}$$

First, we note that (8) provides a "borrowing of strength" across time. More importantly, because $\boldsymbol{\pi}_{tjk}$ are not identifiable so are $\boldsymbol{\mu}_1$ and τ_1. One possible way out of this dilemma is to "center" the nonignorable nonresponse model on the ignorable nonresponse model.

For the time-dependent model, we take

$$\boldsymbol{p}_t \mid \boldsymbol{p}_{t-1}, \tau_2 \overset{\text{iid}}{\sim} \text{Dirichlet}\ (\boldsymbol{p}_{t-1} \tau_2), t = 1, \ldots, T+1, \tag{9}$$

where $\boldsymbol{p}_0$ is also unknown. Note that

$$E\{\boldsymbol{p}_t \mid \boldsymbol{p}_{t-1}, \tau_2\} = \boldsymbol{p}_{t-1}, t = 1, \ldots, T+1;$$

so that $\{p_t\}$, *a priori*, is a martingale vector. Here T is small (*i.e.*, $T = 3$). Thus, this time-dependent structure seems more appropriate, and can potentially provide improved precision. Note also that we have taken $\boldsymbol{p}_0 \sim$ Dirichlet (**1**).

Finally, we specify prior densities for the hyper-parameters. First, we take

$$\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \overset{\text{iid}}{\sim} \text{Dirichlet } (\mathbf{1}), \qquad (10)$$

essentially noniformative prior densities.

Finally, τ_1 and τ_2 are independent and identically distributed random variables from

$$f(x) = 1/(1+x)^2, \; x \geq 0. \qquad (11)$$

Again this is an essentially noniformative prior density. Note that $\boldsymbol{\mu}_1$ and τ_1 do not exist in the ignorable nonresponse model. Gelman (2006) recommended priors like (11) instead of the ill-behaved proper diffuse gamma priors.

For the nonignorable nonresponse models we need to be more careful to specify the prior density of τ_1 because $\boldsymbol{\pi}_{tjk}$ are not identifiable. Here we attempt to "center" the nonignorable nonresponse models on the ignorable nonresponse model. In (8) the parameter τ_1 tells us about the closeness of the nonignorable model to the ignorable model. For example, if τ_1 is small, the $\boldsymbol{\pi}_{tjk}$ will be very different, and if τ_1 is large, the $\boldsymbol{\pi}_{tjk}$ will be very similar. Thus, *a priori* inference will be sensitive to the choice of τ_1, and one has to be careful in choosing τ_1. We would like to choose a prior density for τ_1 so that the nonignorable nonresponse model is kept close to the ignorable nonresponse model. Thus, we take

$$\tau_1 \sim \text{Gamma } (1/c_0^2, \; 1/\mu_0 c_0^2), \qquad (12)$$

where $E(\tau_1) = \mu_0$ and $\text{CV}(\tau_1) = c_0$, with CV the coefficient of variation; both μ_0 and c_0 are to be specified. We use the prior (12) because by an appropriate choice of μ_0 and c_0 it is possible to center the nonignorable nonresponse model on the ignorable nonresponse model. Of course, one can use other convenient proper priors with parameters like μ_0 and c_0 to facilitate the centering. In Section 3.4 we will use samples from the posterior density of τ_1 under the ignorable nonresponse model to specify μ_0 and c_0.

For each of the three models, it is easy to write down the joint prior density of the parameters. For example, for the time-dependent model the joint prior density is

$$p(\boldsymbol{p}, \boldsymbol{\pi}, \boldsymbol{\mu}_1, \tau_1, \tau_2) \propto \tau_1^{1/c_0^2 - 1} e^{-\tau_1/\mu_0 c_0^2} \frac{1}{(1+\tau_2)^2}$$
$$\times \prod_{t=1}^{T} \left\{ \frac{\prod_{j=1}^{r} \prod_{k=1}^{c} p_{tjk}^{p_{t-1jk}\tau_2 - 1}}{D(\boldsymbol{p}_{t-1}\tau_2)} \prod_{j=1}^{r} \prod_{k=1}^{c} \frac{\prod_{s=1}^{4} \pi_{tsjk}^{\mu_{1s}\tau_1 - 1}}{D(\boldsymbol{\mu}_1 \tau_1)} \right\}, \qquad (13)$$

where $D(\cdot)$ is the Dirichlet function.

3.3 Fitting the time-dependent nonignorable nonresponse model

Combining the likelihood function in (5) with the joint prior density in (13) via Bayes' theorem, the joint posterior density of the parameters $\boldsymbol{\pi}, \boldsymbol{p}\ \boldsymbol{\mu}_1, \tau_1\ \tau_2$ and the latent variables $\boldsymbol{y}_{(1)}$ is

$$\pi(\boldsymbol{p}, \boldsymbol{\pi}, \boldsymbol{\mu}_1, \tau_1, \tau_2, \boldsymbol{y}_{(1)} \mid \boldsymbol{y}_1)$$
$$\propto \tau_1^{1/c_0^2 - 1} e^{-\tau_1/\mu_0 c_0^2} \frac{1}{(1+\tau_2)^2} \prod_{t=1}^{T} \left[\prod_{s,j,k}^{4,r,c} \frac{\pi_{tsjk}^{y_{tsjk}}}{y_{tsjk}!} \right] \left[\prod_{j,k}^{r,c} p_{tjk}^{y_{t\cdot jk}} \right]$$
$$\times \prod_{t=1}^{T} \left\{ \frac{\prod_{j=1}^{r} \prod_{k=1}^{c} p_{tjk}^{p_{t-1jk}\tau_2 - 1}}{D(\boldsymbol{p}_{t-1}\tau_2)} \prod_{j=1}^{r} \prod_{k=1}^{c} \frac{\prod_{s=1}^{4} \pi_{tsjk}^{\mu_{1s}\tau_1 - 1}}{D(\boldsymbol{\mu}_1 \tau_1)} \right\} \qquad (14)$$

subject to $\sum_{k=1}^{c} y_{t2jk} = u_{tj}, \; j = 1, \ldots, r, \; \sum_{j=1}^{r} y_{t3jk} = v_{tk}, \; k = 1, \ldots, c$, and $\sum_{j=1}^{r} \sum_{k=1}^{c} y_{t4jk} = w_t, \; t = 1, \ldots, T$.

The posterior density in (14) is complex, so we will use Markov chain Monte Carlo methods to fit it. However, it is easy to fit the time-dependent model using the griddy Metropolis-Hastings sampler (our terminology) as we will describe. Also, in a similar manner using the griddy Gibbs sampler (Ritter and Tanner 1992), it is easy to fit the ignorable and the nonignorable nonresponse models. We obtain a sample from the joint posterior density in order to make inference about the parameters. Specifically, we need to make inference about $\boldsymbol{p}_t$. To run the Metropolis-Hastings sampler, we need the conditional posterior density of each of the parameters given the others.

