The Annals of Applied Statistics
2016, Vol. 10, No. 1, 118–143
DOI: 10.1214/15-AOAS876

BAYESIAN LATENT PATTERN MIXTURE MODELS FOR HANDLING ATTRITION IN PANEL STUDIES WITH REFRESHMENT SAMPLES[1]

BY YAJUAN SI*, JEROME P. REITER† AND D. SUNSHINE HILLYGUS†

University of Wisconsin-Madison and Duke University †*

Many panel studies collect refreshment samples—new, randomly sampled respondents who complete the questionnaire at the same time as a subsequent wave of the panel. With appropriate modeling, these samples can be leveraged to correct inferences for biases caused by nonignorable attrition. We present such a model when the panel includes many categorical survey variables. The model relies on a Bayesian latent pattern mixture model, in which an indicator for attrition and the survey variables are modeled jointly via a latent class model. We allow the multinomial probabilities within classes to depend on the attrition indicator, which offers additional flexibility over standard applications of latent class models. We present results of simulation studies that illustrate the benefits of this flexibility. We apply the model to correct attrition bias in an analysis of data from the 2007–2008 Associated Press/Yahoo News election panel study.

1. Introduction. Many longitudinal or panel surveys, in which the same individuals are interviewed repeatedly at different points in time, suffer from panel attrition. For example, in the American National Election Study, 47% of respondents who completed the first wave in January 2008 failed to complete the follow-up wave in June 2010. Such attrition can result in biased inferences when the attrition generates nonignorable missing data, that is, the reasons for attrition depend on values of unobserved variables [e.g., Behr, Bellgardt and Rendtel (2005), Bhattacharya (2008), Brown (1990), Daniels and Hogan (2008), Diggle and Kenward (1994), Ibrahim, Lipsitz and Chen (1999), Olsen (2005), Scharfstein, Rotnitzky and Robins (1999), Schluchte (1982)].

Unfortunately, it is not possible to determine whether the attrition is ignorable or nonignorable, or the extent to which attrition impacts inferences, using the collected data alone. Consequently, analysts have to rely on strong and generally unverifiable assumptions about the attrition process. Many assume that attrition is a missing at random (MAR) process; for example, MAR assumptions underlie the use of post-stratification to adjust survey weights [e.g., Gelman and Carlin (2001), Henderson, Hillygus and Tompson (2010), Holt and Smith (1979)] and off-the-shelf multiple imputation routines to create completed data sets [e.g., Honaker

Received December 2014; revised August 2015.

[1]Supported in part by NSF Grants SES-10-61241 and SES-11-31897.

Key words and phrases. Panel attrition, refreshment sample, categorical, Dirichlet process, multiple imputation, nonignorable.

and King (2010), Pasek et al. (2009)]. Others allow for specific not missing at random (NMAR) processes, characterizing the attrition with a selection model [Brehm (1993), Hausman and Wise (1979), Kenward (1998), Scharfstein, Rotnitzky and Robins (1999)] or pattern mixture model [Daniels and Hogan (2000), Kenward, Molenberghs and Thijs (2003), Lin, McCulloch and Rosenheck (2004), Little (1993, 1994), Roy (2003), Roy and Daniels (2008)].

Many panel surveys supplement the original panel with refreshment samples. These are cross-sectional, random samples of new respondents given the questionnaire at the same time as a subsequent wave of the panel. For example, refreshment samples are included in the National Educational Longitudinal Study of 1988, which followed a nationally representative sample of 21,500 eighth graders in two year intervals until 2000 and refreshed with cross-sectional samples in 1990 and 1992. Overlapping or rotating panels, in which a new study cohort completes their first wave at the same time a previous cohort completes a second or later wave, offer equivalent information.

Refreshment samples offer information that can be utilized to correct inferences for nonignorable panel attrition [Bartels (1999), Bhattacharya (2008), Deng et al. (2013), Hirano et al. (1998, 2001), Sekhon (2004)]. In particular, analysts can use an additive nonignorable (AN) model, which comprises a model for the survey variables coupled with a selection model for the attrition process [Hirano et al. (1998)]. The selection model must be additive in the variables observed and missing due to attrition so that model parameters are identifiable.

Specifying the models for the survey variables and the attrition indicator can be challenging, even when the data include only a modest number of variables. Consider, for example, a multinomial survey outcome modeled as a function of ten categorical predictors. It is difficult to determine which interaction terms to include in the model, especially in the presence of missing data due to attrition [Erosheva, Fienberg and Junker (2002), Si and Reiter (2013), Vermunt et al. (2008)]. The model specification task is even more complicated when the analyst seeks to model all survey variables jointly, for example, with a log-linear model or sequence of conditional models [e.g., specify $f(a)$, then $f(b \mid a)$, then $f(c \mid a, b)$, and so on]. Joint modeling can be useful when the survey variables suffer from item nonresponse.

Recognizing this, Si, Reiter and Hillygus (2014) propose to use a Dirichlet process mixture of products of multinomial distributions [Dunson and Xing (2009), Si and Reiter (2013)] to model the survey variables. This offers the analyst the potential to capture complex dependencies among variables without selecting interaction effects, as well as to handle item nonresponse among the survey variables. However, for the attrition indicator model, Si, Reiter and Hillygus (2014) use probit regression with only main effects for the survey variables, eschewing the task of selecting interaction effects. While convenient, using a main-effects-only specification makes assumptions about the attrition mechanism that may not be realistic in practice. Furthermore, probit regressions can suffer from the effects of separability and near co-linearity among predictors [Gelman et al. (2008)], which complicates estimation of the AN model.

In this article, we present an alternative approach for leveraging refreshment samples based on Bayesian latent pattern mixture (BLPM) models. We focus on models for categorical variables. The key idea is to use the Dirichlet process mixture of products of multinomial distributions for the survey variables and attrition indicator jointly, thus avoiding specification of an explicit selection model. We note that several other authors [e.g., Lin, McCulloch and Rosenheck (2004), Muthén, Jo and Brown (2003), Roy (2003)] have proposed using mixture models for handling attrition outside of the context of refreshment samples. As we show, the refreshment sample enables us to allow the multinomial vectors within mixture components to depend on attrition indicators, thereby encoding a flexible imputation engine that reduces reliance on conditional independence assumptions.

We were motivated by attrition in the Associated Press/Yahoo 2008 Election Panel (APYN) study, a multi-wave longitudinal survey designed to track the attitudes and opinions of the American public during the 2008 presidential election campaign. The APYN study was the basis of dozens of news stories during the campaign and subsequent academic analyses of the election in the years since. However, the study lost more than one-third of the original sample to attrition by the final wave of data collection, calling into question the accuracy of analyses based on the complete cases. The APYN included a refreshment sample in the final pre-election wave of data collection, which we leverage via the BLPM model to create attrition-adjusted, multiply imputed data sets. We use the multiply imputed data to examine dynamics of public opinion in the 2008 presidential campaign.

The remainder of the article is organized as follows. In Section 2 we introduce the APYN data. In Section 3 we describe pattern mixture models for refreshment samples, including conditions under which model parameters are data-identified. To our knowledge, this is the first description of pattern mixture models in this context. In Section 4 we propose and motivate the BLPM model for refreshment sample contexts. In Section 5 we illustrate properties of the BLPM model with simulation studies. Here, we demonstrate the benefits of allowing the multinomial vectors within mixture components to depend on attrition indicators. In Section 6 we analyze the American electorate in the 2008 presidential election using the BLPM model to account for attrition in the APYN data. Finally, in Section 7 we summarize and discuss future research directions.

2. Description of APYN data. The APYN study included eleven waves of data collection and three refreshment samples spanning the 2008 primary and general U.S. election season. The survey was sampled from the GfK Knowledge Panel, which is one of the nation's only online, probability-based respondent pools designed to be statistically representative of the U.S. population. The respondent pool is recruited via a probability-based sampling method using published sampling frames that cover 96% of the U.S. population. Sampled noninternet households are provided with a laptop and free internet service. Individuals in the respondent pool are then invited to participate in online surveys, such as the APYN panel survey. Surveys from the GfK KnowledgePanel are approved by the Office of Management and Budget for government research and have been used in hundreds of academic publications spanning diverse disciplines, including health and medicine, psychology, social sciences, public policy, and survey and statistical methodology. More information about the survey methodology can be found at http://www.knowledgenetworks.com/ganp/election2008/index.html.

Wave 1 of the APYN was fielded on November 2, 2007, and was completed by 2735 respondents out of 3548 contacted individuals. After the initial wave, these wave 1 respondents were invited to participate in each follow-up wave, even if they failed to respond to the previous one. Consequently, wave-to-wave attrition rates or completion rates vary across the study. Three external refresh cross-sections were also collected: a sample of 697 new respondents in January, 576 new respondents in September, and 464 new respondents in October. Each of the refreshment samples is a random and cross-sectional sample of the GfK respondent pool. Our analysis focuses on wave 1 (November 2007) and the ninth wave with a corresponding refreshment sample (October 2008, the final wave before the election), which we label wave 2 for presentational clarity. As shown in Table 1, of those who completed wave 1, 1011 (37%) respondents failed to complete the October wave. In previous research using the APYN data [Henderson and Hillygus (2011), Henderson, Hillygus and Tompson (2010), Iyengar, Sood and Lelkes (2012), Pasek

TABLE 1

APYN variables from wave 1 (*W*1), *wave* 2 (*W*2) *and refreshment sample* (*Ref*), *with rates of item nonresponse. Item nonresponse arises either from refusals to answer the question* (*respondent proceeded to the next question without giving a response*) *or selection of a "Do not know enough to say" response. We note that* 1011 *of the wave* 1 *participants attrited from the panel by wave* 2, *which could result in attrition bias*

		Item nonresponse counts (%)		
Variable	**Levels**	**W1: 2735**	**W2: 1724**	**Ref: 464**
Obama favorability	2	550 (20.1)	95 (5.5)	20 (4.3)
Party identification (Dem., Rep., Ind.)	3	13 (0.5)		9 (1.9)
Ideology (Lib., Mod., Con.)	3	57 (2.1)		10 (2.2)
Age (18–29, 30–44, 45–59, 60+)	4	0		0
Education (≤HS, Some coll., Coll.)	3	0		0
Race (White, Non-white)	2	0		0
Gender	2	0		0
Income (Ks) (<30, 30–50, 50–75, ≥75)	4	0		0
Married indicator	2	0		0

et al. (2009)], scholars have mostly relied on post-stratification weights to correct for potential panel attrition bias, although Pasek et al. (2009) used standard multiple imputation via `Amelia II` [King et al. (2001)]. Deng et al. (2013) outline the limitations of such approaches—both assume that the attrition is MAR.

