Journal of Survey Statistics and Methodology (2023) ***00****, 1–29*

A COMPREHENSIVE OVERVIEW OF UNIT-LEVEL MODELING OF SURVEY DATA FOR SMALL AREA ESTIMATION UNDER INFORMATIVE SAMPLING

PAUL A. PARKER*
RYAN JANICKI
SCOTT H. HOLAN

Model-based small area estimation is frequently used in conjunction with survey data to establish estimates for under-sampled or unsampled geographies. These models can be specified at either the area-level, or the unit-level, but unit-level models often offer potential advantages such as more precise estimates and easy spatial aggregation. Nevertheless, relative to area-level models, literature on unit-level models is less prevalent. In modeling small areas at the unit level, challenges often arise as a consequence of the informative sampling mechanism used to collect the survey data. This article provides a comprehensive methodological review for unit-level models under informative sampling, with an emphasis on Bayesian approaches.

KEYWORDS: Bayesian analysis; Informative sampling; Pseudo-likelihood; Small area estimation; Survey sampling.

PAUL A. PARKER is an Assistant Professor in the Department of Statistics, University of California Santa Cruz, 1156 High St, Santa Cruz, CA 95064, USA. RYAN JANICKI is a Principal Researcher in the U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233-9100, USA. SCOTT H. HOLAN is Professor in the Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO 65211-6100, USA and U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233-9100, USA.

This research was partially supported by the U.S. National Science Foundation (NSF) and the U.S. Census Bureau under NSF Grant SES-1132031, funded through the NSF-Census Research Network (NCRN) program and NSF grants SES-1853096, NCSE-2215168, and NCSE-2215169. Support for this research at the Missouri Research Data Center (MURDC), through the University of Missouri Population, Education and Health Center Doctoral Fellowship, as well as through the U.S. Census Bureau Dissertation Fellowship Program, is also gratefully acknowledged. 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 article is CBDRB-FY21-331.

*Address correspondence to Paul A. Parker, Department of Statistics, University of California Santa Cruz, 1156 High St, Santa Cruz, CA 95064, USA; E-mail: paulparker@ucsc.edu.

https://doi.org/10.1093/jssam/smad020
Published by Oxford University Press on behalf of the American Association for Public Opinion Research 2023.

Statement of Significance

This article provides a comprehensive review of unit-level modeling approaches for small area estimation in the presence of informative sampling. Specifically, we cover a broad range of approaches, including those that assume or force an ignorable design, those that use pseudo-likelihood approaches, and those that specify differing sample and population models.

1. INTRODUCTION

Government agencies have seen an increase in demand for data products and small-area statistics in recent years as these estimates are often used for influencing government policies and for allocating federal funds (see Rao and Molina 2015, Chapter 1.2). The Small Area Income and Poverty Estimates (SAIPE) program (Bell et al. 2016) and the Small Area Health Insurance Estimates program (Luery 2011; Bauder et al. 2018) within the U.S. Census Bureau are two examples of government programs, which produce county- and sub-county-level estimates for different demographic cross classifications across the entire United States using small area estimation (SAE) methods. Other statistical agencies that produce small area estimates include National Center for Health Statistics, Bureau of Labor Statistics, and Statistics Canada. One trend that has accompanied this demand is the need for granular estimates of parameters of interest at small spatial scales or subdomains of the finite population. Typically, sample surveys are designed to provide reliable estimates of the parameters of interest for large domains. However, for some subpopulations, the area-specific sample size may be too small to produce estimates with adequate precision. The term *small area* is used to refer to any domain of interest, such as a geographic area or demographic cross-classification, for which the domain-specific sample size is not large enough for reliable direct estimation. To improve precision, model-based methods can be used to "borrow strength," by relating the different areas of interest through use of linking models, and by introducing area-specific random effects and covariates.

Models for SAE may be specified either at the area level or the unit level (see Rao and Molina 2015, for an overview of SAE methodology). Area-level models treat the direct estimate (e.g., the survey-weighted estimate of a mean) as the response and typically induce some type of smoothing across areas. In this way, the areas with limited sample sizes may "borrow strength" from areas with larger samples. While area-level models are popular, they are limited, in that it is difficult to make estimates and predictions at a geographic or demographic level that is finer than the level of the aggregated direct estimate.

In contrast, unit-level models use individual survey units as the response data, rather than the direct estimates. Use of unit-level models can overcome some of the limitations of area-level models, as they constitute a bottom-up approach (i.e., they utilize the finest scale of resolution of the data). Since model inputs are at the unit-level (person-level, household-level, or establishment-level), predictions and estimates can be made at the same unit level, or aggregated up to any desired level. Unit-level modeling also has the added benefit of ensuring logical consistency of estimates at different geographic levels. For example, model-based county estimates are forced to aggregate to the corresponding state-level estimates, eliminating the need for ad hoc benchmarking. In addition, because the full unit-level dataset is used in the modeling, rather than the summary statistics used with area-level models, there is potential for improved precision of estimated quantities (Hidiroglou and You 2016).

Although unit-level models may lead to more precise estimates that aggregate naturally across different spatial resolutions, they also introduce new challenges. Perhaps the biggest challenge is accounting for the survey design in the model. With area-level models, the survey design is incorporated into the model through specification of a sampling distribution (typically taken to be Gaussian) and inclusion of direct variance estimates. With unit-level models, accounting for the survey design is not as simple. One challenge is that the sample unit response may be dependent on the probability of selection, even after conditioning on the design variables. When the response variables are correlated with the sample selection variables, the sampling scheme is said to be *informative*, and in these scenarios, to avoid bias, it is critical to capture the sample design in the model by including the survey weights or the design variables used to construct the survey weights.

The choice of whether to use an area or a unit-level model often depends on the goal of the analysis and available data. In particular, a subject matter expert may only have access to area-level tabulations and public-use micro samples. Consequently, if the goal is inference and/or the desired geography is something other than a public-use micro area (PUMA), such as a county or tract, then the analysis would necessitate an area-level model. Conversely, if the goal of the model is to produce tabulations at an official statistics agency, then a unit-level model may be more appropriate. Specifically, depending on the application, the unit-level model may yield greater precision while simultaneously providing consistent aggregation (e.g., avoiding ad hoc benchmarking) and the ability to produce tabulations for any geography (e.g., zip codes and/or American Indian Alaskan Native areas). In other words, unit-level models provide a "bottom-up" approach and, thus, will be internally consistent across various aggregations. This is in contrast to the benchmarking where lower-level direct-estimates are forced to aggregate to higher-level direct-estimates.

The aim of this article is to present a comprehensive literature review of unit-level small area modeling strategies under informative sampling. Some

recent review articles, which give an overview of the unique challenges of modeling survey data collected under informative designs, and the role of survey weights and design variables in statistical models, can be found in Pfeffermann (1993), Gelman (2007), and Lumley and Scott (2017). Pfeffermann (2002) and Pfeffermann (2003) review both area-level and unit-level SAE methods. Chapter 7 of Rao and Molina (2015) provides a review of some commonly used unit-level small area models. The current article adds to this literature by providing a comprehensive review of unit-level small area modeling techniques, with a focus on methods that account for informative sampling designs. We note that many model-based methods are general enough to be implemented in either a frequentist or Bayesian setting, and we highlight some scenarios where Bayesian methodology may be used, as Bayesian methods are becoming more prevalent within statistical agencies (e.g., SAIPE and Voting Rights Act 203b, among others). Importantly, we are not attempting to unify frequentist and Bayesian approaches in the context of finite population inference, though this is an interesting area of research. Related to this topic of research is the work of Little (2012) on *Calibrated Bayes*.

In this article, we focus mainly on model specification and methods that incorporate informative sampling designs into the small area model. Some important, related issues, that will be outside the scope of this article include issues related to measurement error and adjustments for nonresponse. Generally, we assume that observed survey weights have been modified to take into account nonresponse. We also avoid discussion on the relative merits of frequentist versus Bayesian methods for inference.

The remainder of this article is organized as follows. Section 2 introduces the sampling framework and notation to be used throughout the article. We aim to keep the notation internally consistent. This may lead to differences compared to the original authors' notation styles but should lead to easier comparison across methodologies. In section 3, we cover modeling techniques that assume a noninformative survey design. The basic unit-level model is introduced, as well as extensions of this model which incorporate the design variables and survey weights. Methods that allow for an informative design are then discussed, beginning in section 4. Here, we discuss the analytic inference of population parameters under an informative design using pseudo-likelihood methods. Extensions of the pseudo-likelihood to hierarchical, multilevel mixed models are discussed, as well as application to SAE problems. In section 5, we focus on models that use a sample distribution that differs from the population distribution. We conclude the review component of this article in section 6, where we will review models that are specific to a Binomial likelihood, as many variables collected from survey data are binary. Finally, we provide concluding remarks in section 7. A simulation study and application of some of the methodology presented herein, as well as discussion on the practical trade-offs between the methodologies, can also be found in Parker et al. (2023).

2. BACKGROUND AND NOTATION

Consider a finite population $\mathcal{U}$ of size N, which is a subset into m nonoverlapping domains, $\mathcal{U}_i = \{1, \ldots, N_i\}, i = 1, \ldots, m$, where $\sum_{i=1}^{m} N_i = N$. These subgroups will typically be small areas of interest, or socio-demographic cross-classifications, such as age by race by gender within the different geographies. We use y_{ij} to represent a particular response characteristic associated with unit $j \in \mathcal{U}_i$, and $\boldsymbol{x}_{ij}$ a vector of predictors for the response variables y_{ij}.