First, we consider the conditional posterior probability mass functions of $\boldsymbol{y}_{ts}, s = 2, 3, 4, \; t = 1, \ldots, T$ given $\boldsymbol{y}_{t(s)}$, $\boldsymbol{p}_t, \boldsymbol{\pi}_{tjk}, \; j = 1, \ldots, r, k = 1, \ldots, c$. From (14) it is clear that under the conditional posterior density the $\boldsymbol{y}_{ts}, t = 1, \ldots, T, s = 2, 3, 4$, are independent multinomial random vectors. Specifically, letting $\boldsymbol{p} = (p_{tjk}, t = 1, \ldots, T, \; j = 1, \ldots, r, \; k = 1, \ldots, c)$ and $\boldsymbol{\pi} = (\boldsymbol{\pi}_{tjk}, t = 1, \ldots, T, \; j = 1 \ldots, r, \; k = 1, \ldots, c)$,

$$\boldsymbol{y}_{t2j} \mid \{\boldsymbol{y}_{t1}, \boldsymbol{p}, \boldsymbol{\pi}\} \overset{\text{ind}}{\sim} \text{Multinomial}(u_{tj}, \boldsymbol{q}_{tj}^{(2)}) \quad j = 1, \ldots, r),$$

$$\boldsymbol{y}_{t3k} \mid \{\boldsymbol{y}_{t1}, \boldsymbol{p}, \boldsymbol{\pi}\} \overset{\text{ind}}{\sim} \text{Multinomial}(v_{tk}, \boldsymbol{q}_{tk}^{(3)}), \; k = 1, \ldots, c),$$

$$\boldsymbol{y}_{t4} \mid \{\boldsymbol{y}_{t1}, \boldsymbol{p}, \boldsymbol{\pi}\} \sim \text{Multinomial}(w_t, \boldsymbol{q}_t^{(4)}), \qquad (15)$$

where $q_{tjk}^{(2)} = \pi_{t2jk}\, p_{tjk} / \sum_{k'=1}^{c} \pi_{t2jk'}\, p_{tjk'},\ k = 1, \ldots, c,\ q_{tjk}^{(3)} = \pi_{t3jk}\, p_{tjk} / \sum_{j'=1}^{r} \pi_{t3j'k}\, p_{tj'k},\ j = 1, \ldots, r$ and $q_{tjk}^{(4)} = \pi_{t4jk} p_{tjk} / \sum_{j'=1}^{r} \sum_{k'=1}^{c} \pi_{t4j'k'} p_{tj'k'},\ j = 1, \ldots, r,\ k = 1, \ldots, c,\ t = 1, \ldots, T$

The conditional posterior density of $\boldsymbol{\pi}_{tjk}$ is given by

$$\boldsymbol{\pi}_{tjk} \mid \{\boldsymbol{\mu}, \tau, \boldsymbol{y}_t\} \overset{\text{ind}}{\sim} \text{Dirichlet}(y_{t1jk} + \mu_{t1}\tau_1, y_{t2jk} + \mu_{t2}\tau_1, y_{t3jk} + \mu_{t3}\tau_1, y_{t4jk} + \mu_{t4}\tau_1) \quad (16)$$

with independence over $t = 1, \ldots, T,\ j = 1, \ldots, r,\ k = 1, \ldots, c.$

The conditional posterior density for $\boldsymbol{p}_t,\ t = 1, \ldots, T$ is more difficult. We note that

$$\pi(\boldsymbol{p}_0 \mid \text{else}, \boldsymbol{y}_1) \propto \frac{\prod_{j=1}^{r} \prod_{k=1}^{c} p_{1jk}^{p_{0jk}\tau_2 - 1}}{D(\boldsymbol{p}_0 \tau_2)} \quad (17)$$

and

$$\pi(\boldsymbol{p}_t \mid \text{else}, \boldsymbol{y}_1)$$

$$\propto \left\{ \prod_{j=1}^{r} \prod_{k=1}^{c} p_{tjk}^{y_{t\cdot jk} + p_{t-1jk}\tau_2 - 1} \right\} \frac{\prod_{j=1}^{r} \prod_{k=1}^{c} p_{t+1jk}^{p_{tjk}\tau_2 - 1}}{D(\boldsymbol{p}_t \tau_2)},\ t = 1, \ldots, T, \quad (18)$$

where "else" refers to all of the parameters in $(\boldsymbol{p}, \boldsymbol{\pi}, \boldsymbol{\mu}_1, \tau_1, \tau_2, \boldsymbol{y}_{(1)})$ excluding $\boldsymbol{p}_0$ in (17) or $\boldsymbol{p}_t$ in (18). We show how to draw samples from (17) and (18) in Appendix A.

Next, we consider the hyper-parameters. Letting $\delta_s = \prod_{t=1}^{T} \prod_{j=1}^{r} \prod_{k=1}^{c} \pi_{sjk}$, and $\boldsymbol{\pi} = (\pi_{tjk},\ t = 1, \ldots, T,\ j = 1, \ldots, r,\ k = 1, \ldots, c)$, the joint conditional posterior density of $\boldsymbol{\mu}_1, \tau_1$ is

$$p(\boldsymbol{\mu}_1, \tau_1 \mid \boldsymbol{\pi}) \propto \frac{\prod_{s=1}^{4} \delta_s^{\mu_{1s}\tau_1}}{\{D(\boldsymbol{\mu}_1 \tau_1)\}^{rcT}} \tau_1^{1/c_0^2 - 1} e^{-\tau_1/\mu_0 c_0^2},$$

where

$$\sum_{s=1}^{4} \mu_{1s} = 1,\ \mu_{1s} \geq 0,\ s = 1, 2, 3, 4,\ \tau_1 > 0.$$

We do not need to get a sample directly from $p(\boldsymbol{\mu}_1 \mid \tau_1, \boldsymbol{\pi})$. But, letting $\boldsymbol{\mu}_{1(s)}$ denote the vector of all components of $\boldsymbol{\mu}_1$ except μ_{1s}, we have

$$p(\mu_{1s} \mid \boldsymbol{\mu}_{1(s)}, \tau_1, \boldsymbol{\pi})$$

$$\propto \frac{\delta_s^{\mu_{1s}\tau_1}}{\{\Gamma(\mu_{1s}\tau_1)\}^{rcT}} \frac{\delta_4^{(1 - \mu_{11} - \mu_{12} - \mu_{13})\tau_1}}{\{\Gamma((1 - \mu_{11} - \mu_{12} - \mu_{13})\tau_1)\}^{rcT}},$$

$$0 \leq \mu_{1s} \leq 1 - \sum_{s'=1,\, s' \neq s}^{3} \mu_{1s'},\ s = 1, 2, 3. \quad (19)$$