The primary outcome of interest in pre-election polls tends to be evaluations of the candidates, as analysts attempt to gauge levels of candidate support within the electorate. Which candidate is most likely to win the election? Who in the electorate supports each side? Because the earliest waves of the APYN took place before the ballot match-up was known—that is, before Obama and McCain had been selected as their party nominees—we focus on Obama favorability (coded as favorable or not). This variable offers exact comparability in question wording across survey waves and is highly correlated with eventual vote choice (the tetrachoric correlation of the items in wave 2 is 0.97). In examining Obama favorability, we consider standard covariates from the voting behavior literature. These include demographic variables (from "Age" to "Marital status" in Table 1) previously shown to be related to candidate evaluations and/or panel attrition [Frankel and Hillygus (2013)].[2] We also consider two relevant political background variables ("Party identification" and "Ideology" in Table 1) that are typically considered time invariant in the context of a single election cycle [Bartels et al. (2011)].

3. Additive pattern mixture models for refreshment samples. Before introducing the BLPM model and analyzing the APYN data, we review the AN model of Hirano et al. (1998) and present a corresponding pattern mixture model formulation. Suppose the data comprise a two-wave panel of N_p individuals with a refreshment sample of N_r new individuals in the second wave. For all $N = N_p + N_r$ individuals, the data include q_0 time-invariant variables $X = (X_1, \dots, X_{q_0})$, such as demographic or frame variables. Let $Y_1 = (Y_{11}, \dots, Y_{1q_1})$ be the q_1 survey variables of interest collected in wave 1. Let $Y_2 = (Y_{21}, \dots, Y_{2q_2})$ be the corresponding q_2 survey variables collected in wave 2. Here, we assume that Y_1 and Y_2 comprise the same variables collected at different waves, although this is not necessary. Among the N_p individuals, $N_{cp} < N_p$ provide at least some data in the second wave, and the remaining $N_{ip} = N_p - N_{cp}$ individuals drop out of the panel. The refreshment sample includes only (X, Y_2); by design, Y_1 are missing for all the individuals in the refreshment sample. In this section, we presume that X, Y_1 in the panel and Y_2 in the refreshment sample are not subject to nonresponse, although we relax this when analyzing the APYN.

For each individual $i = 1, \dots, N$, let $W_i = 1$ if individual i would remain in wave 2 if included in wave 1, and let $W_i = 0$ if individual i would drop out of

[2]Demographic and political profile variables are collected in profile surveys when a panelist joins the KnowledgePanel and are updated continually; thus, they have few missing values for any one study.

TABLE 2

Structure of panel and refreshment samples. Notation for sample sizes in parentheses. The total number of individuals in both data sets is $N = N_p + N_r$

	Time-invariant	**Wave 1**	**Wave 2**
Panel (N_p)	X	Y_1	Y_2, $W = 1$ (N_{cp}) $Y_2 =?$, $W = 0$ (N_{ip})
Refreshment sample (N_r)		$Y_1 =?$	Y_2, $W =?$

wave 2 if included in wave 1. Here, W_i is an indicator of panel attrition conditional on participation in wave 1; it is not an indicator of item or unit nonresponse among individuals in the refreshment sample. We note that W_i is fully observed for all individuals in the panel but is missing for the individuals in the refreshment sample, since individuals in the refreshment sample are not provided the chance to respond in wave 1. Putting it all together, the concatenated data have the structure illustrated in Table 2.

The AN model requires a joint model for $(Y_1, Y_2 \mid X)$ and a selection model for $(W \mid X, Y_1, Y_2)$, that is,

$$(3.1)\qquad \begin{aligned} (Y_1, Y_2) \mid X &\sim f(Y_1, Y_2|X, \Theta), \\ W \mid Y_1, Y_2, X &\sim f(W|X, Y_1, Y_2, \Theta), \end{aligned}$$

where Θ generically represents the parameters for both models. To enable identification, (3.2) must exclude interactions between Y_1 and Y_2.

As an example of an AN model, suppose Y_1 and Y_2 are binary variables and X is empty, as in Hirano et al. (1998). One specification of the additive nonignorable selection model is

$$(3.2)\qquad Y_{i1} \sim \mathrm{Bern}(\pi_1), \qquad \mathrm{logit}(\pi_1) = \alpha_0,$$

$$(3.3)\qquad Y_{i2} \mid Y_{i1} \sim \mathrm{Bern}(\pi_{i2}), \qquad \mathrm{logit}(\pi_{i2}) = \beta_0 + \beta_1 Y_{i1},$$

$$(3.4)\qquad W_i \mid Y_{i1}, Y_{i2} \sim \mathrm{Bern}(\pi_{iW}), \qquad \mathrm{logit}(\pi_{iW}) = \tau_0 + \tau_1 Y_{i1} + \tau_2 Y_{i2}.$$

For a pattern mixture model representation, we require a model for $(W \mid X)$ and for $(Y_1, Y_2 \mid X, W)$, that is,

$$\begin{aligned} W \mid X &\sim f(W|X, \Theta), \\ (Y_1, Y_2) \mid X, W &\sim f(Y_1, Y_2|X, W, \Theta). \end{aligned}$$

Using the basic example, one specification of the additive pattern mixture (APM) model is

$$(3.5)\qquad \begin{aligned} W_i &\sim \mathrm{Bern}(\pi_W), & \mathrm{logit}(\pi_W) &= \omega_0, \\ Y_{i1} \mid W_i &\sim \mathrm{Bern}(\pi_{i1}), & \mathrm{logit}(\pi_{i1}) &= \delta_0 + \delta_1 W_i, \\ Y_{i2} \mid Y_{i1}, W_i &\sim \mathrm{Bern}(\pi_{i2}), & \mathrm{logit}(\pi_{i2}) &= \gamma_0 + \gamma_1 W_i + \gamma_2 Y_{i1}, \end{aligned}$$

which contains as many free parameters as in (3.2)–(3.4) and thus is data-identified. To enable identification, we exclude interactions between Y_1 and W in (3.5). We note that both the AN and APM models can include interactions with X and readily extend to other data types.

4. Bayesian latent pattern mixture models. We now develop an APM model for categorical data with $q = q_0 + q_1 + q_2$ variables. Let $Z = (X, Y_1, Y_2) = (Z_1, \dots, Z_q)$ comprise all potentially collected variables. We order variables so that $j = 1, \dots, q_0$ for X variables, $j = q_0 + 1, \dots, q_0 + q_1$ for Y_1 variables, and $j = q_0 + q_1 + 1, \dots, q$ for Y_2 variables. For $i = 1, \dots, N$ and $j = 1, \dots, q$, without loss of generality, let $Z_{ij} \in \{1, \dots, d_j\}$ denote the level of variable j for unit i, where $d_j \geq 2$ is the total number of levels for variable j.

We specify the pattern mixture model as $f(W)f(Z \mid W)$, including X in the joint distribution of the survey variables. This facilitates imputation of (ignorable) item nonresponse in X, and allows us to take advantage of computationally efficient latent class representations of categorical data. Specifically, we adapt the truncated Dirichlet process mixture of products of multinomial distributions (DPMPM) developed by Dunson and Xing (2009), used previously for multiple imputation of missing cross-sectional data by Si and Reiter (2013). The DPMPM assumes that each individual is a member of a latent class, and that within each class the variables follow independent multinomial distributions. Averaging the multinomial probabilities over the latent classes induces global dependence among the variables.

