The Annals of Applied Statistics
2022, Vol. 16, No. 2, 887–904
https://doi.org/10.1214/21-AOAS1524

COMPUTATIONALLY EFFICIENT BAYESIAN UNIT-LEVEL MODELS FOR NON-GAUSSIAN DATA UNDER INFORMATIVE SAMPLING WITH APPLICATION TO ESTIMATION OF HEALTH INSURANCE COVERAGE

BY PAUL A. PARKER[1,a], SCOTT H. HOLAN[1,2,b] AND RYAN JANICKI[3,c]

[1]*Department of Statistics, University of Missouri,* [a]*paulparker@mail.missouri.edu,* [b]*holans@missouri.edu*
[2]*Office of the Associate Director for Research and Methodology, U.S. Census Bureau*
[3]*Center for Statistical Research and Methodology, U.S. Census Bureau,* [c]*ryan.janicki@census.gov*

Statistical estimates from survey samples have traditionally been obtained via design-based estimators. In many cases these estimators tend to work well for quantities, such as population totals or means, but can fall short as sample sizes become small. In today's "information age," there is a strong demand for more granular estimates. To meet this demand, using a Bayesian pseudolikelihood, we propose a computationally efficient unit-level modeling approach for non-Gaussian data collected under informative sampling designs. Specifically, we focus on binary and multinomial data. Our approach is both multivariate and multiscale, incorporating spatial dependence at the area level. We illustrate our approach through an empirical simulation study and through a motivating application to health insurance estimates, using the American Community Survey.

1. Introduction. An important dichotomy in the realm of small-area estimation is that of area-level vs. unit-level modeling approaches. In general, area-level models use the design-based direct estimate as a response within a statistical model. These models tend to smooth the noisy direct estimates in some fashion and estimate the true latent population value. In contrast to this, unit-level models treat the individual survey respondents as observations in the statistical model. Predictions can then be made for the entire population and aggregated as necessary to produce the desired estimates. As the need for more granular estimates becomes essential, area-level models may perform poorly, due to underlying direct estimates with extremely small or nonexistent sample sizes. Unit-level approaches offer an attractive alternative by modeling the individual survey responses directly rather than smoothing the direct estimators. Although unit-level methodologies offer many advantages over their area-level counterparts, they also face their own set of challenges.

The primary difficulty with modeling survey data at the unit level is the consideration of informative sampling. Many surveys are sampled in an informative manner, whereby there is dependence between the probability of selection and the response of interest. When this relationship is not accounted for, increased bias may be present in the corresponding estimates (Pfeffermann and Sverchkov (2007)). The basic unit-level model, introduced by Battese, Harter and Fuller (1988), assumes that the sample model holds for the entire population, and thus does not account for informative sampling. Parker, Janicki and Holan (2019) review the current methods for addressing the problem of informative sampling. Of primary interest is the pseudo-likelihood (PL) method (Skinner (1989), Binder (1983)), which exponentially weights each unit's likelihood contribution according to the corresponding survey weight. Savitsky and Toth (2016) extend the PL approach to Bayesian settings and provide theoretical justification. Other methods to account for the survey design include modeling the design

Received September 2020; revised April 2021.
Key words and phrases. Bayesian analysis, informative sampling, Pólya Gamma, pseudolikelihood, small area estimation, Small Area Health Insurance Estimates (SAHIE) Program.

variables (Little (2012)), nonlinear regression on the survey weights (Si, Pillai and Gelman (2015), Vandendijck et al. (2016)) as well as specifying a sample model and weight model to find the implied population model (Pfeffermann and Sverchkov (2007)).

Although the problem of informative sampling has been studied in depth, there are other concerns with unit-level modeling that have received considerably less attention. In general, one major difference between area and unit-level approaches is dimensionality. Modeling survey data at the unit level can result in sample sizes that are magnitudes larger than those considered at the area level. Unit-level models are fit to individual survey responses which can number in the millions for large-scale surveys. In contrast, area-level models are typically fit to aggregated survey statistics, such as survey-weighted means, which may number in the thousands. For example, the American Community Survey (ACS) samples 3.5 million households annually which may reasonably fall under the realm of "big data." With these extremely large sample sizes comes computational concerns that must be addressed in order to make unit-level modeling viable. To further exacerbate the problem, many survey variables are non-Gaussian which can lead to nonconjugate full conditional distributions when modeling dependence relationships using traditional Bayesian hierarchical models. Sampling from these posterior distributions can require Metropolis steps that are not efficient and can be cumbersome to tune.

Bradley, Holan and Wikle (2020) introduce a class of conjugate prior distributions that may be used to model dependence for non-Gaussian data in the natural exponential family. This covers important cases, such as Binomial, Multinomial, and Poisson data. Parker, Holan and Janicki (2020) extend this approach to model-count data at the unit level under informative sampling through the use of a PL. Unfortunately, sampling from the full conditional distributions can be difficult under these approaches when observations fall on the boundary of the data (i.e., zero for Poisson data, zero or one for Bernoulli data, etc.). Parker, Holan and Janicki (2020) work around this by using an importance sampling scheme that works well when there are not an excessive number of boundary values (zeroes for Poisson data). However, many surveys contain a multitude of Binomial or Bernoulli random variables which results in an abundance of boundary counts.

There are a number of data augmentation approaches that have been developed to yield conjugate full-conditional distributions for Bernoulli data. Albert and Chib (1993) use latent Gaussian variables in conjunction with a probit link function to model Bernoulli data. More recently, Polson, Scott and Windle (2013) use latent Pólya-Gamma random variables to model binomial data with a logit link function. This approach may also be used to model negative binomial as well as multinomial data.

In this paper we develop methodology to model binomial and multinomial data at the unit level in a computationally efficient manner, while accounting for informative sampling. This is done through the use of Bayesian hierarchical modeling in order to capture various sources of dependence. As previously alluded to, the weights are essential to account for the sampling design, and without them we could end up with significantly biased estimates of the target tabulations. Conversely, depending on the problem, using the pseudolikelihood can still result in a computationally intensive estimation problem. For this reason we develop a variational Bayes approach, based on using fixed weights from the survey provided by the official statistical agency. As such, we consider both a Gibbs sampling approach with fully conjugate full conditional distributions as well as a variational Bayes approach to model fitting.

As a motivating example we consider the problem of estimation of the proportion of people with health insurance at the county level for different income to poverty ratio (IPR) categories. Currently, the Small Area Health Insurance Estimates (SAHIE) program within the U.S. Census Bureau produces estimates of health insurance rates, using an area-level

small area model, fit to direct survey estimates using ACS data (Bauder, Luery and Szelepka (2018)). The model-based estimates produced by SAHIE are the only source of single year health insurance coverage estimates at the county level. While the estimates are generally more precise than the corresponding direct estimates, there are serious modeling challenges with developing area level models for health insurance coverage. First, there are boundary issues in that many of the direct estimates at the county level are exactly equal to either 0 or 1, making use of continuous models impossible. Second, there are policy requirements to benchmark lower-level county estimates to state-level estimates so that users have confidence in the quality of the data. Third, there are multiple within-county estimates that need to be produced, such as health insurance coverage by income level, and accounting for within-county dependencies in an area-level model can be difficult. Finally, the computational requirements of fitting the model used by SAHIE are enormous, due to the complexity of the model and the number of estimates that are produced, despite the fact that an area-level model is used.

The model proposed in this paper eliminates many of these problems. The boundary issues are resolved by using non-Gaussian likelihoods at the unit level. There is no need to benchmark estimates, as the PL produces predictions at the unit level, which can then be aggregated up to any desired geographic level. Spatial and multivariate dependencies are handled through careful specification of the process model. Finally, computational efficiency is achieved through a variational Bayes approximation. This work builds upon Zhang et al. (2014), who use a pseudolikelihood for binary data in a frequentist context. In particular, we extend to the multinomial setting, which allows for categorical response data, as well as to the Bayesian pseudolikelihood which allows for straightforward uncertainty quantification. This paper provides several contributions to the existing literature. Importantly, our unit-level model provides a multiscale approach, bringing in spatial dependence at the area level, while modeling unit-level responses. Also, through the use of our multinomial specification, we are able to seamlessly combine multiple responses into one coherent modeling framework. In terms of computation, we develop a Gibbs sampling approach to model fitting, through the use of Pólya-Gamma data augmentation, building upon Polson, Scott and Windle (2013). Finally, we extend the variational Bayes approach of Durante and Rigon (2019), which is intended for logistic regression, to be used in the case of our pseudo-likelihood mixed model.