We use a grid method to draw a sample from $p(\mu_{1s} \mid \boldsymbol{\mu}_{1(s)}, \tau_1, \boldsymbol{\pi})$. We started by using 50 grids (*i.e.*, we have divided the range of μ_{1s}, $(0, 1 - \sum_{s'=1,\, s' \neq s}^{3} \mu_{1s'})$, into 50 intervals of equal widths) to form an approximate probability mass function of $\mu_{1s},\ s = 1, 2, 3$. We first draw a random variable from this probability mass function to indicate which of the 50 intervals is selected. Then, for μ_{1s} we draw a uniform random variable in this interval. This procedure is efficient because μ_{1s} is bounded, the intervals are very narrow, and it is very "cheap" to construct the discrete probability mass function for each $\mu_{1s},\ s = 1, 2, 3$. Finally, μ_{14} is obtained from its conditional posterior density by taking $\mu_{14} = 1 - \sum_{s=1}^{3} \mu_{1s}$.

The conditional posterior density of τ_1 is

$$p(\tau_1 \mid \boldsymbol{\mu}_1, \boldsymbol{\pi}) \propto \left[\prod_{s=1}^{4} \frac{\delta_s^{\mu_{1s}\tau_1}}{\{\Gamma(\mu_{1s}\tau_1)\}^{rcT}} \right] \tau_1^{1/c_0^2 - 1} e^{-\tau_1/\mu_0 c_0^2},\ \tau_1 > 0. \quad (20)$$

To draw a random deviate from (20), we proceed in the same manner as for (19), except that we transform τ_1 from the positive half of the real line to $(0, 1)$. (It is more convenient to perform a grid approximation to a density in a bounded interval.) Thus, letting $\tau_1 = \phi/(1 - \phi)$ in (20), we have

$$p(\phi \mid \boldsymbol{\mu}_1, \boldsymbol{\pi})$$

$$\propto \frac{1}{(1 - \phi)^2} \left\{ \left[\prod_{s=1}^{4} \frac{\delta_s^{\mu_{1s}\tau_1}}{\{\Gamma(\mu_{1s}\tau_1)\}^{rcT}} \right] \tau_1^{1/c_0^2 - 1} e^{-\tau_1/\mu_0 c_0^2} \right\}_{\tau_1 = \frac{\phi}{1 - \phi}},\ 0 < \phi < 1.$$

Again, we started by using 50 intervals of equal width to draw ϕ, and the random deviate for τ_1 is $\phi/(1 - \phi)$.

Letting $p = (p_{tjk},\ t = 1, \ldots, T,\ j = 1, \ldots, r,\ k = 1, \ldots, c)$, the conditional posterior density of τ_2 is

$$\pi(\tau_2 \mid \boldsymbol{p}) \propto \frac{1}{(1 + \tau_2)^2} \prod_{t=1}^{T} \left\{ \frac{\prod_{j=1}^{r} \prod_{k=1}^{c} p_{tjk}^{p_{t-1jk}\tau_2 - 1}}{D(\boldsymbol{p}_{t-1}\tau_2)} \right\},\ \tau_2 > 0. \quad (21)$$

A sample is obtained in a manner similar to τ_1 in (20).

We have extensive experience in using the grid approximation. However, one has to be careful in using the grid approximation for parameters close to 0 or 1 in the in the interval $[0, 1]$. One would need to use a grid approximation in an interval near the boundary; this can be obtained by trial and error in looking at the output of the sampler as it progresses. If a parameter in $[0, 1]$ is likely to be away from 0 or 1, then the grid method works fine; this is the case for the μ_{1s}'s. However, for a parameter like τ_1 (can be very large), when transformed to ϕ in the interval $[0, 1]$, ϕ can be very large (near to 1). If the transformed value is like 0.999, one needs to adjust the grid search to be in an interval containing 0.999. This has to be done by trial and error; one needs to look at the output of ϕ as the sampler progresses, and adjust the interval accordingly. For example, if 100 grid points are equally spaced in $[0, 1]$ such as 0.01, 0.02, 0.03, …. 0.99, and the parameter is likely to be around 0.999, although we draw uniformly in the selected grid interval, these grid points are not going to be very efficient.

The Metropolis-Hastings sampler is executed by drawing a random deviate from each of (15), (16), (17), (18), (19), (20) and (21) iterating the entire procedure until convergence. This is an example of the griddy Metropolis-Hastings sampler (Ritter and Tanner 1992). We obtain a sample from the posterior densities corresponding to the ignorable and nonignorable nonresponse models in a similar manner. For all models, we use a sample of $M = 1{,}000$ from the posterior densities to do estimation and prediction. We monitored the algorithm for convergence by looking at the trace plots of each parameter versus iteration order and we studied the autocorrelation coefficient. We used a griddy Gibbs sampler to fit the ignorable and nonignorable nonresponse models. We used a "burn in" of 1,000 iterates and we took every tenth thereafter. This procedure works well.

However, for the time-dependent model, we used a griddy approximation to the conditional posterior of $\boldsymbol{p}_0$, but Metropolis steps for $\boldsymbol{p}_t, t = 1,\ldots,T$. The Metropolis steps did not work well because the jumping probabilities are 0.67, 0.65 and 0.73 for the three conditional posterior densities of $\boldsymbol{p}_1$, $\boldsymbol{p}_2$, and $\boldsymbol{p}_3$, but they are recommended to be between 0.25 and 0.50 (Gelman, Roberts and Gilks 1996); tuning did not help. So we used grid approximations to these three conditional posterior densities as well. The grid approximations are very accurate. In all grid approximations, we started with 50 grids, and we increased the number of grids until our estimates of all $\boldsymbol{p}_T$, $\boldsymbol{p}_{T+1}$, $\boldsymbol{\mu}_2$, τ_2 do not change. We found that 200 grids were adequate in all cases (*i.e.*, for $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \tau_1, \tau_2$). Also, we found that although the Metropolis-Hastings sampler did not work as well as we wanted, the estimates of the cell proportions are virtually the same from both samplers. The Metropolis-Hastings sampler was run for 25,000 iterations with a "burn in" of 5,000 and thinning by choosing every twentieth.