Let $\boldsymbol{Z}$ be a vector of design variables, which characterize the sampling process. For example, $\boldsymbol{Z}$ may contain geographic variables used for stratifying the population, or size variables used in a probability proportional to size sampling scheme. Note that in some cases, $\boldsymbol{x}_{ij}$ may partially or completely include the design variables. A sample $\mathcal{S} \subset \mathcal{U} = \cup \mathcal{U}_i$ is selected according to a known sampling design with inclusion probabilities dependent on the design variables, $\boldsymbol{Z}$. Let $\mathcal{S}_i$ denote the sampled units in small area i, and let $\pi_{ij} = P(j \in \mathcal{S}_i | \boldsymbol{Z})$. The inverse probability sampling weights are denoted with $w_{ij} = 1/\pi_{ij}$. We note that as analysts, we may not have access to the functional form of $P(j \in \mathcal{S}_i | \boldsymbol{Z})$, and may not even have access to the design variables $\boldsymbol{Z}$, so that the only information available to us about the survey design is through the observed values of π_{ij} or w_{ij}, for the sampled units in the population. Finally, we let $D_S = \{\{y_{ij}, \boldsymbol{x}_{ij}, w_{ij}\} : j \in \mathcal{S}_i, i = 1, \ldots, m\}$ represent the observed data. This simply consists of the responses, predictors, and sampling weights for all units included in the sample. In this context, y_{ij} is random, $\boldsymbol{x}_{ij}$ is typically considered fixed and known, and w_{ij} can either be fixed or random depending on the specific modeling assumptions.

The usual inferential goal, and the main focus of this article, is on the estimation of the small area means, $\bar{y}_i = \sum_{j \in \mathcal{U}_i} y_{ij}/N_i$, or totals, $y_i = \sum_{j \in \mathcal{U}_i} y_{ij}$. The best predictor, $\widehat{\bar{y}}_i$ of $\bar{y}_i$, under squared error loss, given the observed data D_S, is (Pfeffermann and Sverchkov 2007; Molina and Rao 2010; Marhuenda et al. 2017)

$$\widehat{\bar{y}}_i = E(\bar{y}_i | D_S) = \frac{1}{N_i} \sum_{j \in \mathcal{U}_i} E(y_{ij} | D_S) = \frac{1}{N_i} \sum_{j \in \mathcal{S}_i} y_{ij} + \frac{1}{N_i} \sum_{j \in \mathcal{S}_i^c} E(y_{ij} | D_S). \quad (1)$$

The first term on the right-hand side of (1) is known from the observed sample. However, computation of the conditional expectation in the second term requires specification of a model, and potentially, depending on the model specified, auxiliary information, such as knowledge of the covariates $\boldsymbol{x}_{ij}$ or sampling weights w_{ij} for the nonsampled units. For the case where the predictors $\boldsymbol{x}_{ij}$ are categorical, the assumption of known covariates for the nonsampled units is not necessarily restrictive if the totals, $N_{i,g}$, for each cross-classification g in each of the small areas i are known. In this case, the last term in (1) reduces to $N_i^{-1} \sum_g (N_{i,g} - n_{i,g}) E(y_{ij} | j \in g, D_S)$, and only predictions for each

cross-classification need to be made. If the cell totals are unknown, they may be estimated, either using design- or model-based methods.

The predictor given in (1) is general, and the different unit-level modeling methods discussed in this article are essentially different methods for predicting the nonsampled units, that is, estimating the conditional expectation in (1), under different model specifications and assumptions on the finite population and the sampling scheme. An entire finite population can then be generated, consisting of the observed, sampled values, along with model-based predictions for the nonsampled individuals. The small area mean can then be estimated by simply averaging appropriately over this population. If the sampling fraction n_i/N_i in each small area is small, inference using predicted values for the entire population will be nearly the same as inference using a finite population consisting of the observed values and predicted values for the nonsampled units. In this situation, it may be more convenient to use a completely model-based approach for the prediction of the small area means (Battese et al. 1988).

3. UNWEIGHTED ANALYSIS

3.1 Ignorable Design

First, assume the survey design is *ignorable* or *noninformative*. For ease of presentation, let $\boldsymbol{X}$ represent the full set of auxiliary information including covariates and observable design variables (i.e., $\boldsymbol{Z}$ may be partially or fully contained in $\boldsymbol{X}$). Ignorable designs, such as simple random sampling with replacement, arise when the sample inclusion variable I is independent of the response variable y, after conditioning on observable design variables and covariates. In this situation, the distribution of the sampled responses will be identical to the distribution of nonsampled responses. That is, after conditioning on relevant covariates, $\boldsymbol{X}$, if a model $f(\cdot|\boldsymbol{X},\boldsymbol{\theta})$ is assumed to hold for all nonsampled units in the population, then it will also hold for the sampled units, since the *sample distribution* of y, $f(y|I=1,\boldsymbol{X},\boldsymbol{\theta}) = f(y|\boldsymbol{X},\boldsymbol{\theta})$ is identical to the population distribution of y. In this case, a model can be fit to the sampled data, and the fitted model can then be used directly to predict the nonsampled units, without needing any adjustments due to the survey design. Conditions for the ignorability of the sample selection process can be found in Rubin (1976) and Sugden and Smith (1984).

The nested error regression model or, using the terminology of Rao and Molina (2015), the basic unit-level model, was introduced by Battese et al. (1988) for the estimation of small area means using data obtained from a survey with an ignorable design. Consider the linear mixed-effects model

$$y_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + v_i + e_{ij}, \tag{2}$$

where $i = 1, \ldots, m$ indexes the different small areas of interest and $j \in \mathcal{S}_i$ indexes the sampled units in small area i. Here, the model errors, v_i, are i.i.d. $N(0, \sigma_v^2)$ random variables, and the sampling errors, e_{ij}, are i.i.d. $N(0, \sigma_e^2)$ random variables, independent of the model errors.

Let $\boldsymbol{V}_i$ be the covariance matrix consisting of diagonal elements $\sigma_v^2 + \sigma_e^2/n_i$ and off-diagonal elements σ_v^2. Assuming (2) holds for the sampled units, and the variance parameters σ_v^2 and σ_e^2 are known, the best linear unbiased predictor of $\bar{y}_i = \sum_{j \in \mathcal{U}_i} y_{ij}/N_i$ is

$$\hat{\bar{y}}_i = \frac{1}{N_i} \sum_{j \in \mathcal{S}_i} y_{ij} + \frac{1}{N_i} \sum_{j \in \mathcal{S}_i^c} \left(\boldsymbol{x}_{ij}^T \tilde{\boldsymbol{\beta}} + \tilde{v}_i \right), \tag{3}$$

where

$$\tilde{\boldsymbol{\beta}} = \left(\sum_{i=1}^{m} \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{X}_i \right)^{-1} \left(\sum_{i=1}^{m} \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{y}_i \right).$$

$\boldsymbol{X}_i$ is the $n_i \times p$ matrix with rows $\boldsymbol{x}_{ij}^T$, and $\tilde{v}_i = (\sigma_v^2/(n_i\sigma_v^2 + \sigma_e^2))$ $\sum_{j \in \mathcal{S}_i} (y_{ij} - \boldsymbol{x}_{ij}^T \tilde{\boldsymbol{\beta}})$. In (3), as in the general expression in (1), the unobserved y_{ij} are replaced by model predictions. Note that evaluation of (3) requires knowledge of the population mean, $\bar{\boldsymbol{X}}_{ip} = \sum_{j \in \mathcal{U}_i} \boldsymbol{X}_{ij}/N_i$, of the covariates.

In practice, the variance components σ_v^2 and σ_e^2 are unknown and need to be estimated. The empirical best linear unbiased predictor (EBLUP) is obtained by substituting estimates, $\hat{\sigma}_e^2$ and $\hat{\sigma}_v^2$, (typically MLE, REML, or moment estimates) of the variance components in the above expressions (Prasad and Rao 1990). Stukel and Rao (1999) derive unbiased method of moment estimates of the variance components as well as mean squared error estimates of the EBLUP, and Hall and Maiti (2006) give bootstrap methods for mean squared error estimation. In addition, this model can easily be fit using Bayesian hierarchical modeling rather than using the EBLUP, which would incorporate the uncertainty from the variance parameters. Datta and Ghosh (1991) developed a Bayesian version of the nested error regression model (2), using a uniform prior on the regression coefficients and gamma priors on the variance components.

The survey weights do not enter into either the nested error regression model (2) or the EBLUPs of the small area means (3). Because of this, the EBLUP is not design-consistent, unless the sampling design is self-weighting within each small area (Rao and Molina 2015).

3.2 Including Design Variables in the Model

Suppose now that the survey design is informative, so that the way in which individuals are selected in the sample depends in an important way on the

value of the response variable y_{ij}. It is well established that when the survey design is informative, that ignoring the survey design and performing unweighted analyses without adjustment can result in substantial biases (Nathan and Holt 1980; Pfeffermann and Sverchkov 2007).

One method to eliminate the effects of an informative design is to condition on all design variables (Gelman et al. 1995, Chapter 7). To see this, decompose the response variables as $\boldsymbol{y} = (\boldsymbol{y}_s, \boldsymbol{y}_{ns})$, where $\boldsymbol{y}_s$ are the observed responses for the sampled units in the population, and $\boldsymbol{y}_{ns}$ represents the unobserved variables corresponding to nonsampled individuals. Let $\boldsymbol{I}$ be the vector of sample inclusion variables, so that $I_{ij} = 1$ if y_{ij} is observed and $I_{ij} = 0$ otherwise. The observed data likelihood, conditional on covariate information $\boldsymbol{X}$, and model parameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, are then

$$\begin{aligned} f(\boldsymbol{y}_s, \boldsymbol{I}|\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{\phi}) &= \int f(\boldsymbol{y}_s, \boldsymbol{y}_{ns}, \boldsymbol{I}|\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{\phi}) d\boldsymbol{y}_{ns} \\ &= \int f(\boldsymbol{I}|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\phi}) f(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\theta}) d\boldsymbol{y}_{ns}. \end{aligned} \tag{4}$$

If $f(\boldsymbol{I}|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\phi}) = f(\boldsymbol{I}|\boldsymbol{X}, \boldsymbol{\phi})$, the inclusion variables $\boldsymbol{I}$ are independent of $\boldsymbol{y}$, conditional on $\boldsymbol{X}$, and the survey design can be ignored. For example, if the design variables $\boldsymbol{Z}$ are included in $\boldsymbol{X}$, the ignorability condition may hold, and inference can be based on $f(\boldsymbol{y}_s|\boldsymbol{X}, \boldsymbol{\theta})$.