The remainder of this paper is organized as follows. Section 2 introduces some necessary background material and then presents our proposed models as well as the methodology used to fit the models. We conduct an empirical simulation study in Section 3. We also provide a data analysis in Section 4 where we estimate the health insurance rate for each county and five different income categories for the entire continental U.S. Finally, we provide concluding remarks and discussion in Section 5. Although the data used herein is confidential microdata, we provide code and an example using ACS public-use microdata at https://github.com/paparker/Unit_Level_Non-Gaussian as well as in the Supplementary Material (Parker, Holan and Janicki (2022)).

2. Methodology. Let $\mathcal{U} = \{1, \ldots, N\}$ be an enumeration of a finite population of interest. Suppose the finite population, $\mathcal{U}$, can be represented as the union of m nonoverlapping subpopulations or small areas, $\mathcal{U}_j = \{1, \ldots, N_j\}$, where $\sum_{j=1}^{m} N_j = N$, and $j \in \{1, \ldots, m\}$ indexes the small areas. Associated to each unit $i \in U_j$ is a characteristic of interest, Z_{ij}, and a vector, $\boldsymbol{x}_{ij}$, of auxiliary information.

A sample, $\mathcal{S} \subset \mathcal{U}$, is selected from the finite population according to a known sampling design. Let $\pi_i = P(i \in \mathcal{S})$ be the sample inclusion probability for unit i in the finite population, and let $w_i = 1/\pi_i$ be the survey weight. A typical inferential goal is estimation of the finite population means

$$\bar{Z}_j = \frac{1}{N_j} \sum_{i=1}^{N_j} Z_{ij} \tag{1}$$

from the observed survey responses. The Horvitz–Thompson estimator (Horvitz and Thompson (1952))

$$\hat{\bar{Z}}_j = \frac{1}{N_j} \sum_{i \in \mathcal{S}_j} w_{ij} Z_{ij}, \tag{2}$$

where $\mathcal{S}_j = \mathcal{S} \cap \mathcal{U}_j$ is a design-unbiased and design-consistent estimator of the finite population mean, $\bar{Z}_j$. We refer to any estimate, which only used the observed survey data, such as (2), as a *direct estimate*.

Let n be the total number of sampled units, and let n_j be the number of sampled units in $\mathcal{S}_j$. In many surveys the overall sample size, n, and many of the area-specific sample sizes, n_j, are large. For these areas the large-sample properties of the Horvitz–Thompson estimator guarantee that (2) will be a precise estimator of the finite population mean (1). However, it is also often the case that, for many of the small areas of interest, n_j will be too small for (2) to be reliable. In such situations, precision can be increased by using models for the survey data which incorporate auxiliary information to "borrow strength" by relating the different small areas and increasing the effective sample sizes.

Models for small area estimation (SAE) often include area-level random effects in order to link the small areas and incorporate spatial dependence. These random effects are typically modeled using a latent Gaussian process (LGP), and Bayesian hierarchical modeling is a common technique used to fit these models. This may be computationally efficient when considering a Gaussian response, as it leads to conjugate full conditional distributions; however, when the data model (likelihood) is non-Gaussian, sampling from the posterior distribution can become difficult, as it may require the use of Metropolis type steps. These sampling mechanisms require tuning that can become unwieldy especially in high-dimensional situations.

Polson, Scott and Windle (2013) use a data augmentation scheme to allow for conjugate sampling under logistic likelihoods. Importantly, this includes both Bernoulli and multinomial responses, which is useful, as binary and categorical data are two often observed types of non-Gaussian survey data. This class also includes the negative-binomial distribution which may be used to model count data.

Specifically, Polson, Scott and Windle (2013) define a random variable X to have a Pólya-Gamma distribution with parameters $b > 0$ and $c \in \mathcal{R}$, denoted $\text{PG}(b, c)$, if X is equal in distribution to

$$\frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{g_k}{(k - 1/2)^2 + c^2/(4\pi^2)},$$

where $g_k \overset{\text{ind}}{\sim} \text{Gamma}(b, 1)$. Furthermore, they show that

$$\frac{(e^{\psi})^a}{(1 + e^{\psi})^b} = 2^{-b} e^{\kappa \psi} \int_0^{\infty} e^{-\omega \psi^2/2} p(\omega)\, d\omega, \tag{3}$$

where $\kappa = a - b/2$ and $p(\omega)$ is a $\text{PG}(b, 0)$ density. They also show that $(\omega|\psi) \sim \text{PG}(b, \psi)$. Thus, with a binomial likelihood, using this data augmentation scheme and Gaussian prior distributions, one can sample from Gaussian full conditional distributions for the parameters, and Pólya-Gamma distributions for the latent augmentation variables. The `BayesLogit` package in `R` provides efficient sampling of Pólya-Gamma random variables (Windle, Polson and Scott (2013))

2.1. *Pseudolikelihoods*. One of the main difficulties when implementing unit-level models for survey data is accounting for an informative sampling design. For example, certain demographic subgroups may be sampled with higher probability, but there may also be a relationship between these subgroups and the response variable of interest. Under this scenario the sample is not representative of the population, and thus the sample likelihood should be adjusted to account for this. Parker, Janicki and Holan (2019) give a review of modern methods for unit-level modeling under informative sampling. One general approach is to use a pseudolikelihood, introduced by Skinner (1989) and Binder (1983), by weighting each unit's likelihood contribution using the reported survey weight w_i,

$$\prod_{i\in\mathcal{S}} f(Z_i \mid \boldsymbol{\theta})^{w_i}, \tag{4}$$

where $\mathcal{S}$ indicates the sample and Z_i represents the response value for unit i.

The PL can be maximized, using maximum-likelihood techniques; however, Savitsky and Toth (2016) show that a PL may also be used in a Bayesian setting, thus generating a pseudo-posterior distribution

$$\hat{\pi}(\boldsymbol{\theta}|\mathbf{Z}, \tilde{\mathbf{w}}) \propto \left\{\prod_{i\in\mathcal{S}} f(Z_i|\boldsymbol{\theta})^{\tilde{w}_i}\right\}\pi(\boldsymbol{\theta}).$$

They emphasize the importance of scaling the weights to sum to the sample size, $\tilde{w}_i = n\frac{w_i}{\sum_{j=1}^{n} w_j}$, in order to prevent contraction of the PL and achieve appropriate variance estimates.

Using a unit-level model such as this, it is simple to generate predictions for any unobserved units, thereby effectively generating the population. It is then straightforward to aggregate units in order to estimate any finite population quantities, such as for SAE purposes. Under a Bayesian framework this can be done for each sample from the posterior distribution, thus yielding a posterior distribution over any desired estimates. In the special case where all covariates are categorical in nature, this approach can be seen as a type of poststratification (see Gelman and Little (1997) and Park, Gelman and Bafumi (2006) for examples of poststratification outside of a pseudo-likelihood framework). For special cases where the poststratification variables include all survey design variables, poststratification alone may be used to account for the sample. However, this is typically not the case for complex survey designs; thus the pseudolikelihood may be used in conjunction with poststratification. Zhang et al. (2014) provide an example of a pseudolikelihood and poststratification combination for small area estimates in a frequentist framework, whereas Parker, Holan and Janicki (2020) take a Bayesian pseudolikelihood and poststratification approach.

Now, an unweighted binomial likelihood has the form

$$\prod_{i\in\mathcal{S}} \frac{(e^{\psi_i})^{Z_i}}{(1+e^{\psi_i})^{n_i}}.$$

By using a pseudolikelihood instead, the form becomes

$$\prod_{i\in\mathcal{S}}\left(\frac{(e^{\psi_i})^{Z_i}}{(1+e^{\psi_i})^{n_i}}\right)^{\tilde{w}_i} = \prod_{i\in\mathcal{S}} \frac{(e^{\psi_i})^{Z_i^*}}{(1+e^{\psi_i})^{n_i^*}}, \tag{5}$$

where $Z_i^* = Z_i \times \tilde{w}_i$ and $n_i^* = n_i \times \tilde{w}_i$. The PL given by (5) is of the same form as that given in (3); thus, we are able to sample from conjugate full conditional distributions, using a binomial type PL with Gaussian prior distributions, and PG data augmentation variables.