Finally, we stored the sample from the joint posterior density for further analysis. Specifically, for the ignorable and the nonignorable nonresponse models, we need the sample of size M from $\{(\boldsymbol{\mu}_2^{(h)}, \tau_2^{(h)}, \boldsymbol{p}_{T-1}^{(h)}, \boldsymbol{p}^{(h)}), h = 1,\ldots,M\}$, and for the time-dependent model we need the sample of size M from $(\tau_2^{(h)}, \boldsymbol{p}_{T-1}^{(h)}, \boldsymbol{p}^{(h)}), h = 1,\ldots,M$.

3.4 Specification of μ_0 and c_0^2

Finally, we describe how to specify μ_0 and c_0 in (12). This is important because it permits us to "center" the nonignorable nonresponse model on the ignorable nonresponse model (*i.e.*, an expansion model). This procedure is in the spirit of Nandram, *et al.* (2005).

We have drawn a sample of $\boldsymbol{\pi}_t^{(h)}, t = 1,\ldots,T, h = 1,\ldots,M$, $M = 1{,}000$ iterates from the ignorable nonresponse model, and computed $\boldsymbol{\pi}^{(h)} = \sum_{t=1}^T \boldsymbol{\pi}_t^{(h)}/T, h = 1,\ldots,M$. Then, using the griddy Gibbs sampler, we fit the model

$$\boldsymbol{\pi}^{(h)} \overset{\text{iid}}{\sim} \text{Dirichlet}(\boldsymbol{\mu}_1\tau_1),$$

$$\boldsymbol{\mu}_1 \sim \text{Dirichlet}(\mathbf{1}),\ p(\tau_1) = 1/(1+\tau_1)^2, \tau_1 > 0,$$

with *a priori* $\boldsymbol{\mu}_1$ and τ_1 independent, to obtain a sample $\tau_1^{(h)}, h = 1,\ldots,M$. We have drawn 1,500 iterates with a "burn in" of 500 to get $M = 1{,}000$ iterates.

Finally, taking $a = M^{-1}\sum_{h=1}^M \tau_1^{(h)}$ and $b = (M-1)^{-1}\sum_{h=1}^M (\tau_1^{(h)} - a)^2$, we set

$$c_0 = \sqrt{b}/a \text{ and } \mu_0 = a.$$

For the election data, our procedure gives $c_0 = 0.031$ and $\mu_0 = 2.431$. This specification will hold the nonignorable nonresponse model close to the ignorable nonresponse model, thereby providing a possible centering mechanism.

To study sensitivity to the misspecification of the prior density of τ_1, we use two constants, κ_1 and κ_2, such that *a priori*

$$\tau_1 \sim \text{Gamma}\,(1/\kappa_1^2 c_0^2,\ 1/\kappa_1^2 \kappa_2 \mu_0 c_0^2)$$

with varying values of κ_1 and κ_2. It is worth noting that $E(\tau_1) = \kappa_2\mu_0$ and $\text{CV}(\tau_1) = \kappa_1 c_0$; thus increasing κ_2 means increasing τ_1 which, in turn, means increasing precision *a priori* but not necessarily *a posteriori*. We will study the sensitivity to the specification of κ_1 and κ_2 when we describe the data analysis.

3.5 Estimation and prediction

We show how to improve estimation (*i.e.*, Rao-Blackwellization) in the October poll, and how to do prediction in the November election.

For the ignorable and nonignorable nonresponse models,

$$\begin{aligned} g(\boldsymbol{p}_T \mid \boldsymbol{y}_1) &= \int g(\boldsymbol{p}_T \mid \boldsymbol{\mu}_2, \tau_2)\,\pi(\boldsymbol{\mu}_2, \tau_2 \mid \boldsymbol{y}_1)\,d\boldsymbol{\mu}_2\,d\tau_2 \\ &\approx \frac{1}{M}\sum_{h=1}^M g(\boldsymbol{p}_T \mid \boldsymbol{\mu}_2^{(h)}, \tau_2^{(h)}), \end{aligned} \tag{22}$$

where $\boldsymbol{p}_T \mid \boldsymbol{\mu}_2, \tau_2 \sim \text{Dirichlet}(\boldsymbol{\mu}_2\tau_2)$, and for the time-dependent model,

$$\begin{aligned} g(\boldsymbol{p}_T \mid \boldsymbol{y}_1) &= \int g(\boldsymbol{p}_T \mid \boldsymbol{p}_{T-1}, \tau_2)\,\pi(\boldsymbol{p}_{T-1}, \tau_2 \mid \boldsymbol{y}_1)\,d\boldsymbol{p}_T\,d\tau_2 \\ &\approx \frac{1}{M}\sum_{h=1}^M g(\boldsymbol{p}_T \mid \boldsymbol{p}_{T-1}^{(h)}, \tau_2^{(h)}), \end{aligned} \tag{23}$$

where $\boldsymbol{p}_T \mid \boldsymbol{p}_{T-1}, \tau_2 \sim \text{Dirichlet}(\boldsymbol{p}_{T-1}\tau_2)$.

We obtain (predict) the cell proportions for November as follows. The ignorable or nonignorable nonresponse model, posterior density of $\boldsymbol{p}_{T+1}$ is

$$\begin{aligned} g(\boldsymbol{p}_{T+1} \mid \boldsymbol{y}_1) &= \int g(\boldsymbol{p}_{T+1} \mid \boldsymbol{\mu}_2, \tau_2)\,\pi(\boldsymbol{\mu}_2, \tau_2 \mid \boldsymbol{y}_1)\,d\boldsymbol{\mu}_2\,d\tau_2 \\ &\approx \frac{1}{M}\sum_{h=1}^M g(\boldsymbol{p}_{T+1} \mid \boldsymbol{\mu}_2^{(h)}, \tau_2^{(h)}), \end{aligned} \tag{24}$$

where $\boldsymbol{p}_{T+1} \mid \boldsymbol{\mu}_2, \tau_2 \sim \text{Dirichlet}(\boldsymbol{\mu}_2 \tau_2)$. For the time-dependent

$$g(\boldsymbol{p}_{T+1} \mid \boldsymbol{y}_1) = \int g(\boldsymbol{p}_{T+1} \mid \boldsymbol{p}_T, \tau_2)\, \pi(\boldsymbol{p}_T, \tau_2 \mid \boldsymbol{y}_1)\, d\,\boldsymbol{p}_T\, d\,\tau_2$$

$$\approx \frac{1}{M} \sum_{h=1}^{M} g(\boldsymbol{p}_{T+1} \mid \boldsymbol{p}_T^{(h)}, \tau_2^{(h)}), \tag{25}$$

where $\boldsymbol{p}_{T+1} \mid \boldsymbol{p}_T, \tau_2 \sim \text{Dirichlet}(\boldsymbol{p}_T \tau_2)$.