Little (2012) advocates for a general framework using unit-level Bayesian modeling that incorporates the design variables. For example, if cluster sampling is used, one could incorporate a cluster-level random effect into the model, or if a stratified design is used, one might incorporate fixed effects for the strata. The idea is that when all design variables are accounted for in the model, the conditional distribution of the response given the covariates for the sampled units is independent of the inclusion probabilities. Because the model is unit-level and Bayesian, the unsampled population can be generated via the posterior predictive distribution. Doing so provides a distribution for any finite population quantity and incorporates the uncertainty in the parameters. For example, if the population response is generated at draw k of a Markov chain Monte Carlo algorithm, $\boldsymbol{y}^{(k)}$, then one has implicitly generated a draw from the posterior distribution of the population mean for a given area i:

$$\bar{y}_i^{(k)} = \frac{\sum_{j=1}^{N} y_j^{(k)} I(j \in \mathcal{U}_i)}{\sum_{j=1}^{N} I(j \in \mathcal{U}_i)}.$$

If there are K total posterior draws, one could then estimate the mean and standard error of $\bar{y}_i$ with

$$\widehat{\bar{y}}_i = \frac{1}{K}\sum_{k=1}^{K} \bar{y}_i^{(k)}$$

and

$$\widehat{\mathrm{SE}(\bar{y}_i)} = \sqrt{\mathrm{Var}(\widehat{\bar{y}}_i)},$$

where $\mathrm{Var}(\widehat{\bar{y}}_i)$ is the sample variance of $\bar{y}_i^{(k)}$.

The problem with attempting to eliminate the effect of the design by conditioning on design variables is often more of a practical one, because neither the full set of design variables nor the functional relationship between the design and the response variables will be fully known. Furthermore, expanding the model by including sufficient design information so as to ignore the design may make the likelihood extremely complicated or even intractable.

3.3 Poststratification

Little (1993) gives an overview of poststratification. Consider the case where $f(\boldsymbol{I}|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\phi}) = f(\boldsymbol{I}|\boldsymbol{X}, \boldsymbol{\phi})$ in (4). Then, the model $f(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\theta})$ holds for both the sample and the population and may be used to generate predictions for unsampled units. To perform poststratification, the population is assumed to contain C categories, or poststratification cells, defined by $\boldsymbol{X}$, such that within each category units are independent and identically distributed. Usually, these categories are cross-classifications of categorical predictor variables such as county, race, and education level. When a regression model is fit relating the response to the predictors, predictions can be generated for each unit within a cell, and thus for the entire population. Importantly, any desired aggregate estimates can easily be generated from the unit-level population predictions.

Gelman and Little (1997) and Park et al. (2006) develop a framework for poststratification via hierarchical modeling. By using a hierarchical model with partial pooling, parameter estimates can be made for poststratification cells without any sampled units, and variance is reduced for cells having few sampled units. Gelman and Little (1997) and Park et al. (2006) provide an example for binary data that uses the following model

$$\begin{aligned} y_{ij}|p_{ij} &\sim \mathrm{Bernoulli}(p_{ij}) \\ \mathrm{logit}(p_{ij}) &= \boldsymbol{x}_{ij}'\boldsymbol{\beta} \\ \boldsymbol{\beta} &= (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_G)' \\ \boldsymbol{\beta}_g &\overset{\text{ind}}{\sim} N_{c_g}(0, \sigma_g^2 \boldsymbol{I}_{c_g}), g = 1, \ldots, G, \end{aligned} \tag{5}$$

where $\boldsymbol{x}_{ij}$ is a vector of dummy variables for G categorical predictor variables with c_g classes in variable g.

Bayesian inference can be performed on this model, leading to a probability, $p(y_{ij} = 1|p_{ij}) = p_{ij} = p_i, \forall j$, that is constant within each cell $i = 1, \ldots, C$. The number of positive responses within cell i can be estimated with $N_i p_i$, and any higher-level aggregate estimates can be made by aggregating the corresponding cells. Depending on the data at hand, it is straightforward to replace the Bernoulli data model in (5) with another data model. In some scenarios, the number of units within each cell may not be known, in which case further modeling would be necessary. After estimating the model parameters, predictions can be made for every unit in the population, in essence creating a synthetic population. Aggregation of this synthetic population is what allows for SAE. Finally, Gao et al. (2021) explore the use of structured priors within a hierarchical modeling and poststratification framework to reduce bias from highly unrepresentative samples.

4. MODELS WITH SURVEY WEIGHT ADJUSTMENTS

Although many of the models in section 3 can be used to handle informative sampling, they do not rely on the survey weights. In this section, we explore techniques that rely on the weights to adjust the sample likelihood.

There have been several methods proposed in the literature which make use of the nested error regression model (2), but which incorporate the survey weights, either as regression variables or as adjustments to the predicted values, so as to protect against a possible informative survey design.

4.1 Survey Weight Adjustments to the Basic Unit-Level Model

Verret et al. (2015) augmented the nested error regression model (2), by including functions of the inclusion probabilities, $g(\pi_{ij})$, as predictors. Care must be taken in the choice of the function g, as the population means, $\bar{G}_i = \sum_{j \in \mathcal{U}_i} g(\pi_{ij})/N_i$, must be known to obtain the EBLUPs from (3). Some suggestions for the choice of g were $g(\pi_{ij}) = \pi_{ij}$, which gives $\bar{G}_i = n_i/N_i$, and $g(\pi_{ij}) = n_i/\pi_{ij}$, which gives $\bar{G}_i = n_i \sum_{j \in \mathcal{U}_i} w_{ij}/N_i$, which may be known in practice. Verret et al. (2015) reported strong performance of the EBLUP using the augmented nested error regression model, in a probability proportional to size simulation study, in terms of bias and mean squared error, for properly chosen augmenting variable g. However, some choices of g, such as $g(\pi_{ij}) = w_{ij}$, could lead to poor performance, except under non-informative sampling, when the chosen function differs from the true relationship. Verret et al. (2015) suggested using scatter plots of residuals from the nested error regression model against different choices of augmenting variables to choose an appropriate model. An alternative to exploring a collection of augmenting variables is to estimate the functional form of g. Zheng and Little (2003) investigated nonparametric estimation of g using penalized splines and found that

predictions of small area means using this modeling framework resulted in large gains in mean squared error over the design-based estimates in their simulation studies.

You and Rao (2002) proposed a pseudo-EBLUP of the small area means $\theta_i = \bar{\boldsymbol{X}}_i^T \boldsymbol{\beta} + v_i$ based on the nested error regression model (2), which incorporates the survey weights. In their approach, the regression parameters $\boldsymbol{\beta}$ in (2) are estimated by solving a system of survey-weighted estimating equations

$$\sum_{i=1}^{m} \sum_{j \in \mathcal{S}_i} w_{ij} \boldsymbol{x}_{ij} \{ y_{ij} - \boldsymbol{x}_{ij}^T \boldsymbol{\beta} - \gamma_{iw} (\bar{y}_{iw} - \bar{\boldsymbol{x}}_{iw}^T \boldsymbol{\beta}) \} = 0, \tag{6}$$

where $\gamma_{iw} = \sigma_v^2 / (\sigma_v^2 + \sigma_\epsilon^2 \delta_i^2)$, $\delta_i^2 = \sum_{j \in \mathcal{S}_i} w_{ij}^2$, $\bar{y}_{iw} = \sum_{j \in \mathcal{S}_i} w_{ij} y_{ij} / \sum_{j \in \mathcal{S}_i} w_{ij}$, and $\bar{\boldsymbol{x}}_{ij} = \sum_{j \in \mathcal{S}_i} w_{ij} \boldsymbol{x}_{ij} / \sum_{j \in \mathcal{S}_i} w_{ij}$. This is an example of the pseudo-likelihood approach to incorporating survey weights, which is later discussed in more detail.

The pseudo-BLUP $\tilde{\boldsymbol{\beta}}_w = \tilde{\boldsymbol{\beta}}_w(\sigma_e^2, \sigma_v^2)$ is the solution to (6) when the variance components σ_e^2 and σ_v^2 are known, and the pseudo-EBLUP, $\widehat{\boldsymbol{\beta}}_w = \tilde{\boldsymbol{\beta}}(\widehat{\sigma}_e^2, \widehat{\sigma}_v^2)$, is the solution to (6) using plug-in estimates $\widehat{\sigma}_e^2$ and $\widehat{\sigma}_v^2$ of the variance components. The pseudo-EBLUP, $\widehat{\theta}_i$ of the small area mean θ_i is then

$$\widehat{\theta}_{iw} = \widehat{\gamma}_{iw} \bar{y}_{iw} + (\bar{X}_i - \widehat{\gamma}_{iw} \bar{x}_{iw})^T \widehat{\boldsymbol{\beta}}_w.$$

Similar to Battese et al. (1988), You and Rao (2002) assumed an ignorable survey design, so that the model (2) holds for both the sampled and nonsampled units. However, You and Rao (2002) showed that inclusion of the survey weights in the pseudo-EBLUP results in a design-consistent estimator. In addition, when the survey weights are calibrated to the population total, so that $\sum_{j \in \mathcal{S}_i} w_{ij} = N_i$, the pseudo-EBLUP has a natural benchmarking property, without any additional adjustment, in the sense that

$$\sum_{i=1}^{m} N_i \widehat{\theta}_{iw} = \widehat{Y}_w + (\boldsymbol{X} - \widehat{\boldsymbol{X}}_w)^T \widehat{\boldsymbol{\beta}}_w,$$

where $\widehat{Y}_w = \sum_{i=1}^{m} \sum_{j \in \mathcal{S}_i} w_{ij} y_{ij}$ and $\widehat{\boldsymbol{X}}_w = \sum_{i=1}^{m} \sum_{j \in \mathcal{S}_i} w_{ij} \boldsymbol{x}_{ij}$. That is, the weighted sum of area-level pseudo-EBLUPs is equal to a generalized regression estimator of the population total. See also Zimmerman and Münnich (2018), which extends the pseudo-EBLUP method of You and Rao (2002) to lognormal sampling models for skewed business survey data.