2.2. *Binomial response model.* Using the Pólya-Gamma data augmentation scheme, we develop a computationally efficient pseudo-likelihood mixed model for binomial survey data (PL-MB) under informative sampling,

$$
\begin{aligned}
\boldsymbol{Z}|\boldsymbol{\beta},\boldsymbol{\eta} &\propto \prod_{i\in S} \mathrm{Bin}(Z_i|n_i,p_i)^{\widetilde{w}_i},\\
\mathrm{logit}(p_i) &= \boldsymbol{x}_i'\boldsymbol{\beta}+\boldsymbol{\phi}_i'\boldsymbol{\eta},\\
\boldsymbol{\eta}|\sigma_\eta^2 &\sim \mathrm{N}_r(\boldsymbol{0}_r,\sigma_\eta^2\boldsymbol{I}_r),\\
\boldsymbol{\beta} &\sim \mathrm{N}_q(\boldsymbol{0}_q,\sigma_\beta^2\boldsymbol{I}_q),\\
\sigma_\eta^2 &\sim \mathrm{IG}(a,b),\\
\sigma_\beta, a, b &> 0,
\end{aligned}
\tag{6}
$$

where Z_i represents the response for unit $i \in \mathcal{S}$. We model the data using a binomial pseudolikelihood, with n_i representing the number of trials and p_i representing the probability of a positive response (e.g., a unit having health insurance) for unit i. In many survey data scenarios, including those explored here, the data is binary; thus, $n_i = 1, \forall i$. The vector $\boldsymbol{x}_i'$ represents a q-dimensional set of covariates, and $\boldsymbol{\beta}$ is the q-dimensional vector of fixed effects. In this work the vector $\boldsymbol{\phi}_i'$ represents either an r-dimensional vector of spatial basis functions or an incidence vector, indicating in which area unit i resides. In this way the r-dimensional vector $\boldsymbol{\eta}$ acts as area-level random effects. Note that the binomial pseudolikelihood can be rewritten using (3). Although we do not present the model this way for the sake of readability, we take advantage of this fact when we construct the Gibbs sampling scheme which introduces a step to sample the latent Pólya-Gamma random variables. The full conditional distributions for Gibbs sampling, which rely on the Pólya-Gamma data augmentation, can be found in Appendix A. As an alternative to Gibbs sampling, for manageable sample sizes Hamiltonian Monte Carlo could be used, for example via Stan (Stan Development Team (2021)).

2.3. *Variational Bayes approximation.* In many high-dimensional settings it can become a computational burden to sample from the posterior distribution via MCMC, even through the use of Gibbs sampling with fully conjugate full conditional distributions. For example, using the Pólya-Gamma data augmentation scheme, a latent random variable must be drawn for every sample observation at every iteration of the MCMC. As sample sizes become very large, this may become infeasible, even after allowing for parallel computing techniques. One popular solution to this computational problem is the variational Bayes approach (Jordan et al. (1999), Wainwright et al. (2008)) for which an approximation to the posterior distribution is used rather than the true posterior distribution. A class of distributions, $\mathcal{D}$, is chosen for $q^*(\boldsymbol{\theta})$, the approximation to the true posterior, $p(\boldsymbol{\theta}|\boldsymbol{x})$. Optimization techniques may then be used to minimize the Kullback–Leibler (KL) divergence between the approximate and true posterior distributions,

$$
q^*(\boldsymbol{\theta}) = \underset{q(\boldsymbol{\theta})\in\mathcal{D}}{\arg\min}\,\mathrm{KL}\big(q(\boldsymbol{\theta})||p(\boldsymbol{\theta}|\boldsymbol{x})\big). \tag{7}
$$

Beal and Ghahramani (2003) focus on a specific case known as the variational Bayes EM algorithm. The approximating distribution can be factored into a product of global parameters and local latent variables, $q(\boldsymbol{\theta}) = q(\boldsymbol{\beta})\prod_{i=1}^n q(\xi_i)$. With this factorization an iterative approach can be used to minimize the KL divergence, where

$$
\begin{aligned}
q(\boldsymbol{\beta})^{(t)} &\propto \exp\{\mathbb{E}_{q^{(t-1)}(\boldsymbol{\xi})}\log[p(\boldsymbol{\beta}|\boldsymbol{Z},\boldsymbol{\xi})]\},\\
q(\xi_i)^{(t)} &\propto \exp\{\mathbb{E}_{q^{(t-1)}(\boldsymbol{\beta})}\log[p(\xi_i|\boldsymbol{Z},\boldsymbol{\xi}_{-i},\boldsymbol{\beta})]\}, \quad i=1,\ldots,n.
\end{aligned}
\tag{8}
$$

Algorithm 1: VB EM algorithm for PL-MB model

Initialize $\tilde{\sigma}_\eta^2$ and $\tilde{\xi}_i,\ i = 1, \ldots, n$;
Let $\boldsymbol{D} = [\boldsymbol{X}, \boldsymbol{\Phi}]$ and $\boldsymbol{\zeta} = (\boldsymbol{\beta}', \boldsymbol{\eta}')$;
for $t = 1$ *until convergence* **do**
  $\tilde{\boldsymbol{\Omega}} = \text{Diag}(\frac{w_1}{2\tilde{\xi}_1}\tanh(\tilde{\xi}_1/2), \ldots, \frac{w_n}{2\tilde{\xi}_n}\tanh(\tilde{\xi}_n/2))$;
  $\tilde{\boldsymbol{\Sigma}} = (\text{blockdiag}(\frac{1}{\sigma_\beta^2}\boldsymbol{I}_p, \frac{a+r/2}{\tilde{\sigma}_\eta^2}\boldsymbol{I}_r) + \boldsymbol{D}'\tilde{\boldsymbol{\Omega}}\boldsymbol{D})^{-1}$;
  $\tilde{\boldsymbol{\Sigma}}_\eta = \tilde{\boldsymbol{\Sigma}}[(p+1):(p+r), (p+1):(p+r)]$;
  $\tilde{\boldsymbol{\mu}} = (\tilde{\boldsymbol{\mu}}_\beta', \tilde{\boldsymbol{\mu}}_\eta')' = \tilde{\boldsymbol{\Sigma}}\boldsymbol{D}'(\boldsymbol{w} \odot (\boldsymbol{Z} - 1/2))$;
  $\tilde{\sigma}_\eta^2 = b + \frac{1}{2}(\tilde{\boldsymbol{\mu}}_\eta'\tilde{\boldsymbol{\mu}}_\eta + \text{tr}(\tilde{\boldsymbol{\Sigma}}_\eta))$;
  for $i = 1$ *to* n **do**
    $\tilde{\xi}_i = (\boldsymbol{D}_i'\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}_i + (\boldsymbol{D}_i'\tilde{\boldsymbol{\mu}})^2)^{1/2}$;
  end
end

In models that use fully conjugate full conditional distributions as well as likelihoods from the exponential family, these factorized approximate distributions are of the same class as their corresponding full conditional distribution. Importantly, this includes the case of logistic regression via Pólya-Gamma data augmentation for which Durante and Rigon (2019) explore a variational Bayes EM algorithm approach.

Algorithm 1 provides an extension of the one explored by Durante and Rigon (2019), which is intended for unweighted logistic regression, and thus not directly applicable to our pseudo-likelihood mixed model. The main extension of this algorithm is the inclusion of the pseudolikelihood rather than the original binomial likelihood. This algorithm may be used in place of MCMC in order to fit the PL-MB model in high-dimensional settings. Independent samples from the variational approximation to the posterior of $\boldsymbol{\zeta} = (\boldsymbol{\beta}', \boldsymbol{\eta}')$ may be drawn by sampling from a $\text{N}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}})$ distribution which may then be used to produce any desired Monte Carlo estimates. We give more details about prediction using poststratification with the variational distribution in Appendix B.

2.4. *Multinomial response model*. In addition to binomial data, multinomial or categorical data is often observed in survey data. In a similar fashion as the PL-MB model, we can write the pseudo-likelihood mixed effect multinomial model (PL-MM) with K categories as

$$
\begin{aligned}
\boldsymbol{Z}|\boldsymbol{\beta}, \boldsymbol{\eta} &\propto \prod_{i \in S} \text{multinomial}(\boldsymbol{Z_i}|n_i, \boldsymbol{p_i})^{\tilde{w}_i}, \\
p_{ik} &= \frac{\exp(\psi_{ik})}{\sum_{k=1}^K \exp(\psi_{ik})}, \\
\psi_{ik} &= \boldsymbol{x_i'}\boldsymbol{\beta_k} + \boldsymbol{\phi_i'}\boldsymbol{\eta_k}, \\
\boldsymbol{\eta_k}|\sigma_{\eta k}^2 &\sim \text{N}_r(\boldsymbol{0_r}, \sigma_{\eta k}^2\boldsymbol{I}_r), \quad k = 1, \ldots, K-1, \\
\boldsymbol{\beta_k} &\sim \text{N}_p(\boldsymbol{0_p}, \sigma_\beta^2\boldsymbol{I}_p), \quad k = 1, \ldots, K-1, \\
\sigma_{\eta k}^2 &\sim \text{IG}(a, b), \quad k = 1, \ldots, K-1, \\
\sigma_\beta, a, b &> 0,
\end{aligned} \tag{9}
$$

where $\boldsymbol{\beta_K}$ and $\boldsymbol{\eta_K}$ are constrained to be equal to zero for identifiability. The K-dimensional vector $\boldsymbol{Z_i}$ represents the number of successful outcomes in each of the K categories for survey unit i, and the K-dimensional vector $\boldsymbol{p}_i$ represents the probability of each category for unit i.