For $i = 1, \dots, N$, let $s_i \in \{1, \dots, K\}$ indicate the class of individual i, and let $\pi_h = \Pr(s_i = h)$ where $h = 1, \dots, K$. We assume that $\pi = (\pi_1, \dots, \pi_K)$ is the same for all individuals. For $j = q_0 + 1, \dots, q_0 + q_1$, let $\psi_{hjz} = \Pr(Z_{ij} = z | s_i = h)$ be the probability of $Z_{ij} = z$ for any value z given that individual i is in class h. For $j = 1, \dots, q_0$ and $j = q_0 + q_1 + 1, \dots, q$, let $\psi^{(1)}_{hjz} = \Pr(Z_{ij} = z | W_i = 1, s_i = h)$ and $\psi^{(0)}_{hjz} = \Pr(Z_{ij} = z | W_i = 0, s_i = h)$ be the probabilities of $Z_{ij} = z$ for any value z given that individual i is in class h for each value of W_i. The complete-data likelihood for (s_i, W_i, Z_i) in the BLPM is as follows;

$$s_i \mid \pi \sim \text{Multinomial}(\pi_1, \dots, \pi_K), \tag{4.1}$$

$$W_i \mid s_i \sim \text{Bernoulli}(\rho_{s_i}). \tag{4.2}$$

When $j \in \{q_0 + 1, \dots, q_0 + q_1\}$, we have

$$Z_{ij} \mid s_i \sim \text{Multinomial}(\{1, \dots, d_j\}, \psi_{s_i j1}, \dots, \psi_{s_i j d_j}). \tag{4.3}$$

When $j \in \{1, \dots, q_0, q_0 + q_1 + 1, \dots, q\}$, we have

$$Z_{ij} \mid s_i, W_i = 1 \sim \text{Multinomial}(\{1, \dots, d_j\}, \psi^{(1)}_{s_i j1}, \dots, \psi^{(1)}_{s_i j d_j}), \tag{4.4}$$

$$Z_{ij} \mid s_i, W_i = 0 \sim \text{Multinomial}(\{1, \dots, d_j\}, \psi^{(0)}_{s_i j1}, \dots, \psi^{(0)}_{s_i j d_j}). \tag{4.5}$$

The BLPM model is a mixture of pattern mixture models, where

$$f(Z_i, W_i) = \sum_{h=1}^{K} \Pr(s_i = h) f(W_i | s_i = h) f(Z_i | W_i, s_i = h).$$

As in the DPMPM, we assume that $(Z_{q_0+1}, \dots, Z_{q_0+q_1})$, that is, Y_1, follow independent, class-specific multinomial distributions that are also independent of W (and X, Y_2). However, we depart from the DPMPM by letting $(Z_1, \dots, Z_{q_0}, Z_{q_0+q_1+1}, \dots, Z_q)$ follow class-specific, independent multinomial distributions that depend on W. Relaxing the conditional independence between Y_2 and W (i.e., Y_2 is independent of W within any latent class) is possible because of information offered by the refreshment sample. We force Y_1 and W to be independent within latent classes to enable identification, following the strategy outlined in Section 3. We allow X to depend on W within classes to offer additional flexibility for settings where the distributions of X are substantially different across attriters and nonattriters. When this is not the case—the distributions of X are observed for both $W = 1$ and $W = 0$—one can specify the model so that X does not depend on W within classes, thereby reducing the number of parameters to estimate.

For the prior distribution on π, we use the stick-breaking representation of a Dirichlet process prior distribution [Sethuraman (1994)], truncating at large K for computational convenience. In particular, we have

$$\pi_h = V_h \prod_{g<h} (1 - V_g), \tag{4.6}$$

$$V_h \sim \text{Beta}(1, \alpha), \qquad \text{for } h = 1, \dots, K-1, \text{ and } V_K = 1, \tag{4.7}$$

$$\alpha \sim \text{Gamma}(a_\alpha, b_\alpha). \tag{4.8}$$

We use uniform prior distributions on all ψ and ρ parameters. We follow Dunson and Xing (2009) and Si and Reiter (2013) and set $a_\alpha = b_\alpha = 0.25$. Setting $a_\alpha + b_\alpha = 0.5$ represents a small prior sample size and hence vague specification, thereby allowing the data to dominate the cluster allocations. In our simulations and the APYN analyses, results are not sensitive to reasonable default choices of (a_α, b_α). We estimate the model using a blocked Gibbs sampler [Ishwaran and James (2001)]; see the online supplement [Si, Reiter and Hillygus (2016)] for an outline of the algorithm.

We set K to be large enough to help the DPMPM to describe the joint distribution reasonably well yet still offer fast computation. Using an initial proposal for K, say, $K = 20$, analysts can examine the posterior distributions of the number of classes with at least one assigned observation across Markov chain Monte Carlo (MCMC) iterates to diagnose if K is large enough. When there is significant posterior mass at a number of classes equal to K, the analyst should add more classes. The analyst can repeat this diagnostic procedure until finding a suitable K. We note that the posterior predictive distributions used to generate imputations typically are very similar for any sufficiently large K.

The usual truncated DPMPM model is based on (4.1)–(4.8) but requires that $\psi^{(0)}_{hjc_j} = \psi^{(1)}_{hjc_j}$ in (4.4) and (4.5) for all (h, j, c_j). This implies that all Z are independent of W within classes, which may not be the case. The refreshment sample offers information that allows us to relax this assumption, particularly for Y_2. Intuitively speaking, the refreshment sample offers information about $f(Y_2 \mid s)$, and the complete cases in the panel offer information about $f(Y_2 \mid s, W = 1)$. These two distributions identify $f(Y_2 \mid s, W = 0)$. Without the refreshment sample, we do not have information to differentiate $f(Y_2 \mid s, W = 0)$ and $f(Y_2 \mid s, W = 1)$; as a consequence, we are forced to make the unverifiable assumption of conditional independence between Y_2 and W. In Section 5, we present simulation studies that illustrate the biases that can result when falsely assuming the conditional independence assumption.

The model can be used for posterior inference or for multiple imputation. For the latter, analysts select m of the sampled completed data sets after convergence of the Gibbs sampler. These data sets should be spaced sufficiently so as to be approximately independent. This involves thinning the MCMC samples so that the autocorrelations among parameters are close to zero. Multiple imputation inferences then can be based on all N units in the concatenated data. Alternatively, as discussed in Deng et al. (2013), some statistical agencies or data analysts may prefer to disseminate or base inferences on only the original panel after using the refreshment sample for imputing the missing values due to attrition. This might be preferable when combining the original and freshened samples complicates interpretation of sampling weights and design-based inference. Additionally, using only the N_p completed panel cases reduces sensitivity of inferences to the specification of the multiple imputation model, which enters the analysis only for completing Y_2 for the attriters. As pointed out by reviewers of this article, survey-weighted analyses of the multiply imputed data can result in biased estimates of variance [Kim et al. (2006)]. This can result from lack of congeniality [Meng (1994a)] of the imputation model and survey-weighted analysis.

5. Simulation studies. In this section we present results of simulation studies that illustrate the potential of the BLPM model to account for nonignorable attrition. We use two data generation mechanisms: one in which Y_2 and W are not independent within classes, and one in which they are independent within classes. We compare the performance of the BLPM model to the usual DPMPM, a model that assumes Y_2 and W are conditionally independent. In each scenario, we set $N_p = 2000$ and $N_r = 1000$. Each wave includes $q_1 = q_2 = 5$ binary variables; for simplicity, we do not include any X variables. Table 3 displays the values of π and the ψ parameters for each scenario. These designs result in nontrivial dependence structures; for example, we ran Pearson's chi-square tests in the true data sets and rejected independence at the 0.05 significance level for 29 out of the 45 paired combinations among the 10 variables.

TABLE 3

Latent class and marginal probabilities for simulations. The first five ψ parameters correspond to Y_{1j} variables, and the last five ψ parameters correspond to Y_{2j} variables. The columns labeled "marginal" are the weighted averages of ψ over the latent classes

	Y_2 and W not Cond. Ind.				Y_2 and W are Cond. Ind.			
Parameter	$h=1$	$h=2$	$h=3$	**Marginal**	$h=1$	$h=2$	$h=3$	**Marginal**
π	0.4	0.3	0.3	–	0.4	0.3	0.3	–
ρ_h	0.80	0.95	0.60	0.78	0.80	0.95	0.60	0.78
$\psi_{h,1,1}$	0.25	0.55	0.85	0.52	0.25	0.55	0.85	0.52
$\psi_{h,2,1}$	0.20	0.50	0.80	0.47	0.20	0.50	0.80	0.47
$\psi_{h,3,1}$	0.15	0.45	0.75	0.42	0.15	0.45	0.75	0.42
$\psi_{h,4,1}$	0.10	0.40	0.70	0.37	0.10	0.40	0.70	0.37
$\psi_{h,5,1}$	0.05	0.35	0.65	0.32	0.05	0.35	0.65	0.32
$\psi^{(0)}_{h,6,1}, \psi^{(1)}_{h,6,1}$	0.76, 0.38	0.46, 0.58	0.16, 0.78	0.49, 0.56	0.38, 0.38	0.58, 0.58	0.78, 0.78	0.56
$\psi^{(0)}_{h,7,1}, \psi^{(1)}_{h,7,1}$	0.77, 0.41	0.47, 0.61	0.17, 0.81	0.50, 0.59	0.41, 0.41	0.61, 0.61	0.81, 0.81	0.59
$\psi^{(0)}_{h,8,1}, \psi^{(1)}_{h,8,1}$	0.78, 0.44	0.48, 0.64	0.18, 0.84	0.51, 0.62	0.44, 0.44	0.64, 0.64	0.84, 0.84	0.62
$\psi^{(0)}_{h,9,1}, \psi^{(1)}_{h,9,1}$	0.79, 0.47	0.49, 0.67	0.19, 0.87	0.52, 0.65	0.47, 0.47	0.67, 0.67	0.87, 0.87	0.65
$\psi^{(0)}_{h,10,1}, \psi^{(1)}_{h,10,1}$	0.80, 0.50	0.50, 0.70	0.20, 0.90	0.53, 0.68	0.50, 0.50	0.70, 0.70	0.90, 0.90	0.68

In each replication of the simulation, we generate a data set with values of (Z, W) for all $N = 3000$ records; we call this the true data. We delete the values of Y_2 for all records in the panel with $W_i = 0$ and the values of (Y_1, W) for all N_r records in the refreshment sample. The resulting data set has the structure in Table 2 without X. We fit the BLPM and DPMPM models using the Gibbs sampler, imputing Y_2 in the panel when $W = 0$ and (Y_1, W) in the refreshment sample in each MCMC iteration. For each scenario, we run 100 independent replications of the simulation.