An alternative pseudo-EBLUP, which is applicable to the estimation of general small area parameters beyond the small area means, was proposed in Guadarrama et al. (2018) (see also Jiang and Lahiri, 2006). Rather than use the genuine best predictor in (1), which conditions on all observed data, Guadarrama et al. (2018) suggested a pseudo-best predictor, which conditions only on the survey-weighted Horvitz–Thompson estimator, $\bar{y}_{iw} = \sum_{j \in \mathcal{S}_i} w_{ij} y_{ij} /$

$\sum_{j \in \mathcal{S}_i} w_{ij}$, of the small area means. Assuming that the nested error regression model (2) holds for all units in the population, there is a simple, closed-form expression for the predictions of out-of-sample variables, y_{ij}, given by

$$E(y_{ij}|\bar{y}_{iw}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \gamma_{iw}(\bar{y}_{iw} - \bar{\boldsymbol{x}}_{iw}^T\boldsymbol{\beta}),$$

using the same notation as in (6). This idea can easily be extended for prediction of general additive parameters, $H_i = \sum_{j \in \mathcal{U}_i} h(y_{ij})/N_i$, by using the conditional expectation $E\left(h(y_{ij})|\bar{y}_{iw}\right)$ in place of the out-of-sample variables.

4.2 Joint Regression and Prediction of the Survey Weights

Prediction of small area quantities using (1) requires estimation of $E(y_{ij}|D_s)$ for all nonsampled units in the population. One of the main difficulties in using unit-level model-based methods is the lack of knowledge of the covariates, sampling weights, or population sizes associated with the nonsampled units and small areas that are needed to make these model-based predictions. To overcome this difficulty, Si et al. (2015) modeled the observed poststratification cells n_i, conditional on $n = \sum_{i=1}^m n_i$, using the multinomial distribution

$$(n_1, \ldots, n_m) \sim \text{Multinomial}\left(n; \frac{N_1/w_1}{\sum_{i=1}^m N_i/w_i}, \ldots, \frac{N_m/w_m}{\sum_{i=1}^m N_i/w_i}\right)$$

for poststratification cells $i = 1, \ldots, m$.

This model assumes that the unique values of the sample weights determine the poststratification cells and that the sampling weight and response are the only values known for sampled units. The authors state that, in general, this assumption is untrue, because there will be cells with a low probability of selection that do not show up in the sample, but the assumption is necessary to proceed with the model. This model yields a posterior distribution over the cell population sizes which can be used for poststratification with their response model, which uses a nonparametric Gaussian process regression on the survey weights,

$$\begin{gathered} y_{ij}|\mu(w_i), \sigma^2 \sim \text{N}(\mu(w_i), \sigma^2) \\ \mu(w_i)|\beta, C(w_i, w_{i'}|\boldsymbol{\theta}) \sim \text{GP}(w_i\beta, C(w_i, w_{i'}|\boldsymbol{\theta})) \\ \pi(\sigma^2, \beta, \boldsymbol{\theta}), \end{gathered}$$

for observation j in poststratification cell i. Here, *GP* denotes a Gaussian process and $C(\cdot, \cdot|\boldsymbol{\theta})$ represents a valid covariance function that depends on

parameters $\boldsymbol{\theta}$. The authors use a squared exponential function, but other covariance functions could be used in its place. The normal distribution placed over y_{ij} could be replaced with another distribution in the case of non-Gaussian data. Specifically, the authors explore the Bernoulli response case. This model implicitly assumes that units with similar weights will tend to have similar response values, which is likely not true in general. However, in the absence of any other information about the sampled units, this may be the most practical assumption. Because Si et al. (2015) assume that only the survey weights and response values are known, this methodology cannot be used for SAE as presented. However, the model can be extended to include other variables such as a geographic indicator, which would allow for area-level estimation.

Vandendijck et al. (2016) extend the work of Si et al. (2015) to be applied to SAE. They assume that the poststratification cells are designated by the unique weights within each area. Rather than using the raw weights, they use the weights scaled to sum to the sample size within each area. They then use a similar multinomial model to Si et al. (2015) to perform poststratification using the posterior distribution of the poststratification cell population sizes. Assuming a Bernoulli response, they use the data model

$$\begin{aligned} y_{ij}|\eta_{ij} &\sim \text{Bernoulli}(\eta_{ij}) \\ \text{logit}(\eta_{ij}) &= \beta_0 + \mu(\tilde{w_{ij}}) + u_i + v_i \end{aligned}$$

for unit j in small area i, with $\tilde{w_{ij}}$ designating the survey weights scaled to sum to the area-specific sample size. Independent area-level random effects are denoted by v_i, whereas u_i denotes spatially dependent area-level random effects, for which the authors use the prior,

$$u_i|u_j, j \in ne(i) \sim N(\bar{u}_i, \sigma_u^2/n_i). \tag{7}$$

Here, $ne(i)$ is the set of neighbors of area i and $\bar{u}_i$ is the mean of the neighboring spatial effects. The spatial model in (7) is known as the intrinsic conditional autoregressive (ICAR) model (Besag 1974). They explore the use of a Gaussian process prior over the function $\mu(\cdot)$ as well as a penalized spline approach. For their Gaussian process prior, they assume a random walk of order one. The multinomial model

$$(n_{1i}, \ldots, n_{L_i i}) \sim \text{Multinomial}\left(n_i; \frac{N_{1i}/w_{(1)i}}{\sum_{l=1}^{L_i} N_{li}/w_{(l)i}}, \ldots, \frac{N_{L_i i}/w_{(L_i)i}}{\sum_{l=1}^{L_i} N_{li}/w_{(l)i}}\right)$$

is used for poststratification, where n_{li} and N_{li} represent the known sample size and unknown population size respectively for poststrata cell l in area i. The cells are determined by the unique weights in area i, with the value of the weight represented by $w_{(l)i}$. Although Vandendijck et al. (2016) implement their model with a Bernoulli data example, this is a type of a generalized linear

model, and thus, other response types in the exponential family may be used as well.

4.3 Multilevel Models for Clustered Data

Suppose the finite population is hierarchically structured, for example, by area, by strata, or by cluster. In this section, we consider sampling designs which take into account this hierarchical structure through use of a two-stage sampling design. Let $\mathcal{U} = \{1, \ldots, m\}$ be an enumeration of the first level, primary sampling units (PSUs). A sample, $\mathcal{S} \subset \mathcal{U}$, is selected with probabilities π_i. Define $w_i = 1/\pi_i$ to be the first-stage sampling weights. Let $\mathcal{U}_i$ be an enumeration of the secondary sampling units (SSUs) within PSU i, and let $\pi_{j|i}$ be the probability that SSU j is sampled, given that PSU i has been sampled. Define $w_{j|i}$ to be the second-stage survey weights, and let y_{ij} be a characteristic of interest associated with SSU j within PSU i. It is possible that the sample design is informative at one stage, both stages, or neither stage.

The pairs (i, j) can be identified with a single index, k, and the unconditional inclusion probabilities and the unconditional survey weights can be written $\pi_k = \pi_{ij} = \pi_i \pi_{j|i}$ and $w_k = w_{ij} = w_j w_{j|i}$, respectively. If the values y_k can be modeled as independent random variables from a parametric superpopulation model, $f_p = f_p(y|\boldsymbol{\theta})$, pseudo-likelihood methods, introduced by Binder (1983) and Skinner (1989), can be used for inference on the superpopulation parameter $\boldsymbol{\theta}$. The pseudo-log-likelihood is defined as

$$\sum_{k \in \mathcal{S}} w_k \log f_p(y_k|\boldsymbol{\theta}); \tag{8}$$

this is simply the Horvitz–Thompson estimator of the population-level log-likelihood $\sum_{k \in \mathcal{U}} \log f_p(y_k|\boldsymbol{\theta})$. The pseudo-maximum likelihood estimator (pMLE), $\widehat{\boldsymbol{\theta}}$, is obtained by maximizing (8), or equivalently, by solving the system

$$\widehat{U}(\boldsymbol{\theta}) = \sum_{k \in \mathcal{S}} w_k \frac{\partial}{\partial \boldsymbol{\theta}} \log f_p(y_k|\boldsymbol{\theta}) = 0.$$

Under regularity conditions, $\widehat{\boldsymbol{\theta}}$ is design and model consistent for $\boldsymbol{\theta}$ under both informative and noninformative sampling designs (Binder 1983; Godambe and Thompson 1986).

Wang et al. (2018) proposed a Bayesian method for finite population inference using pseudo-likelihood methods, which is valid under informative sampling designs. Their method uses the sampling distribution of the pMLE (or more generally, the solution to a survey-weighted estimating equation) and constructs an approximate posterior distribution

$$p\left(\boldsymbol{\theta}|\widehat{U}(\boldsymbol{\theta})\right) \propto g\left(\widehat{U}(\boldsymbol{\theta})|\boldsymbol{\theta}\right)\pi(\boldsymbol{\theta}), \tag{9}$$

where $g\left(\widehat{U}(\boldsymbol{\theta})|\boldsymbol{\theta}\right)$ is the limiting Gaussian distribution of the estimating function $\widehat{U}(\boldsymbol{\theta})$, and $\pi(\boldsymbol{\theta})$ is the prior distribution of $\boldsymbol{\theta}$. Wang et al. (2018) gave a Bernstein–von Misses Theorem for the approximate posterior distribution (9) and showed consistency of the Bayesian estimator based on (9).

When the finite population has a hierarchical structure, there could be dependence between response variables y_{ij} and y_{ik} belonging to a common cluster i. This dependence can be modeled by introducing random effects, v_i, which are shared by units within a cluster, in the superpopulation model $f_p = f_p(y|\boldsymbol{\theta}, v_i)$. The usual choice for the distribution of the random effects, $\varphi(v|\sigma^2)$, is the mean zero normal distribution with unknown variance σ^2. The presence of random effects, the multilevel structure of superpopulation model, and the dependence of variables within a common cluster mean that neither the pseudo-likelihood method nor the related estimating equation approach can be directly applied to SAE problems. However, Grilli and Pratesi (2004), Asparouhov (2006), and Rabe-Hesketh and Skrondal (2006) extended the pseudo-likelihood approach to accommodate models with hierarchical structure by making use of the population log-likelihood and the decomposed survey weights, w_i and $w_{j|i}$.