Although Algorithm 1 is intended for binomial data, a stick-breaking representation of the multinomial distribution can be used to expand the applicability of this VB approach. Specifically, Linderman, Johnson and Adams (2015) show that the multinomial distribution may be written as a product of independent binomial distributions,

$$\text{multinomial}(\boldsymbol{Z}|n, \boldsymbol{p}) = \prod_{k=1}^{K-1} \text{Bin}(Z_k|n_k, \tilde{p}_k), \tag{10}$$

where

$$n_k = n - \sum_{j<k} Z_j, \qquad \tilde{p}_k = \frac{p_k}{1 - \sum_{j<k} p_j}, \quad k = 2, \ldots, K. \tag{11}$$

Under this view of multinomial data, we can rewrite the PL-MM model as

$$\begin{aligned}
\boldsymbol{Z}|\boldsymbol{\beta}, \boldsymbol{\eta} &\propto \prod_{i \in S} \prod_{k=1}^{K-1} \text{Bin}(Z_{ik}|n_{ik}, \tilde{p}_{ik})^{\tilde{w}_i}, \\
\text{logit}(\tilde{p}_{ik}) &= \boldsymbol{x_i'}\boldsymbol{\beta_k} + \boldsymbol{\phi_i'}\boldsymbol{\eta_k}, \\
\boldsymbol{\eta_k}|\sigma^2_{\eta k} &\sim \text{N}_r(\boldsymbol{0_r}, \sigma^2_{\eta k}\boldsymbol{I}_r), \quad k = 1, \ldots, K-1, \\
\boldsymbol{\beta_k} &\sim \text{N}_p(\boldsymbol{0_p}, \sigma^2_{\beta}\boldsymbol{I}_p), \quad k = 1, \ldots, K-1, \\
\sigma^2_{\eta k} &\sim \text{IG}(a, b), \quad k = 1, \ldots, K-1, \\
\sigma_\beta, a, b &> 0,
\end{aligned} \tag{12}$$

where $n_{ik} = n_i - \sum_{j<k} Z_{ij}$ and $\tilde{p}_{ik} = \frac{p_{ik}}{1-\sum_{j<k} p_{ij}}$, $k = 2, \ldots, K$. Thus, the PL-MM model may be fit as a series of $K-1$ independent binomial models, using either MCMC or the VB approach outlined in Algorithm 1. Note that, after fitting the model, the stick breaking probabilities $\tilde{\boldsymbol{p}}_i$ can be transformed back to the original probabilities $\boldsymbol{p}_i$ for inference.

3. Empirical simulation study. In order to mimic a real survey data setting, our simulations revolve around resampling of an existing survey dataset rather than generating a synthetic population from a parametric distribution. Specifically, we treat the existing survey sample as our population and then take a further sample with probability proportional to s_i, a size variable that is constructed in an informative manner. This informative sampling scheme can be validated by comparing the weighted design-based estimator to an unweighted design-based estimator. Under an informative design the unweighted estimator will result in greater bias.

3.1. *Multinomial response simulation.* An important SAE application is the Small Area Health Insurance Estimation (SAHIE) program (Bauder, Luery and Szelepka (2018)). The goal of SAHIE is to estimate the proportion of individuals with health insurance by county for a number of income to poverty ratio (IPR) categories. IPR is defined as family income, divided by the appropriate federal poverty level. The IPR categories under consideration are 0–138%, 138–200%, 200–250%, 250–400%, and 400+%. The thresholds for the first three IPR categories are motivated, in part, by needs of the Centers for Disease Control and Prevention, which provides breast and cervical cancer screenings for low income and uninsured

women. The IPR categories are also relevant to the Affordable Care Act which increased access to health insurance. In participating states, Medicaid programs expanded health insurance access to individuals and families with IPR less than 138% and provided tax credits for those with IPR between 138% and 400%.

The true number of people within each IPR category is unknown and must be estimated. Thus, to create estimates of the proportion with health insurance by IPR category, health insurance and IPR category must be modeled simultaneously. Within each IPR category an individual may be categorized as either having or not having health insurance. In this manner we view individuals as falling into one of 10 distinct categories, $(C_{1,0}, \ldots, C_{5,0}, C_{1,1}, \ldots, C_{5,1})$, where $C_{j,k}$ indicates an individual in IPR category $j = 1, \ldots, 5$ and health insurance indicator $k = 0, 1$. The multivariate structure of the data, with many IPR categories, and the need to estimate both the number in each IPR category along with the proportion with health insurance makes unit-level modeling appealing for this dataset. In addition, there are areas for which there is no sample, areas for which there are direct estimates on the boundary of the parameter space that are exactly equal to zero or one, and direct estimates for which the sampling variance estimates are not well defined, which makes area-level modeling challenging. Since there is no established theory for applying area-level methodology to this type of survey data, we restrict our simulation study to unit-level methods.

To construct health insurance estimates by county and IPR category, we fit the PL-MM with 10 categories, using $n_i = 1$ for all i. We let $\boldsymbol{x}_i$ consist of poststratification variables, including race category, sex, and age category. We also let $\boldsymbol{\phi}_i$ be a vector indicating which county unit i resides in. Thus, the model uses a county level random effect. We use a vague prior distribution over $\boldsymbol{\beta}$ and σ_η^2 by setting $\sigma_\beta^2 = 1000$ and $a = b = 0.5$. A sensitivity analysis, given in Appendix C, confirmed that these prior choices had very little effect on the model outcome, but for other data scenarios this choice should be considered carefully. The model is fit using both the MCMC and VB fitting strategies, with both drawing a posterior sample size of 1000, after discarding 1000 draws as burn-in for MCMC. For MCMC, convergence was assessed visually through the use of traceplots of the sample chains along with the Geweke convergence diagnostic (Geweke (1992)) for which no lack of convergence was detected. After fitting the model on the sample data, predictions are made for all units in the population. The synthesized population is then aggregated to the desired level of the estimates (i.e., county by IPR category). This is done for each posterior draw, giving a posterior predictive distribution for the desired estimates.

To assess the SAE capability of our PL-MM model through simulation, we treat the 2014 one-year American Community Survey (ACS) sample in Minnesota as our population. This data contains roughly 120,000 respondents across Minnesota's 87 counties. We then take a further probability proportional to size sample without replacement, using the Poisson method (Brewer, Early and Hanif (1984)) with an expected sample size of 10,000. We use the size variable $s_i = \exp\{w_i^* + 2\mathrm{I}(H_i = 0)\}$, where w_i^* is the original survey weight for unit i after scaling to have mean zero and standard deviation of one, and H_i indicates whether or not unit i had health insurance. Estimates are constructed using the PL-MM with both MCMC and VB fits. We also construct a Horvitz–Thompson direct estimate as well as an unweighted direct estimate. We repeat the sampling and estimation process 50 times in order to compare MSE and bias across estimators.

A summary of the simulation results in given in Table 1, including average mean squared error (MSE) and squared bias for the competing estimators as well as computation time and 95% credible interval (CI) coverage rates for the two model based estimators. The higher bias of the UW estimator relative to the direct estimator indicates that the sampling scheme was indeed informative. The two model based approaches yield significant reductions to MSE when compared to the direct estimator. Surprisingly, the predictions using the VB approach

TABLE 1

MSE and squared bias of the four estimators averaged across counties based on simulation results. Average computation time in seconds and 95% credible interval coverage rate are also given for the model based estimates

Estimator	MSE	Bias^2	Time (s)	Coverage rate
Model MCMC	7.1×10^{-3}	3.7×10^{-3}	7314	**94%**
Model VB	$\mathbf{2.3 \times 10^{-3}}$	$\mathbf{1.7 \times 10^{-3}}$	**140**	87%
Direct	9.9×10^{-2}	3.8×10^{-2}	–	–
UW direct	1.6×10^{-1}	1.1×10^{-1}	–	–

had even lower MSE than the predictions using MCMC. The reason for the reduced MSE is not entirely clear and is a subject for future research. The downside to the VB approach is that the approximate posterior results in uncertainty estimates that are not optimal. This is reflected in the lower 95% CI coverage rate for the VB approach, compared to the MCMC approach. This is to be expected, as the VB approach only approximates the true posterior distribution. However, the differences are relatively minor and can be justified through the massive decrease in computation time.