To evaluate the potential of the BLPM and DPMPM models to correct for attrition, as well as to compare them with each other, we focus primarily on the completed data estimates of $\Pr(Y_2 = 1)$ in the panel. Let superscript $r = 1, \ldots, 100$ index replications of the simulation, and let superscript $t = 1, \ldots, T$ index MCMC iterations, where T is the number of MCMC iterations used in computation. For all (r, t), and for $i = 1, \ldots, N$ and $j = 1, \ldots, q$, let $z_{ij}^{(rt)}$ be the value of z_{ij} in replication r and MCMC iteration t. Here, if $j > q_1$, $z_{ij}^{(rt)}$ is an observed value for all panel cases with $W_i^{(r)} = 1$ and is an imputed value for all cases with $W_i^{(r)} = 0$. For any variable indexed by $j > q_1$, we compute

$$\bar{z}_j^{(rt)} = \sum_{i=1}^{N_p} I\big(z_{ij}^{(rt)} = 1\big)/N_p, \tilde{z}_j^{(r)} = \text{Median}\big(\bar{z}_j^{(r1)}, \ldots, \bar{z}_j^{(rT)}\big).$$

Let $\bar{z}_j^{(r,\text{true})}$ be the value of $\Pr(Z_j = 1)$ for the panel in the true data associated with replication r. We then compute

$$DIF_j = \left|\sum_{r=1}^{100} \tilde{z}_j^{(r)} \Big/ 100 - \bar{z}_j^{(r,\text{true})}\right|,$$

$$RMSE_j = \left(\sum_{r=1}^{100} \big(\tilde{z}_j^{(r)} - \bar{z}_j^{(r,\text{true})}\big)^2 \Big/ 100\right)^{0.5}.$$

The larger DIF_j and $RMSE_j$, the more inaccurate are the completed-data estimate in the panel. We use only the panel and not the concatenated data to magnify the impact of the models on imputation of the missing data due to nonignorable attrition. We also report values of DIF_j and $RMSE_j$ for the BLPM and DPMPM models for the means of W and Y_1 in the refreshment sample. These are both fully imputed in the models.

For each simulation run, we run MCMC chains for both models with $K = 10$ classes—we obtained very similar results with $K = 20$ and $K = 30$. We run the chains for 20,000 and 30,000 iterations for the BLPM and DPMPM models, respectively, which exploratory runs suggest as sufficient for the chains to converge. We keep every tenth draw from the final 10,000 draws of each chain, leaving $T = 1000$ MCMC draws for inference. To initialize the chains, for all h we set

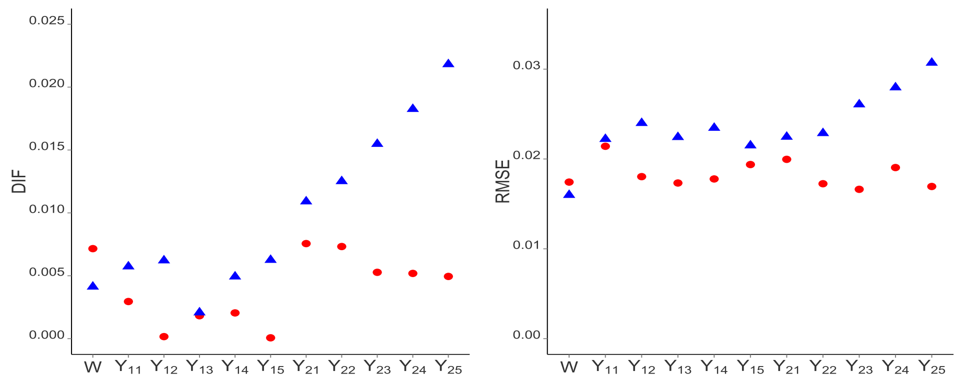

FIG. 1. *Simulation results when the data are generated with Y_2 and W dependent within class. Results for DPMPM displayed with triangles and for BLPM with circles.*

$\rho_h = N_{cp}/N_p$, set all ϕ parameters equal to 0.5, set $\alpha = 1$, and generate all $K - 1$ initial values of V_h from (4.7) using $\alpha = 1$.

Figure 1 summarizes the values of DIF_j and $RMSE_j$ for each quantity for both the BLPM and DPMPM models for the simulation with conditional dependence between Y_2 and W within classes. We also computed the DIF_j and $RMSE_j$ when estimating each $\Pr(Y_{2j} = 1)$ in the panel with only the complete panel cases. For this complete-cases estimator, the average values of DIF and $RMSE$ across the 100 runs are shown in Table 4.

Compared to the results in Table 4, the BLPM and DPMPM tend to offer smaller differences in point estimates, correcting the bias in complete-case analysis due to attrition. When estimating $\Pr(Y_{2j} = 1)$ using the panel data alone, the BLPM tends to be more accurate than the DPMPM. The relative performance of the DPMPM worsens as the magnitude of the attrition bias increases, where by attrition bias we mean the difference in the marginal probabilities of Y_{2j} for nonattriters and attriters, that is, $\sum_h \pi_h \psi_{hj1}^{(1)} - \sum_h \pi_h \psi_{hj1}^{(0)}$. We also tend to see better performance when predicting the missing W and Y_1 in the refreshment sample, although the gaps are not as noticeable as those for Y_2. For all $j > q_1$, the simulated matched pair standard errors are around 0.003 for comparing DIF_j for BLPM and DPMPM, and around 0.005 for comparing DIF_j for BLPM and the complete-case estimator.

TABLE 4
Simulation results for the complete-cases estimator when the data are generated with Y_2 and W dependent within class

$\Pr(Y_{2j} = 1)$	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$
DIF_j	0.031	0.033	0.039	0.042	0.046
$RMSE_j$	0.031	0.033	0.039	0.043	0.047

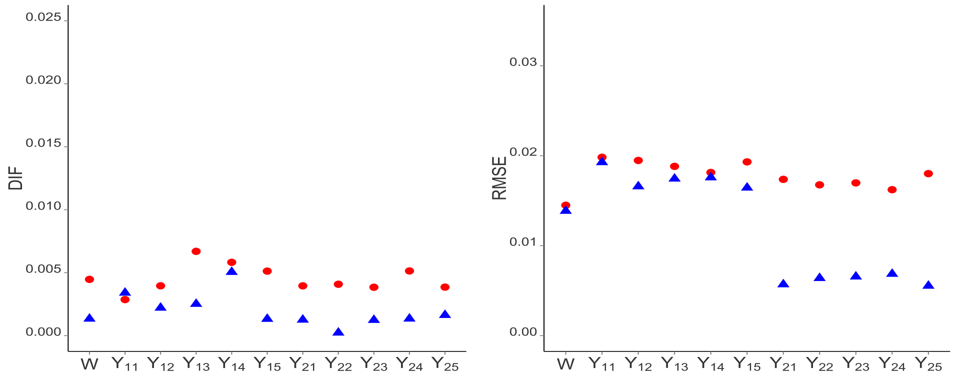

FIG. 2. *Simulation results when the data are generated with Y_2 and W independent within class. Results for DPMPM displayed with triangles and for BLPM with circles.*

Figure 2 summarizes the values of DIF_j and $RMSE_j$ for each quantity for both the BLPM and DPMPM models for the simulation with conditional independence between Y_2 and W within classes. For the complete-cases estimator, across the 100 runs, the average values of $(DIF_1, \dots, DIF_5)$ all equal approximately 0.016 with associated $(RMSE_1, \dots, RMSE_5)$ equal to approximately 0.017. Once again, the BLPM and DPMPM tend to estimate each $\Pr(Y_{2j} = 1)$ using the panel data alone more accurately than the complete-case analysis. When estimating $\Pr(Y_{2j} = 1)$ using the panel data alone, the DPMPM tends to be slightly more accurate than the BLPM, but the differences are modest when compared to those in Figure 1. The differences stem from estimating additional parameters in the BLPM, whereas the DPMPM has the exact specification. For all $j > q_1$, the simulated matched pair standard errors are around 0.002 when comparing DIF_j for BLPM and DPMPM, and 0.002 when comparing DIF_j for BLPM and the complete-case estimator.

In summary, these simulation results suggest that both the BLPM and DPMPM can reduce attrition bias compared to using the complete cases. The BLPM is more flexible than the DPMPM in that it can protect against failure of the conditional independence assumption for Y_2 and W. However, when conditional independence holds, the BLPM estimates can be similar to those based on the DPMPM. A sensible default position with decent sample sizes is to use the BLPM, since the data do not inform whether conditional independence is appropriate.

In our experience, in modest sample sizes both the BLPM and the DPMPM can suffer, as the latent class models will sacrifice higher-order relationships among the variables. Thus, it is crucial to check the fit of the models. We suggest methods for doing so in the analysis of the APYN data (Section 6).

6. Using the BLPM to correct for attrition in the APYN data. We now apply the BLPM model to account for attrition in the APYN data. To begin, we first provide some additional context on the survey design that is relevant for our imputations and analyses. Throughout, we refer to cross-sectional unit nonresponse as nonparticipation or refusal in the wave when an individual is initially surveyed; attrition happens when an individual drops out after participating in a previous wave. For example, the refreshment sample is subject to cross-sectional unit nonresponse but not attrition, as these individuals are only surveyed at wave 2.