The census marginal log-likelihood is obtained by integrating out the random effects from the likelihood:

$$\begin{aligned} \log L(\boldsymbol{\theta}) &= \sum_{i=1}^{m} \log \int \prod_{j\in\mathcal{U}_i} f_p(y_{ij}|\boldsymbol{\theta}, v)\varphi(v|\sigma^2)dv \\ &= \sum_{i=1}^{m} \log \int \exp\left\{\sum_{j\in\mathcal{U}_i} \log f_p(y_{ij}|\boldsymbol{\theta}, v)\right\}\varphi(v|\sigma^2)dv. \end{aligned} \tag{10}$$

The pseudo-log-likelihood for the multilevel model can be defined by replacing $\sum_{j\in\mathcal{U}_i} \log f_p(y_{ij}|\boldsymbol{\theta}, v_i)$ in (10) by the design-unbiased estimate, $\sum_{j\in\mathcal{S}_i} w_{j|i} \log f_p(y_{ij}|\boldsymbol{\theta}, v_i)$, to get

$$\log \widehat{L}(\boldsymbol{\theta}) = \sum_{i=1}^{m} w_i \log \int \exp\left\{\sum_{j\in\mathcal{S}_i} w_{j|i} \log f_p(y_{ij}|\boldsymbol{\theta}, v)\right\}\varphi(v|\sigma^2)dv. \tag{11}$$

Analytical expressions for the maximizer of (11) generally do not exist, so the pMLE, $\widehat{\boldsymbol{\theta}}$, must be found by numerical maximization of (11). Grilli and Pratesi (2004) used the `NLMIXED` procedure within SAS, using appropriately adjusted weights in the `replicate` statement and a bootstrap for mean squared error estimation. Rabe-Hesketh and Skrondal (2006) used an adaptive quadrature

routine using the `gllamm` program within Stata and derived a sandwich estimator of the standard errors, finding good coverage in their simulation studies with this estimate. Kim et al. (2017) proposed an EM algorithm for parameter estimation. Their method involves two steps, where first the random effects are treated as fixed, and a profile likelihood maximum likelihood estimator of the random effects are computed. The second step uses the EM algorithm to estimate the remaining model parameters. Their method relies on a normal approximation to the predictive distribution of the random effects but was found to give good results with moderate cluster sizes in numerical studies. Kim et al. (2017) also gave a method for predicting random effects using the EM algorithm and an approximating predictive distribution that was shown to be valid for sufficiently large cluster sizes, which is needed for the prediction of unobserved variables.

Eideh and Nathan (2009) and Rao et al. (2013) noted that both design consistency and design-model consistency of the variance component, σ^2, require that both the number of areas (or clusters), m, and the number of elements within each cluster, n_i, tend to infinity when the survey design is informative. Furthermore, the relative bias of the estimators can be large when the n_i are small. Rao et al. (2013) showed that consistency of the variance component can be achieved with only m tending to infinity (allowing the n_i to be small) if the joint inclusion probabilities, $\pi_{jk|i} = P(j, k \in \mathcal{S}_i | i \in \mathcal{S})$, are available. Their method uses the marginal joint densities,

$$L_{ijk} = L_{ijk}(\boldsymbol{\theta}, \sigma^2) = \int f_p(y_{ij}|\boldsymbol{\theta}, v) f_p(y_{ij}|\boldsymbol{\theta}, v) \varphi(v|\sigma^2) dv,$$

and estimates $\boldsymbol{\theta}$ and σ^2 by maximizing the design-weighted composite log likelihood

$$l_{wC}(\boldsymbol{\theta}, \sigma^2) = \sum_{i \in \mathcal{S}} w_i \sum_{j<k \in \mathcal{S}_i} w_{jk|i} \log L_{ijk}. \tag{12}$$

It was shown in Yi et al. (2016) that the maximizer of (12), is consistent for both $\boldsymbol{\theta}$ and σ^2, with respect to the joint superpopulation model and the sampling design.

Valid inference using multilevel models when the survey design is informative requires access to two sets of survey weights, w_i and $w_{j|i}$, and in the case of the method of Rao et al. (2013), higher-order inclusion probabilities, $\pi_{jk|i}$, which are not typically provided; in most situations, the analyst will have only the single-inclusion probabilities, π_k, and it seems to be an open question as to whether these single-inclusion probabilities alone are sufficient for valid inference with multilevel models and informative sampling designs. There has been some work in this direction for the specific case of the one-way ANOVA superpopulation model, which has been used by many authors as a

superpopulation model for clustered data (Scott and Smith 1969; Eideh 2012; Slud 2020). The one-way ANOVA model is given by

$$Y_{ij} = \mu + v_i + e_{ij},$$

where μ is a fixed intercept, $v_i \overset{i.i.d.}{\sim} N(0, \sigma_v^2)$, and $e_{ij} \overset{i.i.d.}{\sim} N(0, \sigma_e^2)$. Recently, Slud (2020) gave an expectation-maximization algorithm for consistent estimation of variance components in the one-way ANOVA model, using only single-inclusion weights, w_i and $w_{j|i}$, when the clusters are sampled using an informative sampling design, so long as the units within clusters are sampled using a noninformative sampling design. Furthermore, Slud (2020) showed that none of the existing methods consistently estimate variance components in the one-way ANOVA model using only single-inclusion weights when both the clusters, and the units within clusters, are sampled using an informative design.

The pseudo-log-likelihoods (8) and (11) suggest pseudo-likelihoods

$$\prod_{i}^{m} \prod_{j \in \mathcal{S}_i} f(y_{ij}|\boldsymbol{\theta})^{w_{ij}} \tag{13}$$

for single level models, and

$$\prod_{i=1}^{m} \left\{ \int \prod_{j \in \mathcal{S}_i} f(y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{\theta}, v_j)^{w_{j|i}} \phi(v_j) dv_j \right\}^{w_i}$$

for multilevel models (Asparouhov 2006). The pseudo-likelihood (13) is sometimes called the composite likelihood in general statistical problems, when the weights w_{ij} (not necessarily survey weights) are known positive constants, and its use is popular in problems where the exact likelihood is intractable or computationally prohibitive (Varin et al. 2011).

The pseudo-likelihood (13) is not a genuine likelihood, as it does not incorporate the dependence structure in the sampled data nor the relationship between the responses and the design variables beyond the inclusion of the survey weights. However, the pseudo-likelihood has been shown to be a useful tool for likelihood analysis for finite population inference in both the frequentist and Bayesian frameworks.

By treating the pseudo-likelihood as a genuine likelihood, and specifying a prior distribution $\pi(\boldsymbol{\theta})$ on the model parameters $\boldsymbol{\theta}$, Bayesian inference can be performed on $\boldsymbol{\theta}$. For general models, Savitsky and Toth (2016) showed for certain sampling schemes, and for a class of population distributions, that the pseudo-posterior distribution using the survey-weighted pseudo-likelihood, with survey weights scaled to sum to the sample size, (13) converges in L^1 to the population generating distribution. This result justifies the use of (13) in place of the likelihood in Bayesian analysis of population parameters,

conditional on the observed sampled units, even when the sample design is informative. Predictions of area-level random effects as well as predictions of nonsampled units can then be made as well. Williams and Savitsky (2020) broaden the class of sample designs for which the Bayesian pseudo-likelihood converges to the population-generating distribution, including designs with unattenuated dependence within clusters.

Savitsky and Toth (2016) focus on parameter inference and do not give any advice for making area-level estimates. However, it is straightforward to implement a model with a Bayesian pseudo-likelihood and then apply poststratification after the fact by generating the population, and thus any desired area-level estimates using (1). This type of pseudo-likelihood with poststratification for SAE was demonstrated in the frequentist setting by Zhang et al. (2014).

Ribatet et al. (2012) provide a discussion on the validity of Bayesian inference using the composite likelihood (13) in place of the exact likelihood in Bayes' formula. An example of this method used in the sample survey context can be found in Dong et al. (2014), which used a weighted pseudo-likelihood with a multinomial distribution as a model for binned response variables. They assumed an improper Dirichlet distribution on the cell probabilities and used the associated posterior and posterior predictive distributions for the prediction of the nonsampled population units.

5. LIKELIHOOD-BASED INFERENCE USING THE SAMPLE DISTRIBUTION

The pseudo-likelihood methods discussed in section 4 require the specification of a superpopulation model, which is a distribution, which holds for all units in the finite population. However, validating the superpopulation model based on the observed sampled values is challenging unless the sampling design is not informative, in which case, the distribution for the sampled units is the same as for the nonsampled units. Under an informative sampling design, the model for the population data does not hold for the sampled data. This can be seen by the application of Bayes' theorem. Suppose the finite population values y_{ij} are independent realizations from a population with density $f_p(\cdot|\boldsymbol{x}_{ij}, \boldsymbol{\theta})$, conditional on a vector of covariates $\boldsymbol{x}_{ij}$ and model parameters $\boldsymbol{\theta}$. Given knowledge of this superpopulation model, as well as the distribution of the inclusion variables, the distribution of the sampled values can be derived. Define the sample density, f_s, (Pfeffermann et al. 1998) as the density function of y_{ij}, given that y_{ij} has been sampled, that is,

$$f_s(y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{\theta}, \boldsymbol{\gamma}) = f_p(y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{\theta}, I_{ij} = 1) = \frac{P(I_{ij} = 1|y_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{\gamma}) f_p(y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{\theta})}{P(I_{ij} = 1|\boldsymbol{x}_{ij}, \boldsymbol{\gamma})}, \quad (14)$$

where I_{ij} is a binary variable indicating whether unit j in area i has been sampled and $\boldsymbol{\gamma}$ is a vector of parameters related to the sampling model. From

(14), the sample distribution differs from the population distribution, unless $P(I_{ij} = 1|y_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{\gamma}) = P(I_{ij} = 1|\boldsymbol{x}_{ij}, \boldsymbol{\gamma})$, which occurs in ignorable sampling designs. Note that the inclusion probabilities, π_{ij}, may differ from the probabilities $P(I_{ij} = 1|\boldsymbol{x}_{ij}, y_{ij}, \boldsymbol{\gamma})$ in (14), because the latter are not conditional on the design variables $\boldsymbol{Z}$.