Thus, by (22), (23), (24) and (25), estimation and prediction are straight forward. For example, consider the time-dependent model. For estimation, by (24) for each h, we draw a random deviate $\boldsymbol{p}_T \mid \boldsymbol{p}_{T-1}^{(h)}, \tau_2^{(h)} \sim \text{Dirichlet}(\boldsymbol{p}_{T-1}^{(h)} \tau_2^{(h)})$, denoted by $\boldsymbol{p}_T^{(h)}, h = 1, \ldots, M$. For prediction, by (25) for each h, we draw a random deviate $\boldsymbol{p}_{T+1} \mid \boldsymbol{p}_T^{(h)}, \tau_2^{(h)} \sim \text{Dirichlet}(\boldsymbol{p}_T^{(h)} \tau_2^{(h)})$, denoted by $\boldsymbol{p}_T^{(h)}, h = 1, \ldots, M$. Thus, inference about $\boldsymbol{p}_T$ and $\boldsymbol{p}_{T+1}$ is made in the usual manner. The procedure is similar for the ignorable and nonignorable nonresponse models.

4. Data analysis

In this section we compare our models with those of Chen and Stasny (2003) and the actual (November election) outcomes. We have introduced a new parameter to help predict the outcome of the election. We also study extensively sensitivity of inference to choices of κ_1 and κ_2. Based on our procedure, we have specified the coefficient of variation, $c_0 = 0.031$, and the mean, $\mu_0 = 2.431$, of the prior distribution of τ_1.

In Table 2 we compare inference about the proportions of October voters allocated to the three candidates by our models and those of Chen and Stasny (2003). In this table the results are based on the prior $\tau_1 \sim \text{Gamma}(1/c_0^2, 1/\mu_0 c_0^2)$ (*i.e.*, $\kappa_1 = \kappa_2 = 1$). We also present the actual proportions taken from Chang and Krosnick (2001). The actual proportions are (0.45, 0.50, 0.05) for Fisher, Taft and other. Using our time-dependent nonresponse model these proportions are estimated to be (0.41, 0.50, 0.09). These compare favorably with the actual outcomes. The corresponding estimates are (0.41, 0.51, 0.08) for the ignorable nonresponse model and (0.40, 0.50, 0.09) for the nonignorable nonresponse model. The best result of Chen and Stasny (2003) is obtained from their Model D, and their estimates are (0.42, 0.51, 0.07). We have provided 95% credible intervals for our estimates, but within the approach of Chen and Stasny (2003) it is relatively more difficult to provide similar intervals. Also, in Table 2 we present estimates of the predicted proportions for the November elections. The point predictors are similar to the point estimates except for the predicted proportion going to Taft under the ignorable nonresponse model. However, as expected the 95% credible intervals for the predicted proportions are much wider. For example, under the time-dependent model 95% credible interval for the proportion voting for Taft in the October poll is (0.41, 0.60) and for prediction it is (0.21, 0.78). Thus, while the point estimates and predictions do indicate the winner, the variability indicates no difference between Taft and Fisher. We will look at this further.

Table 2
Comparison of the proportion of likely voters for the October 1998 poll and prediction for November 1998 election for different models with actual outcome

Status	Fisher	Taft	Other
Sample Estimate	0.41	0.51	0.08
Approximate 95% CI	(0.35, 0.47)	(0.45, 0.57)	(0.05, 0.11)
Actual Outcome	0.45	0.50	0.05
a. Estimation			
Chen/Stasny models A,B,C	0.41	0.51	0.08
Chen/Stasny model D	0.42	0.51	0.07
Chen/Stasny model E	0.41	0.51	0.08
Ignorable model	0.41	0.51	0.08
95% CI	(0.35, 0.46)	(0.46, 0.57)	(0.05, 0.12)
Nonignorable model	0.41	0.50	0.09
95% CI	(0.32, 0.51)	(0.40, 0.60)	(0.05, 0.17)
Time-dependent model	0.41	0.50	0.09
95% CI	(0.32, 0.52)	(0.41, 0.60)	(0.05, 0.16)
b. Prediction			
Ignorable model	0.41	0.54	0.05
95% CI	(0.15, 0.70)	(0.25, 0.81)	(0.00, 0.22)
Nonignorable model	0.42	0.52	0.06
95% CI	(0.15, 0.70)	(0.22, 0.79)	(0.00, 0.28)
Time-dependent model	0.41	0.50	0.09
95% CI	(0.15, 0.71)	(0.21, 0.78)	(0.00, 0.31)

NOTE: $\tau_1 \sim \text{Gamma}(1/c_0^2, 1/\mu_0 c_0^2)$, where $c_0 = 0.031$ and $\mu_0 = 2.431$.

Although our estimates from the time-dependent model are close to the actual estimates, the 95% credible intervals for p_{311} and p_{312} overlap, thereby making it difficult to predict Taft is the winner. Although the 95% credible intervals for our other models are shorter, the point estimates are not so good and they still overlap. One weakness in our analysis in Table 2 is that we have ignored the correlation between the two estimates (*i.e.*, we should really study the difference $p_{312} - p_{311}$, the margin of winning).

In Table 3 we present estimates of $\Lambda_e = p_{312} - p_{311}$ and $\Lambda_p = p_{412} - p_{411}$ at $\kappa_1 = \kappa_2 = 1$ for the three models. We have also included the numerical standard error (NSE) which is a measure of how well the numerical results can be reproduced; we have used the batch-means method to compute it. Small NSEs mean that if we repeat the entire computation the same way (*i.e.*, using another 1,000 iterates), we should see very little difference between the two sets of answers. In Table 3 the NSEs are small. The point estimators and predictors are all positive showing that Taft is the winner in both the October poll and the November election. However, the variability dwarfs this result somewhat because the PSD are large, as expected even more so for prediction. This causes the 95% credible intervals for both parameters to contain 0. Thus, again when variability is considered, there is no difference between Taft and Fisher.