6.1. *Survey weights in the APYN.* The APYN data file includes survey weights at each wave. The wave 1 weights are the product of design-based weights and post-stratification adjustments for cross-sectional unit nonresponse at wave 1. These post-stratification adjustments assume the cross-sectional unit nonresponse is missing at random, as is common in the literature [e.g., Bhattacharya (2008), Das, Toepoel and van Soest (2011), Hirano et al. (1998)]. The wave 2 weights for the 1724 panel participants include post-stratification adjustments for attrition in the panel, for cross-sectional unit nonresponse at wave 1 and for cross-sectional unit nonresponse among cases in the refreshment sample; the way that weights are reported does not allow us to disentangle these adjustments. Since we use the BLPM model to account for nonignorable attrition, we disregard the wave 2 weights in all analyses.

The original panel is approximately an equal probability sample, with deviations due primarily to (i) slight oversampling of African American and Hispanic telephone exchanges and (ii) undersampling of areas where the MSN TV service network cannot be used and where there is no access to the internet. The post-strata in wave 1 are based on gender, race, the age groups in Table 1, the education groups in Table 1, census region, metropolitan area and household internet access. We include most of these variables in the BLPM model, thereby accounting for important aspects of the design when making imputations. The geographic variables and internet access are not strong predictors of Obama favorability given all the other variables in Table 1. In a logistic regression with Obama favorability in wave 1 as the dependent variable, a drop in deviance test for the models with and without census region, metropolitan area and internet access (including all other variables in X) results in a p-value of 0.20.[3] Since these variables do not substantially improve our ability to predict the missing Obama favorability values, and are not of substantive interest in our analyses of the American electorate, we exclude them from the imputation model.

We use unweighted analyses to illustrate the attrition effects and describe the behavior of the BLPM model (as in Figures 3 and 4 in Section 6.3), and we use survey weighted analyses when computing finite population quantities (as in Figure 5 in Section 6.3). The survey-weighted estimates account for the sampling design and cross-sectional unit nonresponse in wave 1 only. To make these estimates, we use the wave 1 weights for the 1724 panelists in multiple imputation inferences [Rubin (1987)].

[3]We estimated the model with wave 1 data to avoid any issues from nonignorable attrition.

6.2. *Generating completed data sets*. We run the BLPM with $K = 30$ classes using the Gibbs sampler outlined in the online supplement [Si, Reiter and Hillygus (2016)], treating Obama favorability as (Y_1, Y_2) and all other variables as X. As initial values for W in the refreshment sample, we use independent draws from a Bernoulli distribution with probability $N_{cp}/N_p = 0.63$. For missing data in (X, Y_1, Y_2)—due to item nonresponse and attrition—and W in the refreshment sample, we implement the initialization steps of the MCMC as follows:

- For any missing values in X, sample from the marginal distribution of X computed from the observed cases in the combined panel and the refreshment sample.
- For any missing values in Y_1, sample from the observed marginal distribution of Y_1.
- For missing values in Y_2 in the refreshment sample, sample from the observed marginal distribution of Y_2 in the refreshment sample.
- For missing values in Y_2 in the panel for cases with $W_i = 1$, sample from the observed marginal distribution of Y_2 in the panel.
- For missing values in Y_2 in the panel for cases with $W_i = 0$, sample from independent Bernoulli distributions with probabilities $\Pr(Y_2|W = 0)$, obtained by $[\Pr(Y_2) - \Pr(Y_2|W = 1)\Pr(W = 1)]/\Pr(W = 0)$. Here, $\Pr(Y_2)$ is estimated with the refreshment sample, $\Pr(Y_2|W = 1)$ is estimated with cases with $W_i = 1$ in the panel, and $\Pr(W = 1) = 0.63$.

For the initial values of the parameters, we set $\alpha = 1$; set each $\rho_h = N_{cp}/N_p$; set each ψ parameter equal to the corresponding marginal probability calculated from the initial completed data set; and set $V_h = 0.1$ for $h = 1, \dots, K - 1$. Each record's latent class indicator is initialized from a draw of a multinomial distribution with probability π implied by the set of initial $\{V_h\}$.

We run the MCMC for 150,000 iterations, treating the first 100,000 as burn-in and thinning every 50th iteration. The trace plots of each variable's marginal probability suggest convergence. The posterior mode of the number of distinct occupied classes is 9, and the maximum is 18. This suggests that $K = 30$ classes is sufficient. We collect $m = 50$ completed data sets by keeping every twentieth draw from the $T = 1000$ thinned draws. We use only the N_p records in the completed panels for multiple imputation inferences.

6.3. *Results*. We begin by comparing the distributions of variables in wave 2 among the N_{cp} nonattriters in the panel and the N_r respondents in the refreshment sample; these are summarized in Table 5. Among the nonattriters, 54.9% favor Obama. In the refreshment sample, however, 61.7% favor Obama. This suggests that people who liked Obama may have dropped out with higher frequency than those who did not. As a sense of the magnitude of these differences, the 95% confidence interval limits corresponding to these two percentages are (0.525, 0.573) and (0.572, 0.662), offering evidence that the difference may well be systematic.

TABLE 5

Unweighted percentages of respondents in each category in wave 1 and wave 2 of the panel (W1 and W2), and in the refreshment sample (Ref). Percentages based on available cases only, before imputation of item nonresponse

Variable	W1	W2	Ref.
Favorable to Obama	0.553	0.549	0.617
Democrat	0.327	0.318	0.374
Independent	0.369	0.374	0.312
Liberal	0.223	0.234	0.289
Conservative	0.366	0.370	0.397
Age 18–29	0.148	0.135	0.110
Age 30–44	0.284	0.284	0.213
Age 45–59	0.317	0.320	0.341
HS Edu. or less	0.343	0.325	0.323
College Edu.	0.298	0.333	0.308
Nonwhite	0.230	0.220	0.177
Female	0.548	0.537	0.565
Income <30 K	0.277	0.262	0.170
Income 30–50 K	0.269	0.270	0.306
Income 50–75 K	0.225	0.235	0.211
Married	0.631	0.632	0.647

Of note, compared to the refreshment sample, the N_{cp} nonattriters are less likely to be Democrats and to be liberals, more likely to be nonwhite and to have income below $30,000, and more likely to be below age 45.

These differences in the marginal frequencies reflect the effects of attrition, as well as differential cross-sectional unit nonresponse in the refreshment sample and initial wave. Reassuringly, national cross-sectional polls in October 2008 from Gallup, Fox News and other major polling organizations also put Obama favorability ratings close to 62% (http://www.pollingreport.com/obama_fav.htm), suggesting the respondents in the refreshment sample faithfully represent Obama's favorability ratings at the time. In our analyses, we assume that Obama favorability values missing for reasons other than attrition, that is, due to cross-sectional item and unit nonresponse, are MAR given the variables in the BLPM model. Previous survey methodology research indicates that missingness mechanisms for attrition and cross-sectional nonresponse are distinct [e.g., Groves (2006), Groves and Couper (1998), Loosveldt and Carton (1997), Lynn (2005), Olson and Witt (2011), Smith and Son (2010)], so that one can plausibly consider attrition as potentially nonignorable even when assuming cross-sectional unit nonresponse is MAR. See Schifeling et al. (2015) for further discussion of the effects on inferences of nonignorable cross-sectional unit nonresponse in the initial wave and refreshment sample.

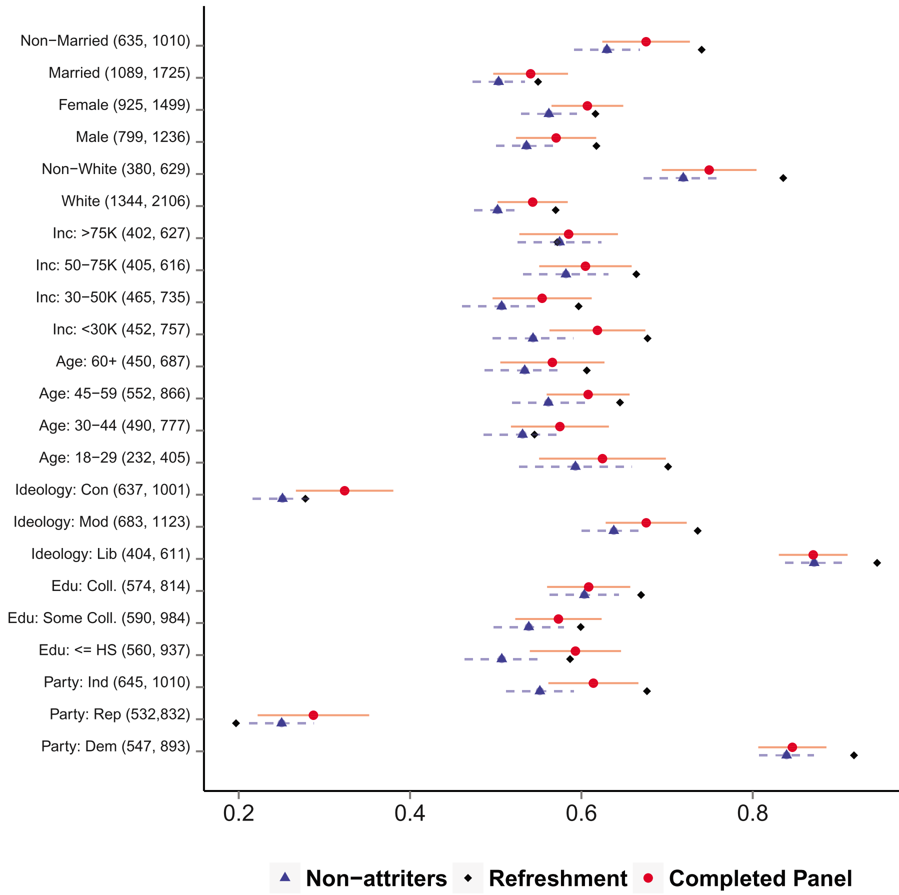

FIG. 3. *Point estimates and 95% confidence intervals for Obama favorability in various subgroups. Results presented for the N_{cp} panel nonattriters, the N_r refreshment samples and the N_p panel participants. Inferences based on unweighted analyses of the $m = 50$ completed data sets, after multiple imputation of missing values via the BLPM model. The numbers in parentheses are the corresponding subgroup sizes, the first being the size among nonattriters and the second being among the completed panel. We randomly select one imputed data set to obtain the sample sizes when the background variables are subject to item nonresponse.*

Figure 3 displays estimated probabilities for Obama favorability for each of the subgroups defined by the time-invariant variables. For many subgroups, the estimates for nonattriters in the panel are noticeably different from those in the refreshment sample. This finding offers an important correction to the prevailing wisdom about the nature of panel attrition in political surveys. Research had previously concluded that attrition bias impacted outcomes related to political engagement (e.g., turnout) but not those related to candidate support (e.g., favorability) [Bartels (1999), Kruse et al. (2009)]. The attrition biases within these subgroups provide evidence to the contrary. It is also noteworthy that the differences are most pronounced for women, low-income respondents, respondents aged 45–59, the least educated and political independents. Many of these are the subpopulations often thought to lack a voice in American politics [Gilens (2005)], and these results suggest that panel attrition may further complicate accurate estimation of their political attitudes and preferences.