Equation (14) can be used for likelihood-based inference if the simplifying assumption that the sampled values are independent is made. While this is not true in general, asymptotic results given in Pfeffermann et al. (1998) justify an assumption of independence of the data for certain sampling schemes when the overall population size is large, and the sample size remains fixed. More recently, Bonnéry et al. (2018) gave precise asymptotic results for the sample maximum likelihood estimator of (14) when both the sample size and the population size increase.

Direct use of (14) for finite population inference requires additional model specifications for the sample inclusion variables $P(I_{ij} = 1|\boldsymbol{x}_{ij}, y_{ij}, \boldsymbol{\gamma})$ as well as $P(I_{ij} = 1|\boldsymbol{x}_{ij}, \boldsymbol{\gamma})$. It was shown in Pfeffermann et al. (1998) that $P(I_{ij} = 1|\boldsymbol{x}_{ij}, y_{ij}, \boldsymbol{\gamma}) = E_p(\pi_{ij}|\boldsymbol{x}_{ij}, y_{ij}, \boldsymbol{\gamma})$, and that $P(I_{ij} = 1|\boldsymbol{x}_{ij}, \boldsymbol{\gamma}) = E_p(\pi_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{\gamma})$, so that a superpopulation model still needs to be specified for likelihood-based inference.

Ideally, one would like to specify a model for the sampled data, and to use this model fit to the sampled data to infer the nonsampled values, without explicit specification of a superpopulation model. Pfeffermann and Sverchkov (2007) showed how to predict small area means by only identifying models for the sampled data. Their work makes use of an important identity derived in Pfeffermann and Sverchkov (1999), which links the moments of the sample and population moments. They showed that

$$P(I_{ij} = 1|y_{ij}, \boldsymbol{x}_{ij}) = E_p(\pi_{ij}|y_{ij}, \boldsymbol{x}_{ij}) = 1/E_s(w_{ij}|y_{ij}, \boldsymbol{x}_{ij}).$$

Similarly, it was shown that

$$P(I_{ij} = 1|\boldsymbol{x}_{ij}) = E_p(\pi_{ij}|\boldsymbol{x}_{ij}) = 1/E_s(w_{ij}|\boldsymbol{x}_{ij}).$$

Combining these results with an application of Bayes' theorem, as was done to arrive at (14), gives the distribution for the nonsampled units in the finite population

$$f_c(y_{ij}|\boldsymbol{x}_{ij}) \equiv f_p(y_{ij}|\boldsymbol{x}_{ij}, I_{ij} = 0) = \frac{E_s(w_{ij} - 1|y_{ij}, \boldsymbol{x}_{ij}) f_s(y_{ij}|\boldsymbol{x}_{ij})}{E_s(w_{ij} - 1|\boldsymbol{x}_{ij})}, \qquad (15)$$

where f_c represents the density function of y_{ij}, given that y_{ij} has not been sampled. This result allows one to specify only a distribution for the sampled responses and a distribution for the sampled survey weights for inference on the nonsampled units, without any hypothetical distribution for the finite population. Importantly, this allows for the identification of the finite population

generating distribution f_p through the sample-based likelihood. It also establishes the relationship between the moments of the sample distribution and the population distribution, allowing for the prediction of nonsampled units.

Pfeffermann and Sverchkov (2007) adapted the sample distribution to multilevel models for SAE of finite population means when both small areas and units within small areas are sampled with unequal probabilities in a two-sage, informative survey design. Following the notation of section 4.3, suppose there are area-specific random effects, v_i, which are shared by all units in the population in small area i, so that the population distribution can be written $f_p(y_{ij}|\boldsymbol{x}_{ij}, v_i, \boldsymbol{\theta})$. Let $f_p(v)$ be the population pdf of the random effects v_i. The first level of the multilevel sample models is

$$f_s(v_i) = f(v_i|I_i = 1) = P(I_i = 1|v_i)f_p(v_i)/P(I_i = 1),$$

where I_i is a binary variable indicating area i has been sampled. The second level is the same as in (14), but with the random effects, v_i, included. Pfeffermann and Sverchkov (2007) showed how small area means can be predicted using the observed unit-level data under a multilevel, informative survey design. Under mild assumptions, Pfeffermann and Sverchkov (2007) showed that

$$E_p(\bar{y}_i|D_s, I_i = 1) = \frac{1}{N_i}\left(\sum_{j\in\mathcal{S}_i} y_{ij} + \sum_{j\notin\mathcal{S}_i} E_s\big(E_c(y_{ij}|\boldsymbol{x}_{ij}, v_i, I_i = 1)|D_s\big)\right).$$

Combining this with (14) and (15) allows for the prediction of the small area means after specification of a model for the sampled responses, $f_s(y_{ij}|\boldsymbol{x}_{ij}, v_i)$, and a model for the sampled weights, $f_s(w_{ij}|y_{ij}, \boldsymbol{x}_{ij}, v_i)$. A similar expression was derived for nonsampled areas.

The model for the survey weights can be specified conditionally on the response variables to account for the informativeness of the survey design. Possible models for the sample weights considered in the literature include the linear model (Beaumont 2008)

$$w_{ij} = a_0 + a_1 y_{ij} + a_2 y_{ij}^2 + \boldsymbol{x}_{ij}^T\boldsymbol{\alpha} + \epsilon_{ij},$$

and the exponential model for the mean (Pfeffermann et al. 1998; Kim 2002; Eideh and Nathan 2006; Beaumont 2008; Berg and Lee 2019),

$$E_s(w_{ij}|\boldsymbol{x}_{ij}, y_{ij}) = k_i \exp\left(by_{ij} + \boldsymbol{x}_{ij}^T\boldsymbol{\beta}\right). \tag{16}$$

Pfeffermann and Sverchkov (2007) considered the case of continuous response variables, y_{ij}, and modeled the sampled response data using the nested error regression model (2). The exponential model for the survey weights in (16)

was used to model the informative survey design. Under this modeling framework, they showed that the best predictor of $\bar{Y}_i$ is approximately

$$E_p(\bar{Y}_i|D_s) = N_i^{-1}\left[(N_i - n_i)\widehat{\theta}_i + n_i\{\bar{y}_i + (\bar{\boldsymbol{X}}_i - \bar{\boldsymbol{x}}_i)^T\boldsymbol{\beta}\} + (N_i - n_i)b\sigma_e^2\right], \quad (17)$$

where $\widehat{\theta}_i = \widehat{u}_i + \bar{\boldsymbol{X}}_i^T\boldsymbol{\beta}$. The term $(N_i - n_i)b\sigma_e^2$ in (17) is an additional term from the usual best predictor in the nested error regression model (2), which gives a bias correction proportional to the sampling error variance σ_e^2. Finally, there have been examples of flexible nonparametric methods, such as p-splines, used to model the weights (Zheng and Little 2003).

León-Novelo and Savitsky (2019) take a fully Bayesian approach by specifying a population-level model for the response, $f_p(y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{\theta})$, as well as a population-level model for the inclusion probabilities, $f_p(\pi_{ij}|y_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{\gamma})$. Through a Bayes rule argument similar to (14), they show that the implied joint distribution for the sampled units is

$$f_s(y_{ij}, \pi_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{\theta}, \boldsymbol{\gamma}) = \frac{\pi_{ij} f_p(\pi_{ij}|y_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{\gamma})}{E_{y_{ij}|\boldsymbol{x}_{ij},\boldsymbol{\theta}}\{E(\pi_{ij}|y_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{\gamma})\}} \times f_p(y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{\theta}). \quad (18)$$

This joint likelihood for the sample can then be used in a Bayesian model by placing a prior distribution on $(\boldsymbol{\theta}, \boldsymbol{\gamma})$. Note that $\boldsymbol{x}_{ij}$ can be split into two vectors corresponding to $f_p(y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{\theta})$ and $f_p(\pi_{ij}|y_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{\gamma})$ if desired. Consequently, the covariates for the response model and the inclusion probability model need not be the same.

Two computational concerns arise when using the likelihood as in (18). The first issue is that in general, the structure will not lead to conjugate full conditional distributions. To this effect, the authors recommend using the probabilistic programming language Stan (Carpenter et al. 2017), which implements Hamiltonian Monte Carlo (HMC) (Duane et al. 1987) for efficient mixing. The second concern is that the integral involved in the expectation term of (18) needs to be solved for every sampled observation at every iteration of the sampler. If the integral is intractable, it will need to be evaluated numerically, greatly increasing the necessary computation time. They show that if the log-normal distribution is used for the population inclusion probability model, then a closed form can be found for the expectation. Specifically, let $f_p(\pi_{ij}|y_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{\gamma}) = f(\log \pi_{ij}|\mu = g(y_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{\gamma}) + t(\boldsymbol{x}_{ij}, \boldsymbol{\gamma}), \sigma^2 = \sigma_\pi^2)$, where $f(\cdot|\mu, \sigma^2)$ represents a normal distribution with mean μ and variance σ^2, while $g(\cdot)$ and $t(\cdot)$ are some functions. Then

$$f_s(y_{ij}, \pi_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{\theta}, \boldsymbol{\gamma}) = \frac{f(\log \pi_{ij}|\mu = g(y_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{\gamma}) + t(\boldsymbol{x}_{ij}, \boldsymbol{\gamma}), \sigma^2 = \sigma_\pi^2)}{\exp\{t(\boldsymbol{x}_{ij}, \boldsymbol{\gamma}) + \sigma_\pi^2/2\} E_{y_{ij}|\boldsymbol{x}_{ij},\boldsymbol{\theta}}[\exp\{g(y_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{\gamma})\}]} \times f_p(y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{\theta}).$$

In other words, the moment-generating function of the population response model can be used to find the analytical form of the expression, as long as the moment-generating function is defined on the real line. This includes important cases such as the Gaussian, Bernoulli, and Poisson distributions, which are commonly used in the context of survey data.