Table 3
Comparison of the three models for estimation and prediction using the posterior means (PM), posterior standard deviations (PSD), numerical standard errors (NSE) and 95% credible intervals for Λ_e (Λ_p) and Δ_e (Δ_p)

	Model	PM	PSD	NSE	Interval
Λ_e	Ignorable	0.105	0.055	0.002	(-0.002, 0.209)
	Nonignorable	0.097	0.099	0.006	(-0.100, 0.280)
	Time-dependent	0.093	0.101	0.007	(-0.098, 0.276)
Λ_p	Ignorable	0.071	0.154	0.004	(-0.240, 0.362)
	Nonignorable	0.058	0.150	0.005	(-0.252, 0.369)
	Time-dependent	0.050	0.134	0.005	(-0.244, 0.314)
Δ_e	Ignorable	0.688	0.175	0.008	(0.295, 0.958)
	Nonignorable	0.663	0.200	0.012	(0.222, 0.959)
	Time-dependent	0.632	0.148	0.014	(0.336, 0.901)
Δ_p	Ignorable	0.688	0.175	0.008	(0.295, 0.960)
	Nonignorable	0.663	0.193	0.009	(0.253, 0.972)
	Time-dependent	0.648	0.155	0.011	(0.341, 0.923)

NOTE: See note to Table 2; $\Lambda_e = p_{312} - p_{311}$ (estimation, difference between Taft and Fisher for the October poll); $\Lambda_p = p_{412} - p_{411}$ (prediction, difference between Taft and Fisher for the November election); $\Delta_e = \Pr(p_{312} > p_{311} | p_{311} + p_{312} + p_{313}, \boldsymbol{\alpha})$; and $\Delta_p = \Pr(p_{412} > p_{411} | p_{411} + p_{412} + p_{413}, \boldsymbol{\alpha})$; see (26).

We seek an alternative parameter looking to help us predict the winner more convincingly. We pose the following question: "What is the probability that the proportion of Taft's voters in the October poll and the November election is larger than that of Fisher's voters?"

Thus, we consider the parameter $\Delta_e = \Pr(p_{312} > p_{311} | p_{311} + p_{312} + p_{313}, \boldsymbol{\alpha})$ where $\alpha_{jk} = \mu_{jk}\tau_2$, $j = 1, \ldots, r$, $k = 1, \ldots, c$, for the ignorable and nonignorable nonresponse models, and $\alpha_{jk} = p_{2jk}\tau_2$, $j = 1, \ldots, r$, $k = 1, \ldots, c$, for the time-dependent model. In either case, letting $q_1 = p_{311}/p_{31.}$, $q_2 = p_{312}/p_{31.}$, and $q_3 = p_{313}/p_{31.}$ with $p_{31\cdot} = \sum_{k=1}^{3} p_{31k}$ and $\sum_{k=1}^{3} q_k = 1$, it is easy to show that $(q_1, q_2, q_3) \sim \text{Dirichlet}(\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\alpha}_3)$, where $\tilde{\alpha}_1 = \alpha_{11}$, $\tilde{\alpha}_2 = \alpha_{12}$ and $\tilde{\alpha}_3 = \alpha_{13} + \sum_{k=1}^{c} \alpha_{2k}$. Therefore, we have

$$\Delta_e = \Pr(q_2 > q_1 \mid \boldsymbol{\alpha})$$

$$= \int_0^{1/2} \left\{ \int_{q_1}^{1-q_1} \frac{q_1^{\tilde{\alpha}_1^{-1}} q_2^{\tilde{\alpha}_2^{-1}} (1 - q_1 - q_2)^{\tilde{\alpha}_3^{-1}}}{D(\tilde{\alpha}_1, \tilde{\alpha}_2\ \tilde{\alpha}_3)} dq_2 \right\} dq_1.$$

Then, it is easy to show that

$$\Delta_e = 1 - F_{\tilde{\alpha}_1, \tilde{\alpha}_2 + \tilde{\alpha}_3}(1/2) \int_0^{1/2} F_{\tilde{\alpha}_2, \tilde{\alpha}_3}\{q_1/(1-q_1)\}$$

$$\left\{ \frac{q_1^{\tilde{\alpha}_1 - 1}(1-q_1)^{\tilde{\alpha}_2 + \tilde{\alpha}_3 - 1}}{F_{\tilde{\alpha}_1, \tilde{\alpha}_2 + \tilde{\alpha}_3}(1/2) B(\tilde{\alpha}_1, \tilde{\alpha}_2 + \tilde{\alpha}_3)} \right\} dq_1, \quad (26)$$

where

$$F_{\tilde{\alpha}_1, \tilde{\alpha}_2 + \tilde{\alpha}_3}(a) = \int_0^a \frac{x^{\tilde{\alpha}_1 - 1}(1-x)^{\tilde{\alpha}_2 + \tilde{\alpha}_3 - 1}}{B(\tilde{\alpha}_1, \tilde{\alpha}_2 + \tilde{\alpha}_3)} dx$$

and

$$F_{\tilde{\alpha}_2, \tilde{\alpha}_3}(a) = \int_0^a \frac{x^{\tilde{\alpha}_2 - 1}(1-x)^{\tilde{\alpha}_3 - 1}}{B(\tilde{\alpha}_2, \tilde{\alpha}_3)} dx.$$

We note that Δ_e is the probability that Talft received a higher proportion of the votes in the October poll, and Δ_p is the probability that Taft received a higher proportion of the votes in the November election. These parameters can be very useful for estimation (e) and prediction (p). Parameters like Δ_e or Δ_p are difficult to analyze in the non-Bayesian approach such as that of Chen and Stasny (2003); indeed this is a great strength of the Bayesian paradigm.

It is easy to compute (26) using Monte Carlo integration. For each $\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\alpha}_3$, Δ_e $q_1 \sim \text{Beta}(\tilde{\alpha}_1, \tilde{\alpha}_2 + \tilde{\alpha}_3)$ truncated to the $(0, 1/2)$ is used as an importance function. Thus, for each $\tilde{\boldsymbol{\alpha}}^{(h)}$, $h = 1, \ldots, M$, $M = 1{,}000$ from the Metropolis-Hastings sampler (or Gibbs sampler), we can compute $\Delta_e^{(h)}$. *A posteriori* inference about Δ_e is obtained in the standard empirical manner. For prediction, we have also considered $\Delta_p = \Pr(p_{412} > p_{411} | p_{411} + p_{412} + p_{413}, \boldsymbol{\alpha})$, where $\alpha_{jk} = \mu_{jk}\tau_2$, $j = 1, \ldots, r$, $k = 1, \ldots, c$, for the ignorable and nonignorable nonresponse models, and $\alpha_{jk} = p_{3jk}\tau_2$, $j = 1, \ldots, r$, $k = 1, \ldots, c$, for the time-dependent model. Note that Δ_e and Δ_p are the same for the ignorable and nonignorable nonresponse models.