Figure 3 also reveals how the BLPM can correct for attrition bias. In particular, for most subgroups, the point estimate for the N_p panel participants is shrunk toward the refreshment sample estimate, that is, the BLPM model corrects the bias due to attrition. The BLPM-corrected intervals tend to be wider than those computed with the nonattriters. This results from two sources of variability, namely, the estimation of the model parameters based on a modest-sized refreshment sample and the imputation of the $N_{ip} = 1011$ values of Y_2.

Figure 4 displays inferences for several smaller subgroups of substantive interest. Here, the BLPM's advantage over AN models is particularly prescient, as we are able to fit the BLPM model without having to specify (perhaps arbitrarily) a selection model with interaction effects. The attrition biases do appear to differ across the groups, suggesting the importance of using models that can capture interaction effects. Interestingly, high-income males appear not to experience substantial attrition bias, whereas various low-income and less educated groups appear to experience sizable underestimations of Obama favorability. As in Figure 3, for most groups the BLPM generally shrinks point estimates toward those in the refreshment sample.

Of course, evaluating potential attrition bias is not the end goal of our analyses. Rather, having created attrition-adjusted imputations with the BLPM model, we now use the m completed panel data sets to better understand the American electorate during the 2008 campaign. Here, we use survey-weighted analysis as follows. For each population percentage of interest and in each of the m completed panel data sets, we compute the standard ratio estimate of the population percentage and the usual estimated variance based on the formula for unequal probability sampling with replacement [Lohr (1999)]. We obtain estimates with the *survey* package [Lumley (2012)] in R. We then combine the point and variance estimates using the multiple imputation rules [Rubin (1987)].

Accounting for the wave 1 survey weights, the marginal estimate for Obama favorability in the last days before Election Day (wave 2) was 0.615 (0.576, 0.655), indicating Obama enjoyed the level of candidate support necessary to win the November election. As can be seen in Figure 5, Obama enjoyed higher levels of favorability among some expected subgroups—liberals, nonwhites and Democrats—in the weighted analysis for both waves. His high levels of favorability among other subgroups, especially moderates and Independents, offers the clearest signal of the likely election outcome. It was only among self-reported Republicans and conservatives that Obama found favorability levels fall below 0.5.

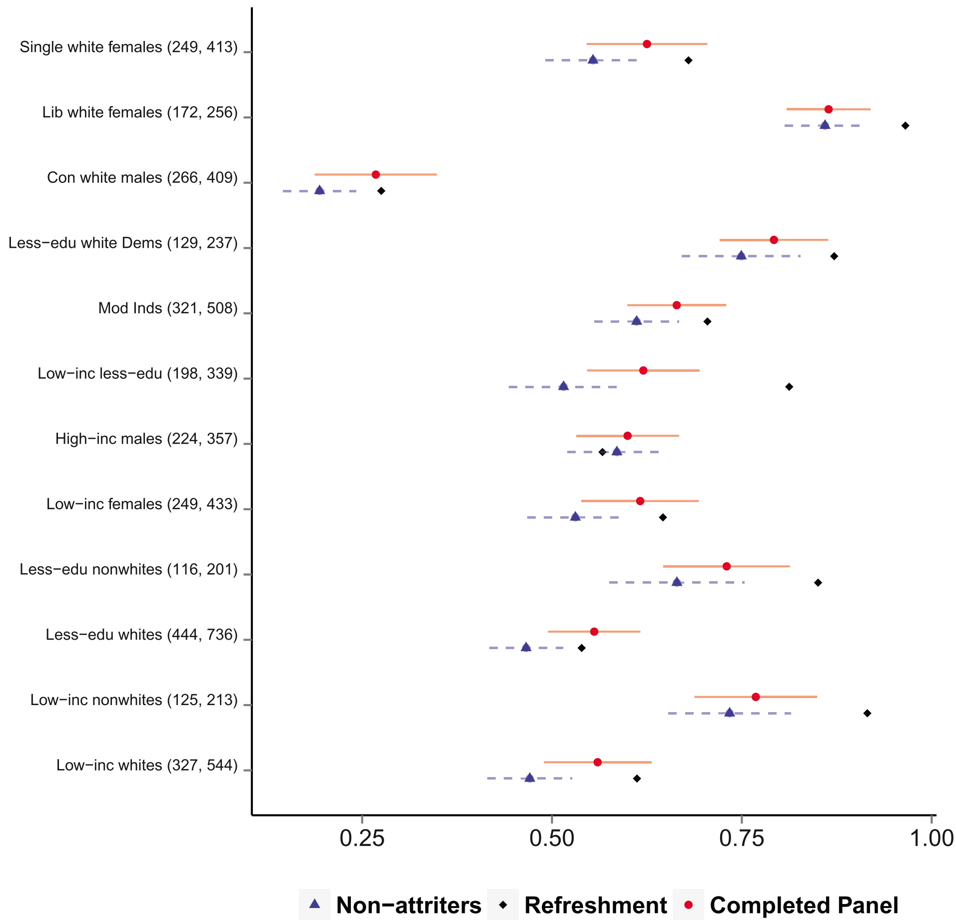

FIG. 4. *Point estimates and 95% confidence intervals for Obama favorability in additional subgroups. Results presented for the N_{cp} panel nonattriters, the N_r refreshment samples and the N_p panel participants. Inferences based on unweighted analyses of the $m = 50$ completed data sets, after multiple imputation of missing values via the BLPM model. The numbers in parentheses are the corresponding subgroup sizes, the first being the size among nonattriters and the second being among the completed panel. We randomly select one imputed data set to obtain the sample sizes when the background variables are subject to item nonresponse.*

Comparing estimates across waves also suggests that the American electorate grew more favorable toward Obama as the campaign unfolded—the average marginal favorability in wave 1 is 0.569 (0.542, 0.597), as illustrated in Figure 5. The increase in marginal favorability rating across waves is 0.046 (0.003, 0.089). In terms of attitude changes during the campaign among the various subgroups, most became slightly more favorable over time, with the exception of conservatives and Republicans who became slightly less favorable from wave 1 and wave 2. Most of these changes are not statistically significant due to sample size issues. The

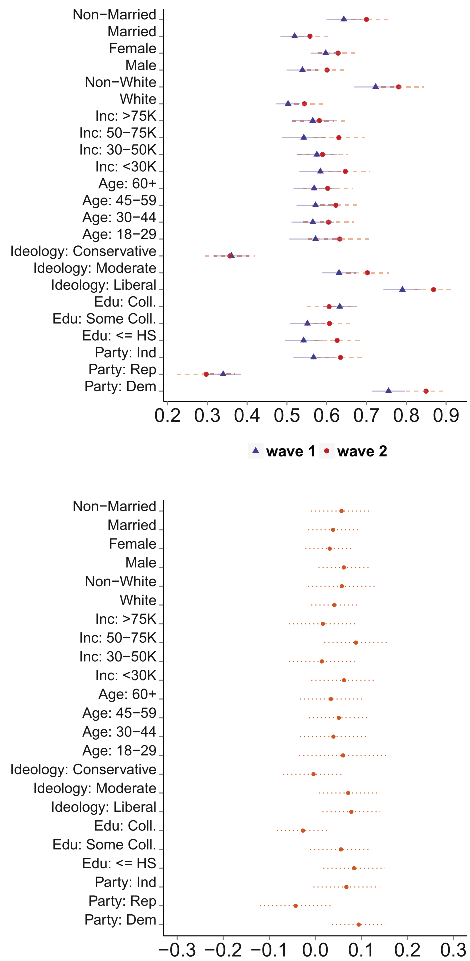

FIG. 5. *Dynamics of Obama favorability ratings between wave* 1 *and wave* 2. *Top plot compares the marginal estimates in wave* 1 *and wave* 2. *Bottom plot presents the differences between wave* 2 *and wave* 1. *Results based on the* N_p *panel participants after multiple imputation via the BLPM model. Inference based on survey-weighted estimation.*

statistically significant changes in attitudes are among Democrats, liberals, moderates, less educated, individuals with middle income, and males, who showed substantial increases in favorability toward Obama between wave 1 and wave 2. Overall, these patterns suggest that the partisan polarization in evaluations of Obama that characterize American politics today actually started during the 2008 presidential campaign [Burden and Hillygus (2009)].