We also show the MSE by county and IPR category for each estimator in Figure 1. The largest reductions in MSE through model-based estimation tend to occur for the more rural and sparsely populated regions of the state. These counties tend to have smaller sample sizes, resulting in more erratic direct estimates. The model-based estimates borrow strength from sampled units in all counties, resulting in more stable (i.e., lower MSE) estimates.

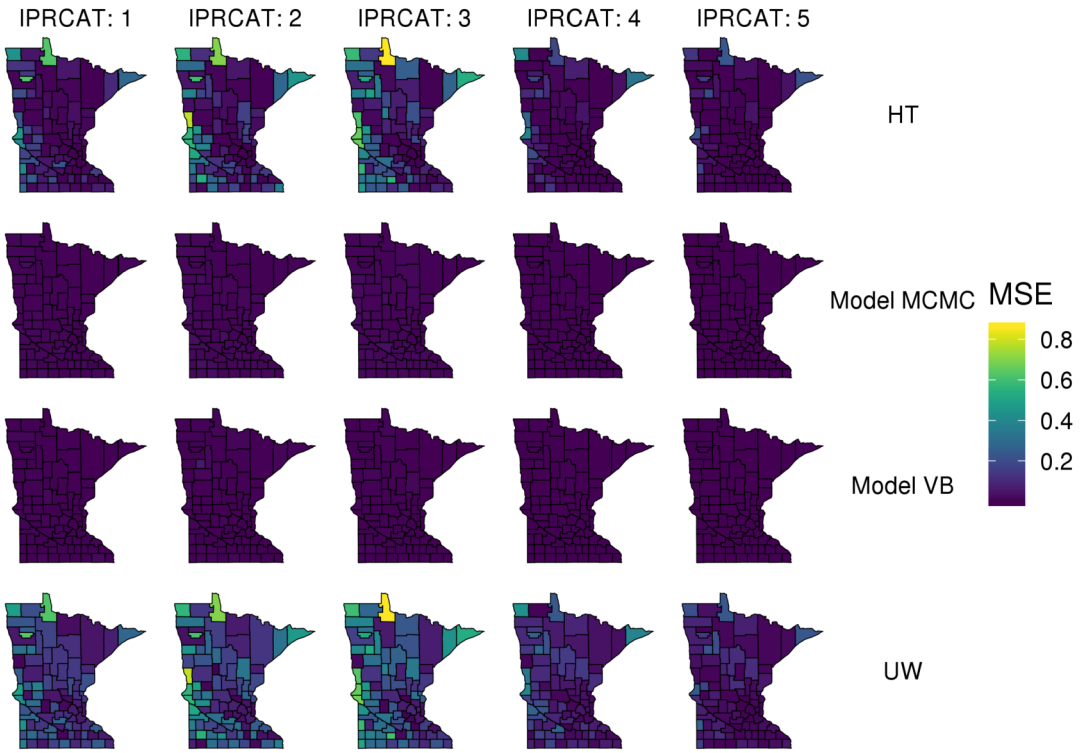

FIG. 1. *Empirical mean squared error by county across the simulation based estimates for the state of Minnesota. Columns represent the different IPR categories and rows represent the different estimators.*

4. Data analysis. The simulation in Section 3 illustrates how the PL-MM model may be used to generate SAHIE type estimates for a single state. However, the SAHIE program is tasked with creating estimates for the entirety of the U.S. rather than a single state. The bottleneck in the MCMC approach to the PL-MM model is the generation of Pólya-Gamma random variables for every sample observation at every MCMC iteration. Although this approach is feasible at a state level, it becomes unwieldy at the national level, where the ACS samples 3.5 million households annually. For this reason we rely on the VB approach to the PL-MM model in order to create estimates of health insurance by county and IPR category for the entire continental U.S.

Again, we use the PL-MM model with 10 categories and $n_i = 1$ for all i. We also use the same prior distribution and poststratification variables that were considered in Section 3. There are over 3000 counties in the U.S., compared to only 87 in Minnesota; thus, we require a form of dimension reduction for ϕ_i rather than using county indicators. To do this, we let ϕ_i be equal to a set of spatial basis functions evaluated for unit i. Specifically, for illustration we use the first 307 (10%) eigenvectors of the county adjacency matrix as our spatial basis functions. This choice was motivated, in part, by the suggestion of Hughes and Haran (2013) to use 10% of the available eigenvectors as well as by the need for substantial dimension reduction with respect to the random effects. In this problem, due to modeling at the unit level, some form of dimension reduction is needed to avoid having memory issues that would result from an approximately 4.5 million $\times$ 3000 dimensional matrix. Choosing the number of basis functions is problem specific and constitutes an ongoing area of research; for example, see Bradley, Cressie and Shi (2016) and the references therein.

We fit the PL-MM model, using the VB approach, with a sample size of roughly 4.5 million. We then take 1000 independent draws from the variational posterior distribution in order to construct the posterior predictive distribution of our estimates. Treating the posterior predictive mean as our point estimates, we plot the model-based estimates alongside the direct estimates in Figure 2. In order to satisfy the disclosure avoidance requirements of the U. S. Census Bureau, a small amount of noise was added to the direct estimates shown in the maps in Figure 2. However, this is the only instance of any additional noise being added to data or estimates. All models were fit to the raw ACS data, and all other results are presented without any additional noise. Visually, the direct estimates are quite noisy, due to the very small sample sizes in many counties. The model-based estimates are able to provide a degree of smoothing through the use of borrowed information in the hierarchical model structure. This results in model based estimates that have the same general spatial pattern as the direct estimates without as much noise. We also plot the health insurance estimates by county without regard to IPR category in Figure 3. Similar patterns can be noticed here.

We plot the ratio of the model-based standard errors to the direct-estimate standard errors by county and IPR category in Figure 4. For the vast majority of estimates, the model-based approach provides quite substantial reductions in standard error, with the largest advantage occurring in the more sparsely populated southern and western regions of the country.

This example demonstrates how the PL-MB and PL-MM models may be used to model complex dependence structures with non-Gaussian data in a computationally efficient manner. The VB approach specifically was able to generate estimates for over 15,000 county and IPR category combinations, utilizing a sample size of over four million, in roughly 17 hours. These estimates are much less noisy than direct estimates, with substantially lower standard errors. Furthermore, the simulation results of Section 3 indicate that these model-based estimates should have much lower MSE. In addition to advantages over the direct estimate, this approach has many advantages over area-level modeling approaches, such as the one currently in use for SAHIE. For example, unit-level models allow for easy aggregation to multiple domains. A single PL-MM model may be used to give county- and state-level estimates, whereas area-level modeling strategies require two separate models and often rely on