In Table 3 we also present estimates of Δ_e and Λ_p for the three models. First, note again that the NSEs are all small. The estimates of these parameters are similar for the three models, and larger than 0.60, but the 95% credible intervals contain 0.5. Thus, again the posterior means indicate that Taft is the winner, but variation is nullifying the effect of Taft being the winner. We note again that the time-dependent model provides sharper inference, not enough though. The parameters Δ_e and Δ_p are more sensible because they restrict inference to a smaller region by conditioning on $p_{311} + p_{312} + p_{313}$ and $p_{411} + p_{412} + p_{413}$, and from a probabilistic view these parameters are more appropriate.

Finally, we study sensitivity to inference about Λ_p and Δ_p for the nonignorable nonresponse model and the time-dependent model. We do not present results for Λ_e and Δ_e because they are similar to Λ_p and Δ_p. Also, we have dropped the ignorable model as well, and we do not present 95% credible intervals because the posterior densities are roughly symmetric. Our results are presented in Table 4 by model, κ_1 and κ_2. The posterior means of Λ_p and Δ_p are respectively very similar for different values of κ_1 and κ_2. Note that *a priori*

Table 4
Sensitivity of the posterior means (PM) and the posterior standard deviations (PSD) of Λ_p and Δ_p with respect to changes in κ_1 and κ_2 by model

		κ_2							
		1		5		25		50	
Model	κ_1	**PM**	**PSD**	**PM**	**PSD**	**PM**	**PSD**	**PM**	**PSD**
a. Λ_p									
Nonignorable	1	0.058	0.150	0.046	0.153	0.060	0.148	0.054	0.147
	2	0.051	0.153	0.046	0.146	0.062	0.151	0.054	0.145
	3	0.058	0.152	0.059	0.145	0.053	0.149	0.055	0.149
	4	0.055	0.151	0.057	0.148	0.049	0.148	0.043	0.154
Time-dependent	1	0.050	0.134	0.044	0.144	0.048	0.136	0.050	0.130
	2	0.049	0.136	0.052	0.140	0.056	0.129	0.047	0.137
	3	0.039	0.139	0.049	0.137	0.045	0.139	0.052	0.133
	4	0.037	0.138	0.042	0.138	0.041	0.141	0.051	0.129
b. Δ_p									
Nonignorable	1	0.663	0.200	0.650	0.194	0.666	0.186	0.670	0.182
	2	0.663	0.197	0.661	0.188	0.667	0.185	0.659	0.181
	3	0.663	0.199	0.647	0.196	0.666	0.184	0.669	0.180
	4	0.641	0.202	0.668	0.191	0.643	0.197	0.650	0.195
Time-dependent	1	0.648	0.155	0.642	0.123	0.657	0.099	0.661	0.095
	2	0.660	0.151	0.652	0.127	0.659	0.102	0.657	0.099
	3	0.622	0.153	0.636	0.137	0.649	0.120	0.648	0.115
	4	0.610	0.162	0.636	0.152	0.646	0.132	0.644	0.127

NOTE: We have taken $\tau_1 \sim \text{Gamma}\,(1/\kappa_1^2 c_0^2,\ 1/\kappa_2\mu_0\kappa_1^2 c_0^2)$ and we studied sensitivity with respect to $\boldsymbol{\kappa_1}$ and $\boldsymbol{\kappa_2}$. See note to Table 3.

$$\tau_1 \sim \text{Gamma}\left(\frac{1}{\kappa_1^2\, c_0^2},\ \frac{1}{\kappa_2\, \mu_0\, \kappa_1^2\, c_0^2}\right),$$

$E(\tau_1) = \kappa_2\,\mu_0$ and $\text{SD}(\tau_1) = \kappa_1\,\kappa_2\,c_0\,\mu_0$; so clearly, *a priori* $E(\tau_1)$ increases with κ_2 and $\text{SD}(\tau_1)$ increases with either κ_1 or κ_2, but not necessarily *a posteriori*. These changes do not have a lot of effect on inference *a posteriori*. For almost all combinations of κ_1 and κ_2, under the time-dependent model posterior standard deviations of Λ_p are smaller (but not substantially) than under the nonignorable nonresponse model. Under the time-dependent model posterior standard deviations of Δ_p are substantially smaller than under the nonignorable nonresponse model for all combinations of κ_1 and κ_2.

Concluding Remarks

The main contribution in this paper is the construction and analysis of a time-dependent nonignorable nonresponse model and its application to the Ohio polling data. We have done two additional things as well. First, we have compared the time-dependent model with an extended version (to include time) of the ignorable and nonignorable nonresponse models of Nandram, *et al.* (2005). Second, we have constructed a new parameter to help predict the winner; however, this parameter did not make an enormous difference partly because there are only three time points in the time-dependent model.

Our time-dependent model provides posterior inferences that are closer to the truth than the ignorable and nonignorable nonresponse models as well as those of Chen and Stasny (2003). It is natural for voters' preference to change as new information, detrimental or supportive, is revealed into the public place. Thus, our time-dependent model, which takes care of changes over time and provides improved precision, is to be preferred. The uncertainty in the prediction can be reduced in two ways. First, with an increased number of polls there will be increased precision in the parameters, which in turn, can lead to improved prediction. Second, with more prior information (*e.g.*, exit polling) about the November election, one can also improve the prediction.

Our 95% credible intervals can be shortened by using prior information on the proportion of voters going to Taft or Fisher. A referee suggested, "The major-party voting proportions are between 35% and 65% in general elections, and in specific states an objective political scientist could generally provide an even tighter prior." However, this is a complex problem because with truncated prior distributions on the $\boldsymbol{p}$ s, there is a normalization constant which is a function of τ_2. Thus, when τ_2 is drawn from its conditional posterior density, we need to perform a Monte Carlo integration to compute the normalization constant at each iterate. While this will be a useful contribution, we prefer not to pursue this problem here.

The number of days to an election has an important impact on poll accuracy and that this effect can vary substantially across different campaign contexts (*e.g.*, DeSart and Holbrook 2003). Thus, it is really difficult to predict the outcome of an election weeks before it actually occurs, unless there exists an absolute margin. Someone who wishes to predict the outcome of an election must take into consideration additional information near the actual election. Our prediction assumes that there is no catastrophic change near the election; such an abrupt change in public opinion can occur. For example, in 1988 Dukakis lost the election against George Bush for various reasons: he spent the last week in Massachusetts, his cold personality, and Bush's attack on his liberal position. Also, an effective campaign can mobilize undecided voters near the election (*e.g.*, Truman and Dewey in 1948). One way to capture a possible catastrophe is to use mixture distributions or other heavy-tailed distributions (as researchers use Levy distributions in mathematical finance).