We also fit the BLPM model assuming that Y_1 and X are conditionally independent of W within latent classes. Reassuringly, the conclusions from this version of the BLPM are similar to those presented previously.

For comparison, we fit two additional models: the DPMPM model described in Section 5 that does not have Y_2 depend on W, and a MAR imputation model based on the DPMPM [as in Si and Reiter (2013)] that disregards W entirely. The results for both models, reported in Section 3 of the online supplement [Si, Reiter and Hillygus (2016)], are similar to each other but different from the BLPM results. These two alternative models generally result in point estimates quite similar to those from the nonattriters; in other words, they suggest that panel attrition bias in Obama favorability is ignorable. This seems implausible given the differences in Obama favorability seen in the nonattriters and the refreshment samples.

We also fit the semiparametric AN model of Si, Reiter and Hillygus (2014), which assumes a probit regression for W conditional on (X, Y_1, Y_2) and a DPMPM model for (X, Y_1, Y_2). Results are reported in Section 4 of the online supplement [Si, Reiter and Hillygus (2016)]. Both the semiparametric AN and BLPM models suggest that the attrition is nonignorable. Point estimates for the quantities in Figures 3 and 4 differ slightly; however, the differences are modest relative to the multiple imputation variances. We prefer the BLPM results, as the model diagnostics of Section 6.4 suggest that the BLPM fits the data more effectively than the semiparametric AN model. We further note that the semiparametric AN model is computationally more intensive than the BLPM, as the probit regression for W requires auxiliary data augmentation and Metropolis steps that are not necessary in the BLPM.

6.4. *Model diagnostics*. To check the fit of the models, we follow the advice in Deng et al. (2013) and use posterior predictive checks [Burgette and Reiter (2010), Gelman et al. (2005), He et al. (2010), Meng (1994b)]. We use the BLPM model to generate $T^0 = 500$ data sets with no missing data in (X, Y_1, Y_2, W), randomly sampling from the T=1000 available completed data sets. Let $\{D^{(1)}, \dots, D^{(T^0)}\}$ be the collection of the T^0 completed data sets. For each $D^{(t)}$, we also use the model to generate new values of Y_2 for all cases in the panel, including cases with $W_i = 1$, and in the refreshment sample. This can be done after running the MCMC to convergence as follows. For given draws of parameter values and any item missing data in (X, Y_1), sample new values for the observed and imputed Y_2 using the distributions in the online supplement [Si, Reiter and Hillygus (2016)]. Let $\{R^{(1)}, \dots, R^{(T^0)}\}$ be the collection of the T^0 replicated data sets.

We then compare statistics of interest in $\{R^{(1)}, \dots, R^{(T^0)}\}$ to those in $\{D^{(1)}, \dots, D^{(T^0)}\}$. Specifically, suppose that S is some statistic of interest, such as a marginal or conditional probability in our context. For $t = 1, \dots, T^0$, let $S_{R^{(t)}}$ and $S_{D^{(t)}}$ be the values of S computed from $R^{(t)}$ and $D^{(t)}$, respectively. We compute the two-sided posterior predictive probability,

$$\text{ppp} = \frac{2}{T^0} * \min\left(\sum_{t=1}^{T^0} I(S_{R^{(t)}} - S_{D^{(t)}} > 0), \sum_{t=1}^{T^0} I(S_{D^{(t)}} - S_{R^{(t)}} > 0)\right).$$

When the value ppp is small, for example, less than 5%, this suggests the replicated data sets are systematically different from the observed data set, with respect to that statistic. When the value of ppp is not small, the imputation model generates data that look like the completed data for that statistic. Recognizing the limitations of posterior predictive probabilities [Bayarri and Berger (1998)], we interpret the resulting ppp values as diagnostic tools rather than as evidence from hypothesis tests that the model is "correct."

As statistics, we select $\Pr(Y_2 = 1)$ in the refreshment sample, $\Pr(Y_2 = 1 \mid W = 1)$ in the panel, $\Pr(Y_1 = 1, Y_2 = 1 \mid W = 1)$ in the panel, and $\Pr(Y_2 = 1 \mid X, W = 1)$ in the panel for all conditional probabilities involved in the subgroup analyses in Figures 3 and 4. This results in 38 quantities of interest. A histogram of the 38 values of ppp is displayed in Section 4 in the online supplement [Si, Reiter and Hillygus (2016)]. The analysis does not reveal any serious lack of model fit, as none of the ppp values are below 0.20.

We repeat the same model diagnostics on the semiparametric AN model of Si, Reiter and Hillygus (2014). Many of the posterior predictive probabilities are uncomfortably small. We believe the differences in the semiparametric AN and BLPM models result because the predictor function in the AN model for W used by Si, Reiter and Hillygus (2014) includes only main effects, whereas the BLPM model does not a priori enforce a model for attrition.

7. Concluding remarks. The proposed Bayesian latent pattern mixture model offers a flexible way to leverage the information in refreshment samples in categorical data sets, helping to adjust for bias due to nonignorable attrition. We have used this approach in analyzing the APYN study to better understand the preferences of the American electorate during the 2008 presidential campaign. Our findings suggest that panel attrition biased downward estimates of Obama favorability among many subgroups in the electorate. With a more accurate assessment of voter attitudes, we find that Obama had sufficiently high levels of favorability among key subgroups—independents and moderates—to suggest that the election outcomes were not really in doubt by late October.

The BLPM approach has key advantages over existing applications of additive nonignorable models. The BLPM avoids the difficult tasks of specifying a binary regression model for the attrition process. Unlike standard latent class models, the BLPM fully utilizes the information in the refreshment sample by allowing for conditional dependence within latent classes between wave 2 variables and the attrition indicator. We note that a wide range of existing surveys have data structure amenable to BLPM modeling, including the General Social Survey, the 2008 American National Election Study, the Survey of Income and Program Participation, and the National Educational Longitudinal Study, to name just a few.

As with other modeling strategies for refreshment samples, the validity of the BLPM depends on several overarching assumptions. First, the initial wave of the panel and the refreshment sample should be representative of the same population of interest. Put another way, the units in the target population should not change substantially between wave 1 and wave 2, although certainly the distributions of the substantive variables can do so. Second, any unit (or item) nonresponse other than that due to attrition is missing at random. Third, to ensure identifiability, we assume conditional independence between wave 1 survey variables and the attrition indicator within classes. When this assumption is unreasonable, the BLPM model—and any additive pattern mixture model—could fail to correct for attrition bias. Unfortunately, the data do not provide information about the plausibility of this assumption. Methods for assessing the sensitivity of results to violations of this assumption, as well as to violations of the two representativeness assumptions, are important areas for research.

SUPPLEMENTARY MATERIAL

Bayesian latent pattern mixture models for handling attrition in panel studies with refreshment samples (DOI: 10.1214/15-AOAS876SUPP; .pdf). The supplement includes the MCMC algorithms for the BLPM and DPMPM models, additional analyses of the APYN data using the DPMPM model and semi-parametric AN model, and details of the BLPM model diagnostics.

REFERENCES

BARTELS, L. M. (1999). Panel effects in the American national election studies. *Polit. Anal.* **8** 1–20.

BARTELS, B. L., BOX-STEFFENSMEIER, J. M., SMIDT, C. D. and SMITH, R. M. (2011). The dynamic properties of individual-level party identification in the United States. *Elect. Stud.* **30** 210–222.

BAYARRI, M. J. and BERGER, J. O. (1998). Quantifying surprise in the data and model verification. In *Bayesian Statistics*, 6 (*Alcoceber*, 1998) (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) 53–82. Oxford Univ. Press, New York. MR1723493

BEHR, A., BELLGARDT, E. and RENDTEL, U. (2005). Extent and determinants of panel attrition in the European community household panel. *Eur. Sociol. Rev.* **21** 489–512.

BHATTACHARYA, D. (2008). Inference in panel data models under attrition caused by unobservables. *J. Econometrics* **144** 430–446. MR2436950

BREHM, J. (1993). *The Phantom Respondents*. Univ. Michigan Press, Ann Arbor, MI.

BROWN, C. H. (1990). Protecting against nonrandomly missing data in longitudinal studies. *Biometrics* **46** 143–155.

BURDEN, B. C. and HILLYGUS, D. S. (2009). Polls and elections: Opinion formation, polarization, and presidential reelection. *Pres. Stud. Q.* **39** 619–635.

BURGETTE, L. F. and REITER, J. P. (2010). Multiple imputation for missing data via sequential regression trees. *Am. J. Epidemiol.* **172** 1070–1076.

DANIELS, M. J. and HOGAN, J. W. (2000). Reparameterizing the pattern mixture model for sensitivity analyses under informative dropout. *Biometrics* **56** 1241–1248. MR1815627

DANIELS, M. J. and HOGAN, J. W. (2008). *Missing Data in Longitudinal Studies*. Chapman & Hall/CRC, Boca Raton, FL. MR2459796

DAS, M., TOEPOEL, V. and VAN SOEST, A. (2011). Nonparametric tests of panel conditioning and attrition bias in panel surveys. *Sociol. Methods Res.* **40** 32–56. MR2758298

DENG, Y., HILLYGUS, D. S., REITER, J. P., SI, Y. and ZHENG, S. (2013). Handling attrition in longitudinal studies: The case for refreshment samples. *Statist. Sci.* **28** 238–256. MR3112408

DIGGLE, P. and KENWARD, M. G. (1994). Informative drop-out in longitudinal data analysis. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **43** 49–93.