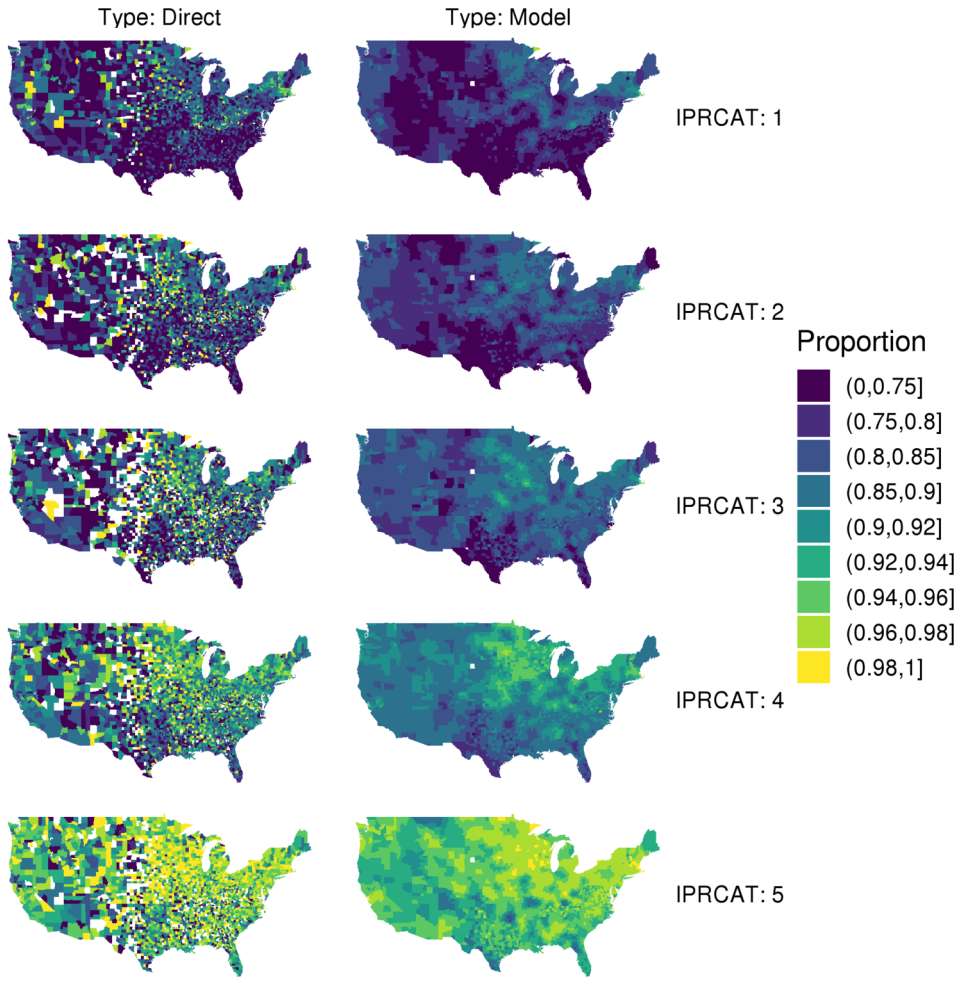

FIG. 2. *Direct- and model-based estimates of the proportion of the population with health insurance by county and IPR category for the continental United States.*

ad hoc benchmarking techniques. Another advantage is that unit-level models do not require a direct estimate for a given area in order to construct an estimate, in contrast to area-level models.

5. Discussion. This paper establishes a framework for modeling binomial and multinomial unit-level survey data, specifically under an informative sample. We envision this methodology being used to create area-level estimates of population proportions with health insurance (SAHIE) as our motivating example. The current methodology used to generate SAHIE estimates is conducted at the area level which can cause a number of problems that are alleviated through the use of unit-level modeling. Our unit-level approach is able to generate multiple levels of estimates through a single model without the need for benchmarking techniques. We demonstrate this by producing health insurance estimates by county as well as by IPR category within each county for the entire continental U.S. Our approach is also able to produce very precise estimates, compared to traditional direct estimators, as demonstrated by our empirical simulation study. Finally, these estimates can be produced in a very computationally efficient manner either through the use of either Gibbs sampling with fully conjugate full-conditional distributions or through a VB approximation to the posterior distribution.

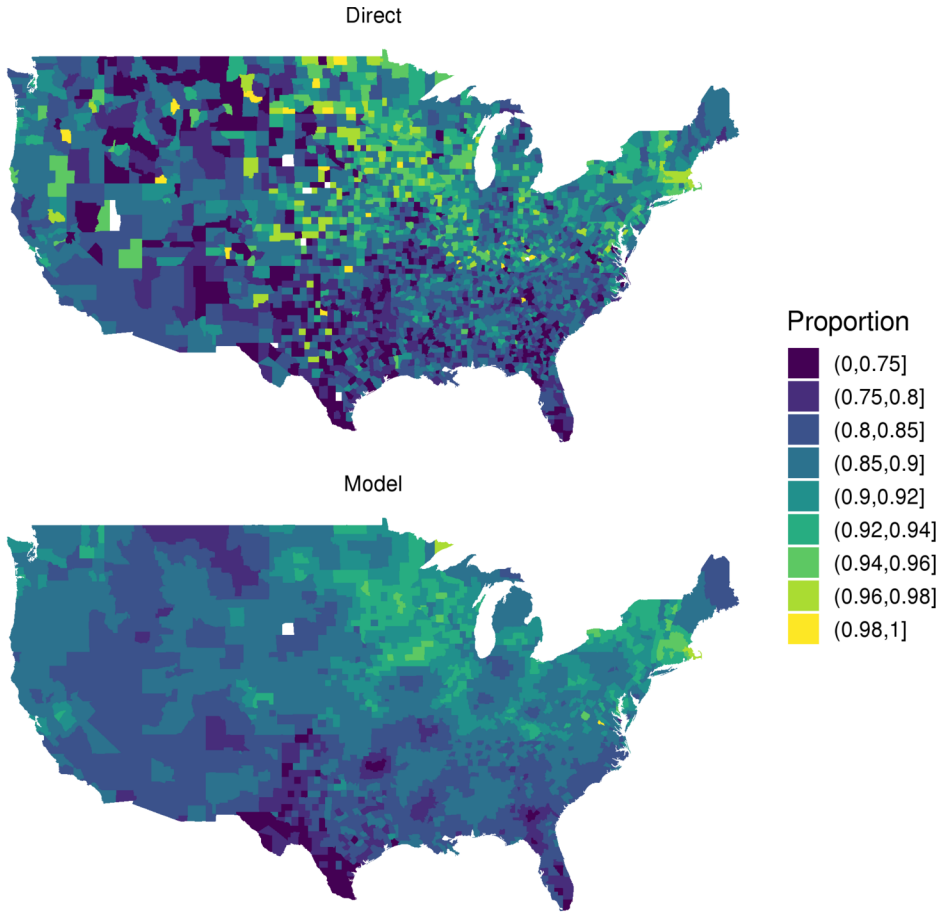

FIG. 3. *Direct- and model-based estimates of the proportion of the population with health insurance by county for the continental United States.*

Although this paper provides a methodological step forward for small-area estimates of health insurance, further work would be necessary to create estimates that might replace the current SAHIE program. For example, the current SAHIE methodology considers a number of important covariates that were not considered here, due to disclosure limitations, including data from the Supplemental Nutrition Assistance Program as well as Medicaid. Furthermore, the method considered here is a type of generalized linear model, but there is potential for improvement through the use nonlinear modeling techniques which is the subject of future work.

APPENDIX A: FULL CONDITIONAL DISTRIBUTIONS FOR PL-MB MODEL

Let $\boldsymbol{\Omega} = \operatorname{diag}(\omega_1, \ldots, \omega_n)$ and $\boldsymbol{\kappa} = (\tilde{w}_1 * (y_1 - n_1/2), \ldots, \tilde{w}_n * (y_n - n_n/2))'$. Note that $\boldsymbol{\kappa}/\boldsymbol{\omega}$ represents elementwise division,

$$\omega_i|\cdot \sim \mathrm{PG}(\tilde{w}_i * n_i, \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{\psi}_i'\boldsymbol{\eta}), \quad i = 1, \ldots, n,$$

$$\boldsymbol{\eta}|\cdot \propto \prod_{i=1}^{n} \exp\left(\kappa_i \boldsymbol{\phi}_i'\boldsymbol{\eta} - \frac{1}{2}\omega_i(\boldsymbol{\phi}_i'\boldsymbol{\eta})^2 - \omega_i(\boldsymbol{\phi}_i'\boldsymbol{\eta})(\boldsymbol{x}_i'\boldsymbol{\beta})\right)$$

$$\times \exp\left(-\frac{1}{2\sigma_\eta^2}\boldsymbol{\eta}'\boldsymbol{\eta}\right),$$

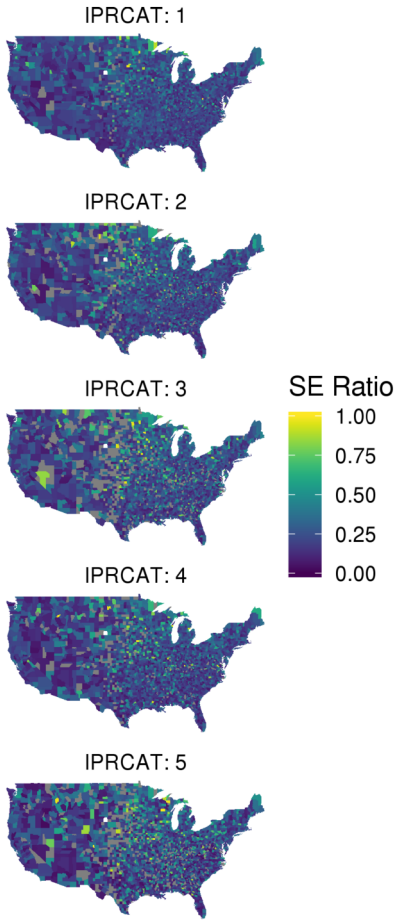

FIG. 4. *Ratio of model-based standard errors to direct-estimate standard errors by county and IPR category for the continental United States. Counties with no available direct estimate are shown in gray.*