Acknowledgement

This work was done while Balgobin Nandram was on sabbatical leave at the National Center for Health Statistics, Hyattsville, Maryland, 2003-2004.

Appendix A

Time-dependent model: Conditional posterior densities of $p_t, t = 0, \ldots, T$

We show how to draw a sample from the conditional posterior density of $\boldsymbol{p}_0$ in (17) using a grid method, and how to draw a sample from the conditional posterior densities of $\boldsymbol{p}_t, t = 1, \ldots, T$ in (18) using Metropolis steps, each with an independence chain.

First, we show how to draw a sample from the conditional posterior density of $\boldsymbol{p}_0$ in (17) using a grid method. Letting $(q_{01}, \ldots, q_{0L}) = (p_{01}, \ldots, p_{0rc})$ and $(q_{11}, \ldots, q_{1L}) = (p_{11}, \ldots, p_{1rc})$ where $L = rc$, with $\sum_{\ell=1}^{L-1} q_{0\ell} \le 1$, we have

$$\pi(q_{01}, \ldots, q_{0L-1} \mid \text{else}, \boldsymbol{y}_1)$$

$$\propto \frac{q_{1L}^{(1-\sum_{\ell=1}^{L-1} q_{0\ell})\tau_2 - 1}}{\Gamma((1-\sum_{\ell=1}^{L-1} q_{0\ell})\tau_2)} \prod_{\ell=1}^{L-1} \frac{q_{1\ell}^{q_{0\ell}\tau_2 - 1}}{\Gamma(q_{0\ell}\tau_2)}, 0 \le q_{0\ell} \le 1, \ell = 1, \ldots, L-1,$$

and it is easy to show that

$$\pi(q_{0\ell} \mid \text{else}, \boldsymbol{y}_1) \propto \frac{q_{1L}^{(1-\sum_{\ell=1}^{L-1} q_{0\ell})\tau_2 - 1}}{\Gamma((1-\sum_{\ell=1}^{L-1} q_{0\ell})\tau_2)} \frac{q_{1\ell}^{q_{0\ell}\tau_2 - 1}}{\Gamma(q_{0\ell}\tau_2)},$$

$$0 \le q_{0\ell} \le 1 - \sum_{\ell'=1,\ \ell' \ne \ell}^{L-1} q_{0\ell'}, \ \ell = 1, \ldots, L-1.$$

For each ℓ we divide the range $0 \le q_{0\ell} \le 1 - \sum_{\ell'=1,\ \ell' \ne \ell}^{L-1} q_{0\ell'}$ into a number of subintervals. To obtain a random deviate $q_{0\ell}$ from its conditional posterior density, we select an interval proportional to its area, and draw a uniform random deviate from this interval.

Second, we show how to draw a sample from the conditional posterior densities of $\boldsymbol{p}_t, t = 1, \ldots, T$ in (18) using Metropolis steps, each with an independence chain. Consider $\boldsymbol{p}_t \mid \boldsymbol{p}_{t-1}, \tau_2, \boldsymbol{y}, t = 1, \ldots, T.$ We use the candidate generating density

$$\boldsymbol{p}_t \mid \boldsymbol{p}_{t-1}, \tau_2, \boldsymbol{y} \sim \text{Dirichlet}(\boldsymbol{a}_t),$$

where

$$a_{tjk} = y_{t \cdot jk} + \tau_2 p_{t-1jk}, \ t = 1, \ldots, T, \ j = 1, \ldots, r, \ k = 1, \ldots, c.$$

Then, the acceptance probability is $A_{s,s+1} = \min(1, \psi_{s+1}/\psi_s)$ where

$$\psi_s = \prod_{j=1}^{r} \prod_{k=1}^{c} p_{t+1jk}^{p_{tjk}^{(s)}\tau_2 - 1} / D(\boldsymbol{p}_t^{(s)}\tau_2).$$

References

Chang, L.C., and Krosnick, J.A. (2001). Improving election forecasting. *Technical Report*, Department of Psychology, The Ohio State University.

Chen, Q.L., and Stasny, E.A. (2003). Handling undecided voters: Using missing data methods in election forecasting. *Technical Report*, Department of Statistics, The Ohio State University.

Chen, T., and Fienberg, S.E. (1974). Two-dimensional contingency tables with both completely and partially cross-classified data. *Biometrics*, 30, 629-642.

DeSart, J., and Holbrook, T.M. (2003). Campaigns, polls, and the states: Assessing the accuracy of statewide presidential trial-heat polls. *Political Research Quarterly*, 56, 431-439.

Draper, D. (1995). Assessment and propagation of model uncertainty (with discussion). *Journal of the Royal Statistical Society*, Series B, 57, 45-97.

Gelman, A. (2006). Prior distribution for variance parameters in hierarchical models. *Bayesian Analysis*, 1, 515-533.

Gelman, A., Roberts, G.O. and Gilks, W.R. (1996), Efficient Metropolis jumping rules. In *Bayesian Statistics*, (Eds., J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith), Oxford, U.K.: Oxford University Press, 599-607.

Ghosh, M., and Meeden, G. (1997). *Bayesian Methods for Finite Population Sampling*. New York: Chapman & Hall.

Kalton, G., and Kasprzyk, D. (1986). The Treatment of Missing Survey Data. *Survey Methodology*, 12, 1-16.

Lavrakas, P.J. (1993). *Telephone Survey Methods*: *Sampling*, *Selection*, *and Supervision*. Newbury Park, CA: Sage Publications.

Little, R.J.A., and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*. 2nd edition, New York: John Wiley & Sons Inc.

Nandram, B. (1998). A Bayesian analysis of the three-stage hierarchical multinomial model. *Journal of Statistical Computation and Simulation*, 61, 97-126.

Nandram, B., and Choi, J.W. (2002a). Hierarchical Bayesian nonresponse models for binary data from small areas with uncertainty about ignorability. *Journal of the American Statistical Association*, 97, 381-388.

Nandram, B., and Choi, J.W. (2002b). A Bayesian analysis of a proportion under nonignorable nonresponse. *Statistics in Medicine*, 21, 1189-1212.

Nandram, B., Cox, L.H. and Choi, J.W. (2005). Bayesian analysis of nonignorable missing categorical data: An application to bone mineral density and family income. *Survey Methodology*, 31, 213-225.

Ritter, C., and Tanner, M.A. (1992). The Gibbs stopper and the Griddy Gibbs sampler. *Journal of the American Statistical Association*, 87, 861-868.

Waksberg, J. (1978). Sampling methods for random digit dialing. *Journal of the American Statistical Association*, 73, 40-46.