6. BINOMIAL LIKELIHOOD SPECIAL CASES

The special case of binary responses is of particular interest to survey statisticians, as many surveys focus on the collection of data corresponding to characteristics of sampled individuals, with a goal of estimating the population proportion or count in a small area for a particular characteristic. In this section, some techniques for modifying a working Bernoulli or binomial likelihood using unit-level weights to account for an informative sampling design are discussed. We note that in certain cases, Bernoulli data at the unit level may be collapsed into binomial data at the area level. Thus, some of the methods included below may be viewed either through a unit-level or area-level lens.

Suppose the responses y_{ij} are binary, and the goal is estimation of finite population proportions in each of the small areas $i = 1, \ldots, m$,

$$p_i = \frac{1}{N_i} \sum_{j \in \mathcal{U}_i} y_{ij}.$$

The pseudo-likelihood methods (Binder 1983; Skinner 1989), which were discussed in detail in Section 4 can be directly applied to construct a working likelihood of independent Bernoulli distributions for the sampled survey responses. Zhang et al. (2014) used these ideas to fit a survey-weighted logistic regression model, with random effects included at both the county level and the state level, using the GLIMMIX procedure within SAS, to estimate chronic obstructive pulmonary disease by age race and sex categories within US counties. Another example can be found in Congdon and Lloyd (2010), who used a Bernoulli pseudo-likelihood to estimate diabetes prevalence within US states by demographic groups. Their model formulation was similar to that used by Zhang et al. (2014), but they included an additional random effect to account for spatial correlation.

Malec et al. (1999) proposed a method that is similar in spirit to the pseudo-likelihood method, which uses the survey weights to modify the shape of the binomial likelihood function. Suppose there are D demographic groups of interest and let $\mathcal{S}_d$ be the sampled individuals belonging to the demographic group $d = 1, \ldots, D$. Also, let n_{id} represent the sample size in area i and domain d, while m_{id} represents the corresponding number of positive responses in the sample. Instead of the usual independent binomial likelihood

$\prod_{id} p_{id}^{m_{id}}(1-p_{id})^{n_{id}-m_{id}}$, Malec et al. (1999) proposed a sample-adjusted likelihood

$$\prod_{id} \frac{p_{id}^{m_{id}}(1-p_{id})^{n_{id}-m_{id}}}{(p_{id}/\bar{w}_{1d}+(1-p_{id})/\bar{w}_{0d})^{n_{id}}}, \tag{19}$$

where

$$\bar{w}_{1d} = \sum_{(i,j)\in\mathcal{S}_d} w_{ijd}y_{ijd} \Big/ \sum_{(i,j)\in\mathcal{S}_d} y_{ijd}$$

and

$$\bar{w}_{0d} = \sum_{(i,j)\in\mathcal{S}_d} w_{ijd}(1-y_{ijd}) \Big/ \sum_{(i,j)\in\mathcal{S}_d} (1-y_{ijd}).$$

The quantities $\bar{w}_{1d}$ and $\bar{w}_{0d}$ are used to represent sampling weights for a demographic group d averaged over all individuals with and without a characteristic of interest, respectively. The justification of the denominator of (19) as an adjustment to the likelihood to account for informative sampling is presented in Malec et al. (1999) through use of Bayes' rule and by considering the empirical distribution of the inclusion probabilities.

An alternative approach to the pseudo-likelihood method is to attempt to construct a new, approximate likelihood with independent components, which matches the information contained in the survey sample. Let

$$\widehat{p}_i = \frac{\sum_{j\in\mathcal{S}_i} w_{ij}y_{ij}}{\sum_{j\in\mathcal{S}_i} w_{ij}},$$

be the direct estimate of p_i and let $\widehat{V}_i$ be the estimated variances of $\widehat{p}_i$. Under a simple random sampling design, the variance of the direct estimate $\widehat{p}_i$ is $V_{\text{SRS}}(\widehat{p}_i) = p_i(1-p_i)/n_i$, which can be estimated by $\widehat{V}_{\text{SRS}}(\widehat{p}_i) = \widehat{p}_i(1-\widehat{p}_i)/n_i$. In complex sampling designs with unequal inclusion probabilities or clustering, elements that belong to a common area may be correlated. Because of this, the information in the sample from a complex survey is not equivalent to the information in a simple random sample of the same size. The design effect for $\widehat{p}_i$ is the ratio

$$d_i = d_i(\widehat{p}_i) = \frac{\widehat{V}_D(\widehat{p}_i)}{\widehat{V}_{\text{SRS}}(\widehat{p}_i)} = \frac{n_i\widehat{V}_D(\widehat{p}_i)}{\widehat{p}_i(1-\widehat{p}_i)},$$

and is a measure of the extent to which the variability under the survey design differs from the variability that would be expected under simple random sampling.

The effective sample size, n_i', is defined as the ratio of the sample size to the design effect

$$n_i' = \frac{n_i}{d_i} = \frac{p_i(1-p_i)}{\widehat{V}_D(\widehat{p}_i)}.$$

The effective sample size is an estimate of the sample size required under a noninformative simple random sampling scheme to achieve the same precision to that observed under the complex sampling design. Typically, the effective sample size n_i' will be less than n_i for complex sample designs.

Often the design effect is not available, either due to the lack of available information with which to compute it or due to computational complexity. In such cases, design weights can be used for the estimation of the effective sample size. A simple estimate of the effective sample size, which uses only the design weights was derived by Kish (1965), and is given by

$$n_i' = \frac{\left(\sum_{j\in\mathcal{S}_i} w_{ij}\right)^2}{\sum_{j\in\mathcal{S}_i} w_{ij}^2}.$$

Other estimates of the design effect which use the survey weights, sample sizes, and population totals and are appropriate for stratified sampling designs, can be found in Kish (1992).

Chen et al. (2014) and Franco and Bell (2015) used the design effect and effective sample size to define the "effective number of cases," $y_i^* = n_i'\widehat{p}_i$. The effective number of cases, y_i^*, was then modeled using a binomial, logit-normal hierarchical structure. The sample model for the effective number of cases is then

$$y_i^*|p_i \sim \text{Binomial}(n_i', p_i), i = 1, \ldots, m,$$

with a linking model of

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \boldsymbol{x}_i^T\boldsymbol{\beta} + v_i,$$

where the v_i are area-specific random effects. Using the effective number of cases and the effective sample size in a binomial model is an attempt to construct a likelihood, which is valid under a simple random sampling design, and will produce approximately equivalent inferences as when using the exact, but possibly unknown or computationally intractable likelihood.

Different distributional assumptions on the random effects can be made to accommodate aspects of the data or different correlation structures particular to sampled geographies. Noting that it might be expected that areas which are close to each other might share similarities, Chen et al. (2014) decomposed the

random effects v_i into spatial and a non-spatial components, so that $v_i = u_i + \varepsilon_i$, where $\varepsilon_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma_\varepsilon^2)$, and an ICAR prior is placed over u_i.

$$u_i | u_j, j \in ne(i) \sim N(\bar{u}_i, \sigma_u^2 / m_i),$$

where $ne(i)$ is the set of geographies which are neighbors of area i, m_i is the size of $ne(i)$, and $\bar{u}_i = \sum_{j \in ne(i)} u_j / m_i$.

Franco and Bell (2015) introduced a time dependence structure into the random effect v_i for situations in which there are data from multiple time periods available and applied their model to the estimation of poverty rates using multiple years of American Community Survey data. In their formulation, the random effects have an AR(1) correlation structure, so that the model becomes

$$\begin{aligned} y_{i,t}^* | p_{i,t} &\sim \text{Binomial}(n'_{i,t}, p_{i,t}),\ i = 1, \ldots, m,\ t = 1, \ldots, T \\ \text{logit}(p_{i,t}) &= \boldsymbol{x}_{i,t}^T \boldsymbol{\beta}_t + \sigma_t^2 v_{i,t} \\ v_{i,t} &= \phi v_{i,t-1} + \varepsilon_{i,t} \end{aligned}$$

where $|\phi| < 1$, and the $\varepsilon_{i,t}$ are assumed to be i.i.d. $N(0, 1 - \phi^2)$ random variables. The unknown parameters $\boldsymbol{\beta}_t$ and σ_t^2 are allowed to vary over time. Franco and Bell (2015) showed that the reductions in prediction uncertainty can be meaningful when the autoregressive parameter ϕ is large, but that the reduction in prediction uncertainty is more modest when $|\phi| < 0.4$. As noted by Chen et al. (2014), the inclusion of spatial or spatiotemporal random effects has the added benefit that the dependent random effects can serve as a surrogate for the variables responsible for dependency in the data.

The above methods use the survey weights either to modify the shape of an independent likelihood (Malec et al. 1997; Zheng and Little 2003) to account for the informative design, or to estimate a design effect in an attempt to match the information contained in the survey sample to the information implied by an independent likelihood by adjusting the sample size (Chen et al. 2014; Franco and Bell 2015). Alternatively, one could specify a working independence model for the sampled units and incorporate the survey design by using the survey weights as predictors (Zheng and Little 2003), and to induce dependence through a latent process model.

7. CONCLUSION

Unit-level models pose many advantages relative to area-level models. These advantages include increased precision and straightforward spatial aggregation (the so-called benchmarking problem), among others. Estimation of unit-level models requires attention to the specific sampling design. That is, the unit response may be dependent on the probability of selection, even after

conditioning on the design variables. In this sense, the sampling design is said to be *informative* and care must be taken to avoid bias. Applications of these methods are of particular interest (e.g., see Parker et al. (2023) for a comprehensive illustration).