$$\propto \exp\left(-\frac{1}{2}(\boldsymbol{\kappa}/\boldsymbol{\omega} - \boldsymbol{X\beta} - \boldsymbol{\Phi\eta})'\boldsymbol{\Omega}(\boldsymbol{\kappa}/\boldsymbol{\omega} - \boldsymbol{X\beta} - \boldsymbol{\Phi\eta}) - \frac{1}{\sigma_\eta^2}\boldsymbol{\eta}'\boldsymbol{\eta}\right),$$

$$\boldsymbol{\eta}|\cdot \sim \mathrm{N}_r\left(\boldsymbol{\mu} = \left(\boldsymbol{\Phi}'\boldsymbol{\Omega}\boldsymbol{\Phi} + \frac{1}{\sigma_\eta^2}\boldsymbol{I}_r\right)^{-1}\boldsymbol{\Phi}'\boldsymbol{\Omega}(\boldsymbol{\kappa}/\boldsymbol{\omega} - \boldsymbol{X\beta}), \boldsymbol{\Sigma} = \left(\boldsymbol{\Phi}'\boldsymbol{\Omega}\boldsymbol{\Phi} + \frac{1}{\sigma_\eta^2}\boldsymbol{I}_r\right)^{-1}\right),$$

$$\boldsymbol{\beta}|\cdot \propto \prod_{i=1}^{n}\exp\left(\kappa_i\boldsymbol{x}_i'\boldsymbol{\beta} - \frac{1}{2}\omega_i(\boldsymbol{x}_i'\boldsymbol{\beta})^2 - \omega_i(\boldsymbol{x}_i'\boldsymbol{\beta})(\boldsymbol{\phi}_i'\boldsymbol{\eta})\right)$$

$$\times \exp\left(-\frac{1}{2\sigma_\beta^2}\boldsymbol{\beta}'\boldsymbol{\beta}\right)$$

$$\propto \exp\left(-\frac{1}{2}(\boldsymbol{\kappa}/\boldsymbol{\omega} - \boldsymbol{\Phi\eta} - \boldsymbol{X\beta})'\boldsymbol{\Omega}(\boldsymbol{\kappa}/\boldsymbol{\omega} - \boldsymbol{\Phi\eta} - \boldsymbol{X\beta}) - \frac{1}{\sigma_\beta^2}\boldsymbol{\beta}'\boldsymbol{\beta}\right),$$

$$\boldsymbol{\beta}|\cdot \sim \mathrm{N}_p\left(\boldsymbol{\mu} = \left(\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{X} + \frac{1}{\sigma_\beta^2}\boldsymbol{I}_p\right)^{-1}\boldsymbol{X}'\boldsymbol{\Omega}(\boldsymbol{\kappa}/\boldsymbol{\omega} - \boldsymbol{\Phi\eta}), \boldsymbol{\Sigma} = \left(\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{X} + \frac{1}{\sigma_\beta^2}\boldsymbol{I}_p\right)^{-1}\right),$$

$$\sigma_\eta^2|\cdot \propto (\sigma_\eta^2)^{-\frac{r}{2}}\exp\left(-\frac{1}{2\sigma_\eta^2}\boldsymbol{\eta}'\boldsymbol{\eta}\right)$$

$$\times (\sigma_\eta^2)^{-a-1} \exp\left(-\frac{1}{\sigma_\eta^2} b\right)$$

$$\propto (\sigma_\eta^2)^{-(a+\frac{r}{2})-1} \exp\left(-\frac{1}{\sigma_\eta^2}\left(b + \frac{\boldsymbol{\eta}'\boldsymbol{\eta}}{2}\right)\right),$$

$$\sigma_\eta^2|\cdot \sim \text{IG}\left(a + \frac{r}{2}, b + \frac{\boldsymbol{\eta}'\boldsymbol{\eta}}{2}\right).$$

APPENDIX B: POSTSTRATIFICATION ROUTINE WITH THE VARIATIONAL DISTRIBUTION

To construct our estimates, we require response predictions for every unit in the population. We will assume for the sake of exposition that the responses are binary, but these same techniques can be applied to categorical responses as well.

Because our model utilizes only categorical rather than continuous covariates, computation can be simplified through the use of poststratification cells. Specifically, let $j = 1, \ldots, J$ index the J unique poststratification cells (e.g., the unique combinations of categorical covariates and county indicators). Each cell is also associated with a population size, N_j. Within each cell, population units are exchangeable, and predicted responses can be generated from the same distribution. To estimate p_j, the probability of a successful outcome in cell j, we require estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ as well as the vector of cell covariates, $\boldsymbol{x}_j$ and the vector of spatial basis functions for the cell, $\boldsymbol{\phi}_j$.

To begin, we work with our variational distribution for $\boldsymbol{\zeta} = (\boldsymbol{\beta}', \boldsymbol{\eta}')$. We can sample from this distribution by sampling from a $\text{N}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}})$ distribution, where $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\Sigma}}$ are estimated from the variational Bayes procedure outlined in Algorithm 1. We take R total posterior samples, yielding $\boldsymbol{\beta}^{(r)}$ and $\boldsymbol{\eta}^{(r)}$ for $r = 1, \ldots, R$. Then, for each posterior sample r and each cell j, we generate the population (i.e., the number of positive responses) within the cell, $s_j^{(r)}$, by sampling from $\text{Bin}(N_j, p_j^{(r)})$, where $p_j^{(r)} = \text{logit}^{-1}(\boldsymbol{x}_j'\boldsymbol{\beta}^{(r)} + \boldsymbol{\phi}_j'\boldsymbol{\eta}^{(r)})$. Having effectively generated a synthetic population, we can aggregate the units within a given domain to generate a population estimate. For example, for a given iteration r, we can create an estimate of the population proportion in county c as

$$p_c^{(r)} = \frac{\sum_{j \in c} s_j^{(r)}}{\sum_{j \in c} N_j}.$$

Then, for our point estimate of the population proportion in county c, we use the posterior mean,

$$\hat{p}_c = \frac{1}{R} \sum_{r=1}^{R} p_c^{(r)}.$$

APPENDIX C: PRIOR SENSITIVITY ANALYSIS

The choice of prior distributions in our model construction was based on computational efficiency; however, other distributions could be considered. Our choice of hyperparameters was intended to induce a vague prior distribution, but we conduct a prior sensitivity analysis to confirm that our results are not sensitive to this choice.

For one of the simulated datasets in Section 3, we fit our model, using the variational Bayes procedure, under a variety of different hyperparameter settings. In particular, we consider the grid over $\sigma_\beta^2 = \{10, 1000, 10{,}000\}$ and $a = b = \{0.1, 0.5, 2\}$. This results in eight alternative hyperparameter settings that we can compare to our original choice ($\sigma_\beta^2 = 1000$

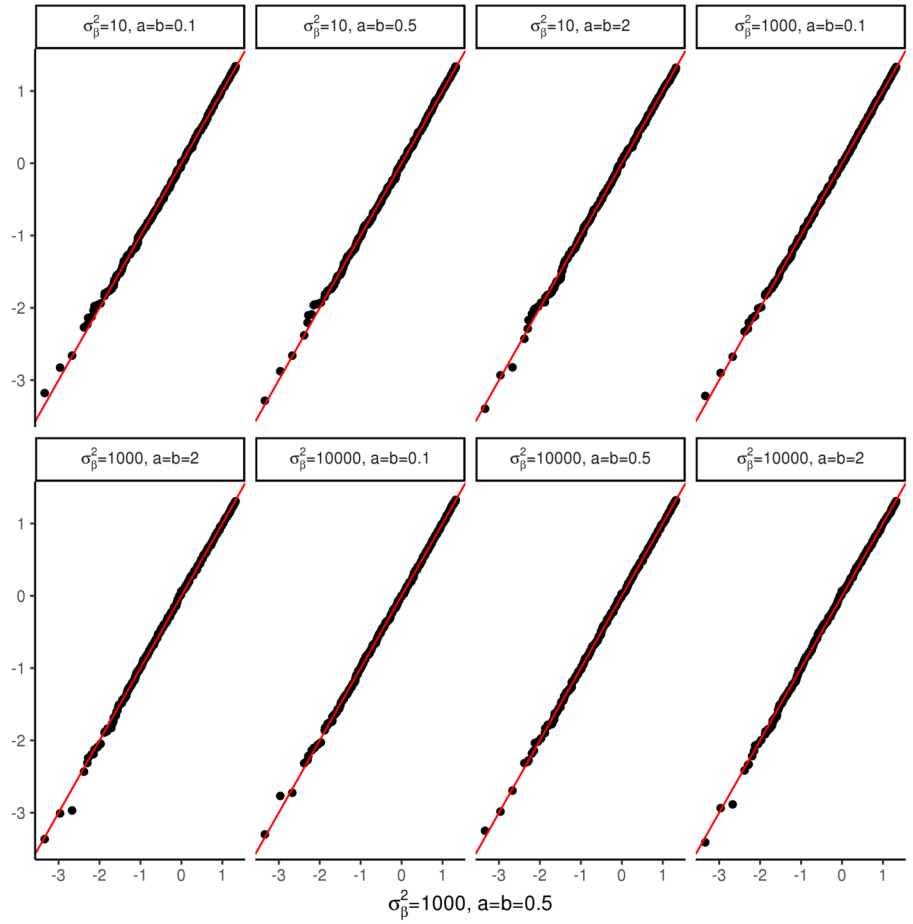

FIG. 5. *Q-Q plots of standardized estimates for a simulated sample dataset under various hyperparameter settings, compared to the specification chosen for analysis, where* $\sigma_\beta^2 = 1000$ *and* $a = b = 0.5$.