DUNSON, D. B. and XING, C. (2009). Nonparametric Bayes modeling of multivariate categorical data. *J. Amer. Statist. Assoc.* **104** 1042–1051. MR2562004

EROSHEVA, E. A., FIENBERG, S. E. and JUNKER, B. W. (2002). Alternative statistical models and representations for large sparse multi-dimensional contingency tables. *Ann. Fac. Sci. Toulouse Math.* (6) **11** 485–505. MR2032354

FRANKEL, L. and HILLYGUS, D. S. (2013). Looking beyond demographics: Panel attrition in the ANES and GSS. *Political Analysis* **22** 1–18.

GELMAN, A. and CARLIN, J. B. (2001). Poststratification and weighting adjustments. In *Survey Nonresponse* (R. Groves, D. Dillman, J. Eltinge and R. Little, eds.) Wiley, New York.

GELMAN, A., VAN MECHELEN, I., VERBEKE, G., HEITJAN, D. F. and MEULDERS, M. (2005). Multiple imputation for model checking: Completed-data plots with missing and latent data. *Biometrics* **61** 74–85. MR2135847

GELMAN, A., JAKULIN, A., PITTAU, M. G. and SU, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *Ann. Appl. Stat.* **2** 1360–1383. MR2655663

GILENS, M. (2005). Inequality and democratic responsiveness. *Public Opinion Quarterly* **69** 778–796.

GROVES, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly* **70** 646–675.

GROVES, R. M. and COUPER, M. P. (1998). *Nonresponse in Household Interview Surveys*. Wiley, New York.

HAUSMAN, J. A. and WISE, D. A. (1979). Attrition bias in experimental and panel data: The gary income maintenance experiment. *Econometrica* **47** 455–473.

HE, Y., ZASLAVSKY, A. M., LANDRUM, M. B., HARRINGTON, D. P. and CATALANO, P. (2010). Multiple imputation in a large-scale complex survey: A practical guide. *Stat. Methods Med. Res.* **19** 653–670. MR2744515

HENDERSON, M. and HILLYGUS, D. S. (2011). The dynamics of health care opinion, 2008–2010: Partisanship, self-interest, and racial resentment. *Journal of Health Politics, Policy and Law* **36** 945–960.

HENDERSON, M., HILLYGUS, D. S. and TOMPSON, T. (2010). "Sour grapes" or rational voting? Voter decision making among thwarted primary voters in 2008. *Public Opinion Quarterly*. **74** 499–529.

HIRANO, K., IMBENS, G. W., RIDDER, G. and RUBIN, D. B. (1998). Combining panel data sets with attrition and refreshment samples. Technical Report No. 230, National Bureau of Economic Research, New York.

HIRANO, K., IMBENS, G. W., RIDDER, G. and RUBIN, D. B. (2001). Combining panel data sets with attrition and refreshment samples. *Econometrica* **69** 1645–1659. MR1865224

HOLT, D. and SMITH, T. M. F. (1979). Post stratification. *J. Roy. Statist. Soc. Ser. A* **142** 33–46.

HONAKER, J. and KING, G. (2010). What to do about missing values in time-series cross-section data. *Am. J. Polit. Sci.* **54** 561–581.

IBRAHIM, J. G., LIPSITZ, S. R. and CHEN, M.-H. (1999). Missing covariates in generalized linear models when the missing data mechanism is non-ignorable. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **61** 173–190. MR1664045

ISHWARAN, H. and JAMES, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc.* **96** 161–173. MR1952729

IYENGAR, S., SOOD, G. and LELKES, Y. (2012). Affect, not ideology: A social identity perspective on polarization. *Public Opin. Q.* **76** 405–431.

KENWARD, M. G. (1998). Selection models for repeated measurements with non-random dropout: An illustration of sensitivity. *Stat. Med.* **17** 2723–2732.

KENWARD, M. G., MOLENBERGHS, G. and THIJS, H. (2003). Pattern-mixture models with proper time dependence. *Biometrika* **90** 53–71. MR1966550

KIM, J. K., BRICK, J. M., FULLER, W. A. and KALTON, G. (2006). On the bias of the multiple-imputation variance estimator in survey sampling. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 509–521. MR2278338

KING, G., HONAKER, J., JOSEPH, A. and SCHEVE, K. (2001). Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *Am. Polit. Sci. Rev.* **95** 49–69.

KRUSE, Y., CALLEGARO, M., DENNIS, J. M., SUBIAS, S., LAWRENCE, M., DISOGRA, C. and TOMPSON, T. (2009). Panel conditioning and attrition in the AP-yahoo! News election panel study. In 64*th Conference of the American Association for Public Opinion Research*. Hollywood, FL.

LIN, H., MCCULLOCH, C. E. and ROSENHECK, R. A. (2004). Latent pattern mixture models for informative intermittent missing data in longitudinal studies. *Biometrics* **60** 295–305. MR2066263

LITTLE, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *J. Amer. Statist. Assoc.* **88** 125–134.

LITTLE, R. J. A. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika* **81** 471–483. MR1311091

LOHR, S. L. (1999). *Sampling*: *Design and Analysis*. Duxbury Press, New York.

LOOSVELDT, G. and CARTON, A. (1997). Evaluation of nonresponse in the Belgian election panel study 1991–1995. In *JSM Proceedings*, *Section on Survey Research Methods* 1017–1022. Amer. Statist. Assoc., Anaheim, CA.

LUMLEY, T. (2012). Survey: Analysis of complex survey samples. R package version 3.28-2.

LYNN, P. (2005). A review of methodological research pertinent to longitudinal survey design and data collection. Technical report, Institute for Social and Economic Research, University of Esses, United Kingdom.

MENG, X. L. (1994a). Multiple-imputation inferences with uncongenial sources of input (Disc: P558-573). *Statist. Sci.* **9** 538–558.

MENG, X.-L. (1994b). Posterior predictive p-values. *Ann. Statist.* **22** 1142–1160. MR1311969

MUTHÉN, B., JO, B. and BROWN, C. H. (2003). Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York city [with comment]. *J. Amer. Statist. Assoc.* **98** 311–314.

OLSEN, R. J. (2005). The problem of respondent attrition: Survey methodology is key. *Mon. Labor Rev.* **128** 63–71.

OLSON, K. and WITT, L. (2011). Are we keeping the people who used to stay? Changes in correlates of panel survey attrition over time. *Soc. Sci. Res.* **40** 1037–1050.

PASEK, J., TAHK, A., LELKES, Y., KROSNICK, J. A., PAYNE, B. K., AKHTAR, O. and TOMPSON, T. (2009). Determinants of turnout and candidate choice in the 2008 US presidential election illuminating the impact of racial prejudice and other considerations. *Public Opin. Q.* **73** 943–994.

ROY, J. (2003). Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. *Biometrics* **59** 829–836. MR2025106

ROY, J. and DANIELS, M. J. (2008). A general class of pattern mixture models for nonignorable dropout with many possible dropout times. *Biometrics* **64** 538–545, 668. MR2432424

RUBIN, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley, New York. MR0899519

SCHARFSTEIN, D. O., ROTNITZKY, A. and ROBINS, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J. Amer. Statist. Assoc.* **94** 1096–1146. MR1731478

SCHIFELING, T., CHENG, C., REITER, J. P. and HILLYGUS, D. S. (2015). Accounting for nonignorable unit nonresponse and attrition in panel studies with refreshment samples. *Journal of Survey Statistics and Methodology* **3** 265–295.

SCHLUCHTE, M. D. (1982). Methods for the analysis of informatively censored longitudinal data. *Stat*. *Med*. **11** 1861–1870.

SEKHON, J. (2004). The varying role of voter information across democratic societies. Working paper, Univ. California, Berkeley, Berkeley, CA.

SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statist*. *Sinica* **4** 639–650. MR1309433

SI, Y. and REITER, J. P. (2013). Nonparametric Bayesian multiple imputation for incomplete categorical variables in large-scale assessment surveys. *J*. *Educ*. *Behav*. *Stat*. **38** 499–521.

SI, Y., REITER, J. P. and HILLYGUS, D. S. (2014). Semi-parametric selection models for potentially non-ignorable attrition in panel study with refreshment sample. *Polit*. *Anal*. **23** 92–112.

SI, Y., REITER, J. P. and HILLYGUS, D. S. (2016). Supplement to "Bayesian latent pattern mixture models for handling attrition in panel studies with refreshment samples." DOI:10.1214/15-AOAS876SUPP.

SMITH, T. W. and SON, J. (2010). An analysis of panel attrition and panel change on the 2006–2008 general social survey panel, Technical report, GSS Methodological Report No. 118. Chicago: NORC.

VERMUNT, J. K., GINKEL, J. R. V., DER ARK, L. A. V. and SIJTSMA, K. (2008). Multiple imputation of incomplete categorical data using latent class analysis. *Sociol*. *Method*. **38** 369–397.

Y. SI
DEPARTMENT OF BIOSTATISTICS & MEDICAL INFORMATICS
AND
DEPARTMENT OF POPULATION HEALTH SCIENCES
UNIVERSITY OF WISCONSIN-MADISON
MADISON, WISCONSIN 53726
USA
E-MAIL: ysi@biostat.wisc.edu

J. P. REITER
DEPARTMENT OF STATISTICAL SCIENCE
DUKE UNIVERSITY
DURHAM, NORTH CAROLINA 27708
USA
E-MAIL: jerry@stat.duke.edu

D. S. HILLYGUS
DEPARTMENT OF POLITICAL SCIENCE
DUKE UNIVERSITY
DURHAM, NORTH CAROLINA 27708
USA
E-MAIL: hillygus@duke.edu