and $a = b = 0.5$). For each hyperparameter setting, we construct a quantile-quantile (Q-Q) plot of the standardized estimates, compared to the standardized estimates under our original specification. These are given in Figure 5. In every case the points lie nearly perfectly on the one-to-one line, confirming that the distribution of our estimates is not sensitive to the choice of hyperparameters.

Acknowledgments. We thank the Editor, Beth Ann Griffin, and two anonymous referees for valuable comments that helped improve this paper.

Funding. Support for this research at the Missouri Research Data Center (MURDC) and through the Census Bureau Dissertation Fellowship program is gratefully acknowledged.

This research was partially supported by the U.S. National Science Foundation (NSF) under NSF Grant SES-1853096. This article is released to inform interested parties of ongoing research and to encourage discussion. The views expressed on statistical issues are those of the authors and not those of the NSF or U.S. Census Bureau. The DRB approval number for this paper is CBDRB-FY20-355.

SUPPLEMENTARY MATERIAL

Supplement to "Computationally efficient Bayesian unit-level models for non-Gaussian data under informative sampling with application to estimation of health

insurance coverage" (DOI: 10.1214/21-AOAS1524SUPP; .zip). We provide code as well as a data example using public use microdata.

REFERENCES

ALBERT, J. H. and CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88** 669–679. MR1224394

BATTESE, G. E., HARTER, R. M. and FULLER, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.* **83** 28–36.

BAUDER, M., LUERY, D. and SZELEPKA, S. (2018). Small area estimation of health insurance coverage in 2010–2016. Technical Report. Small Area Methods Branch, Social, Economic, and Housing Statistics Division, U. S. Census Bureau.

BEAL, M. J. and GHAHRAMANI, Z. (2003). The variational Bayesian EM algorithm for incomplete data: With application to scoring graphical model structures. In *Bayesian Statistics*, 7 (*Tenerife*, 2002) 453–463. Oxford Univ. Press, New York. MR2003189

BINDER, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *Int. Stat. Rev.* **51** 279–292. MR0731144 https://doi.org/10.2307/1402588

BRADLEY, J. R., CRESSIE, N. and SHI, T. (2016). A comparison of spatial predictors when datasets could be very large. *Stat. Surv.* **10** 100–131. MR3527662 https://doi.org/10.1214/16-SS115

BRADLEY, J. R., HOLAN, S. H. and WIKLE, C. K. (2020). Bayesian hierarchical models with conjugate full-conditional distributions for dependent data from the natural exponential family. *J. Amer. Statist. Assoc.* **115** 2037–2052. MR4189775 https://doi.org/10.1080/01621459.2019.1677471

BREWER, K. R. W., EARLY, L. J. and HANIF, M. (1984). Poisson, modified Poisson and collocated sampling. *J. Statist. Plann. Inference* **10** 15–30. MR0752450 https://doi.org/10.1016/0378-3758(84)90029-6

DURANTE, D. and RIGON, T. (2019). Conditionally conjugate mean-field variational Bayes for logistic models. *Statist. Sci.* **34** 472–485. MR4017524 https://doi.org/10.1214/19-STS712

GELMAN, A. and LITTLE, T. C. (1997). Poststratification into many categories using hierarchical logistic regression. *Surv. Methodol.* **23** 127–135.

GEWEKE, J. F. (1991). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Staff Report No. 148, Federal Reserve Bank of Minneapolis.

HORVITZ, D. G. and THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47** 663–685. MR0053460

HUGHES, J. and HARAN, M. (2013). Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 139–159. MR3008275 https://doi.org/10.1111/j.1467-9868.2012.01041.x

JORDAN, M. I., GHAHRAMANI, Z., JAAKKOLA, T. S. and SAUL, L. K. (1999). An introduction to variational methods for graphical models. *Mach. Learn.* **37** 183–233.

LINDERMAN, S., JOHNSON, M. J. and ADAMS, R. P. (2015). Dependent multinomial models made easy: Stick-breaking with the Pólya-Gamma augmentation. In *Advances in Neural Information Processing Systems* 3456–3464.

LITTLE, R. J. (2012). Calibrated Bayes, an alternative inferential paradigm for official statistics. *J. Off. Stat.* **28** 309.

PARK, D. K., GELMAN, A. and BAFUMI, J. (2006). State-level opinions from national surveys: Poststratification using multilevel logistic regression. In *Public Opinion in State Politics* Stanford Univ. Press, Stanford, CA.

PARKER, P. A., HOLAN, S. H. and JANICKI, R. (2020). Conjugate Bayesian unit-level modelling of count data under informative sampling designs. *Stat* **9** e4267, 9. MR4104231 https://doi.org/10.1002/sta4.267

PARKER, P. A., HOLAN, S. H. and JANICKI, R. (2022). Supplement to "Computationally efficient Bayesian unit-level models for non-Gaussian data under informative sampling with application to estimation of health insurance coverage." https://doi.org/10.1214/21-AOAS1524SUPP

PARKER, P. A., JANICKI, R. and HOLAN, S. H. (2019). Unit level modeling of survey data for small area estimation under informative sampling: A comprehensive overview with extensions. Preprint. Available at arXiv:1908.10488.

PFEFFERMANN, D. and SVERCHKOV, M. (2007). Small-area estimation under informative probability sampling of areas and within the selected areas. *J. Amer. Statist. Assoc.* **102** 1427–1439. MR2412558 https://doi.org/10.1198/016214507000001094

POLSON, N. G., SCOTT, J. G. and WINDLE, J. (2013). Bayesian inference for logistic models using Pólya-Gamma latent variables. *J. Amer. Statist. Assoc.* **108** 1339–1349. MR3174712 https://doi.org/10.1080/01621459.2013.829001

SAVITSKY, T. D. and TOTH, D. (2016). Bayesian estimation under informative sampling. *Electron. J. Stat.* **10** 1677–1708. MR3522657 https://doi.org/10.1214/16-EJS1153

SI, Y., PILLAI, N. S. and GELMAN, A. (2015). Bayesian nonparametric weighted sampling inference. *Bayesian Anal.* **10** 605–625. MR3420817 https://doi.org/10.1214/14-BA924

SKINNER, C. J. (1989). Domain means, regression and multivariate analysis. In *Analysis of Complex Surveys* (C. J. Skinner, D. Holt and T. M. F. Smith, eds.) 80–84. Wiley, Chichester.

STAN DEVELOPMENT TEAM (2021). Stan modeling language users guide and reference manual, version 2.26. https://mc-stan.org.

VANDENDIJCK, Y., FAES, C., KIRBY, R. S., LAWSON, A. and HENS, N. (2016). Model-based inference for small area estimation with sampling weights. *Spat. Stat.* **18** 455–473. MR3575502 https://doi.org/10.1016/j.spasta.2016.09.004

WAINWRIGHT, M. J., JORDAN, M. I. et al. (2008). Graphical models, exponential families, and variational inference. *Found. Trends Mach. Learn.* **1** 1–305.

WINDLE, J., POLSON, N. and SCOTT, J. (2013). BayesLogit: Bayesian logistic regression. http://cran.r-project.org/web/packages/BayesLogit/index.html. R package version 0.2-4.

ZHANG, X., HOLT, J. B., LU, H., WHEATON, A. G., FORD, E. S., GREENLUND, K. J. and CROFT, J. B. (2014). Multilevel regression and poststratification for small-area estimation of population health outcomes: A case study of chronic obstructive pulmonary disease prevalence using the behavioral risk factor surveillance system. *Am. J. Epidemiol.* **179** 1025–1033.