With these tools at hand, there are many opportunities for future research. For example, including administrative records into the previous model formulations constitutes one area of active research as care needs to be taken to probabilistically account for the record linkage. Methods for disclosure avoidance in unit-level models also provide another avenue for future research. Finally, in many cases, the level of informativeness is unknown. Methods to assess the degree of informativeness, especially after conditioning on a set of covariates, are an important avenue of future research. In short, there are substantial opportunities for improving the models presented herein. In doing so, the aim is to provide computationally efficient estimates with improved precision. In particular, as data size and the number of domains become large, the computation can become a burden with unit-level data, especially when considering various dependence structures. More research into scalable approaches such as the use of INLA (e.g., Orozco-Acosta et al. 2021) or variational Bayes approximations (e.g., Parker et al. 2022) will be of substantial value. Ultimately, this will provide additional tools for official statistical agencies, survey methodologists, and subject-matter scientists.

REFERENCES

Asparouhov, T. (2006), "General Multi-Level Modeling with Sampling Weights," *Communications in Statistics—Theory and Methods*, 35, 439–460. 3

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," *Journal of the American Statistical Association*, 83, 28–36.

Bauder, M., Luery, D., and Szelepka, S. (2018), "Small Area Estimation of Health Insurance Coverage in 2010–2016," Technical Reports, Small Area Methods Branch, Social, Economic, and Housing Statistics Division, U.S. Census Bureau. Available at https://www2.census.gov/programs-surveys/sahie/technical-documentation/methodology/2008-2016-methods/sahie-tech-2010-to-2016.pdf.

Beaumont, J.-F. (2008), "A New Approach to Weighting and Inference in Sample Surveys," *Biometrika*, 95, 539–553.

Bell, W. R., Basel, W. W., and Maples, J. J. (2016), "An Overview of the U. S. Census Bureau's Small Area Income and Poverty Estimates Program," in *Analysis of Poverty Data by Small Area Estimation*, ed. M. Pratesi, New York: Wiley, pp. 349–378.

Berg, E., and Lee, D. (2019), "Small Area Prediction of Quantiles for Zero-Inflated Data and an Informative Sample Design," *Statistical Theory and Related Fields*, 3, 114–128.

Besag, J. (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems (with Discussion)," *Journal of the Royal Statistical Society. Series B*, 36, 192–225.

Binder, D. A. (1983), "On the Variances of Asymptotically Normal Estimators from Complex Surveys," *International Statistical Review*, 51, 279–292.

Bonnéry, D., Breidt, F. J., and Coquet, F. (2018), "Asymptotics for the Maximum Sample Likelihood Estimator under Informative Selection from a Finite Population," *Bernoulli*, 24, 929–955.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017), "Stan: A Probabilistic Programming Language," *Journal of Statistical Software*, 76, 1–32.

Chen, C., Wakefield, J., and Lumely, T. (2014), "The Use of Sampling Weights in Bayesian Hierarchical Models for Small Area Estimation," *Spatial and Spatio-Temporal Epidemiology*, 11, 33–43.

Congdon, P., and Lloyd, P. (2010), "Estimating Small Area Diabetes Prevalence in the US Using the Behavioral Risk Factor Surveillance System," *Journal of Data Science*, 8, 235–252.

Datta, G. S., and Ghosh, M. (1991), "Bayesian Prediction in Linear Models: Applications to Small Area Estimation," *The Annals of Statistics*, 19, 1748–1770.

Dong, Q., Elliott, M. R., and Raghunathan, T. E. (2014), "A Nonparametric Method to Generate Synthetic Populations to Adjust for Complex Sampling Design Features," *Survey Methodology*, 40, 29–46.

Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987), "Hybrid Monte Carlo," *Physics Letters B*, 195, 216–222.

Eideh, A. (2012), "Fitting Variance Component Model and Fixed Effects Model for One-Way Analysis of Variance to Complex Survey Data," *Communications in Statistics – Theory and Methods*, 41, 3278–3300.

Eideh, A., and Nathan, G. (2006), "Fitting Time Series Models for Longitudinal Survey Data under Informative Sampling," *Journal of Statistical Planning and Inference*, 136, 3052–3069.

———. (2009), "Two-Stage Informative Cluster Sampling–Estimation and Prediction with Application for Small-Area Models," *Journal of Statistical Planning and Inference*, 139, 3088–3101.

Franco, C., and Bell, W. R. (2015), "Borrowing Information Overtime in Binomial/Logit Normal Models for Small Area Estimation," *Statistics in Transition*, 16, 563–584.

Gao, Y., Kennedy, L., Simpson, D., and Gelman, A. (2021), "Improving Multilevel Regression and Poststratification with Structured Priors," *Bayesian Analysis*, 16, 719–744.

Gelman, A. (2007), "Struggles with Survey Weighting and Regression Modeling," *Statistical Science*, 22, 153–164.

Gelman, A., and Little, T. C. (1997), "Poststratification into Many Categories Using Hierarchical Logistic Regression," *Survey Methodology*, 23, 2, 127–135.

Gelman, A., Vehtari, A., Carlin, J., Stern, H. S., Dunson, D., and Rubin, D. (1995), *Bayesian Data Analysis*. London: Chapman and Hall.

Godambe, V., and Thompson, M. E. (1986), "Parameters of Superpopulation and Survey Population: Their Relationships and Estimation," *International Statistical Review*, 54, 127–138.

Grilli, L., and Pratesi, M. (2004), "Weighted Estimation in Multilevel Ordinal and Binary Models in the Presence of Informative Sampling Designs," *Survey Methodology*, 30, 93–103.

Guadarrama, M., Molina, I., and Rao, J. N. K. (2018), "Small Area Estimation of General Parameters under Complex Sampling Designs," *Computational Statistics and Data Analysis*, 121, 20–40.

Hall, P., and Maiti, T. (2006), "Nonparametric Estimation of Mean-Squared Prediction Error in Nested-Error Regression Models," *The Annals of Statistics*, 34, 1733–1750.

Hidiroglou, M. A., and You, Y. (2016), "Comparison of Unit Level and Area Level Small Area Estimators," *Survey Methodology*, 42, 41–61.

Jiang, J., and Lahiri, P. (2006), "Estimation of Finite Population Domain Means: A Model-Assisted Empirical Best Prediction Approach," *Journal of the American Statistical Association*, 101, 301–311.

Kim, D. H. (2002), "Bayesian and Empirical Bayesian Analysis under Informative Sampling." *Sankhyā: The Indian Journal of Statistics, Series B*, 64, 267–288.

Kim, J. K., Park, S., and Lee, Y. (2017), "Statistical Inference Using Generalized Linear Mixed Models under Informative Cluster Sampling," *Canadian Journal of Statistics*, 45, 479–497.

Kish, L. (1965), *Survey Sampling*, New York: Wiley.

———. (1992), "Weighting for Unequal P*i*," *Journal of Official Statistics*, 8, 183.

León-Novelo, L. G., and Savitsky, T. D. (2019), "Fully Bayesian Estimation under Informative Sampling," *Electronic Journal of Statistics*, 13, 1608–1645.

Little, R. J. (1993), "Post-Stratification: A Modeler's Perspective," *Journal of the American Statistical Association*, 88, 1001–1012.
———. (2012), "Calibrated Bayes, an Alternative Inferential Paradigm for Official Statistics," *Journal of Official Statistics*, 28, 3, 309.
Luery, D. M. (2011), "Small Area Income and Poverty Estimates Program," in Proceedings of 27th SCORUS Conference, Jurmala, Latvia, pp. 93–107.
Lumley, T., and Scott, A. (2017), "Fitting Regression Models to Survey Data," *Statistical Science*, 32, 265–278.
Malec, D., Davis, W. W., and Cao, X. (1999), "Model-Based Small Area Estimates of Overweight Prevalence Using Sample Selection Adjustment," *Statistics in Medicine*, 18, 3189–3200.
Malec, D., Sedransk, J., Moriarity, C. L., and LeClere, F. B. (1997), "Small Area Inference for Binary Variables in the National Health Interview Survey," *Journal of the American Statistical Association*, 92, 815–826.
Marhuenda, Y., Molina, I., Morales, D., and Rao, J. N. K. (2017), "Poverty Mapping in Small Areas under a Twofold Nested Error Regression Model," *Journal of the Royal Statistical Society, Series A*, 180, 1111–1136.
Molina, I., and Rao, J. N. K. (2010), "Small Area Estimation of Poverty Indicators," *The Canadian Journal of Statistics*, 38, 369–385.
Nathan, G., and Holt, D. (1980), "The Effect of Survey Design on Regression Analysis," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 42, 377–386.
Orozco-Acosta, E., Adin, A., and Ugarte, M. D. (2021), "Scalable Bayesian Modelling for Smoothing Disease Risks in Large Spatial Data Sets Using INLA," *Spatial Statistics*, 41, 100496.
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*, ed. E. Cohen, Stanford, CA: Stanford University Press, pp. 209–228.
Parker, P. A., Holan, S. H., and Janicki, R. (2022), "Computationally Efficient Bayesian Unit-Level Models for non-Gaussian Data under Informative Sampling with Application to Estimation of Health Insurance Coverage," *The Annals of Applied Statistics*, 16, 887–904.
Parker, P. A., Janicki, R., and Holan, S. H. (2023), "Comparison of Unit Level Small Area Estimation Modeling Approaches for Survey Data Under Informative Sampling," *Journal of Survey Statistics and Methodology*, DOI: 10.1093/jssam/smad022.
Pfeffermann, D. (1993), "The Role of Sampling Weights When Modeling Survey Data," *International Statistical Review*, 61, 317–337.
———. (2002), "Small Area Estimation – New Developments and Directions," *International Statistical Review*, 70, 125–143.
———. (2003), "New Important Developments in Small Area Estimation," *Statistical Science*, 28, 40–68.
Pfeffermann, D., Krieger, A. M., and Rinott, Y. (1998), "Parametric Distributions of Complex Survey Data under Informative Probability Sampling," *Statistica Sinica*, 8, 1087–1114.
Pfeffermann, D., and Sverchkov, M. (1999), "Parametric and Semi-Parametric Estimation of Regression Models Fitted to Survey Data." *Sankhyā*," *The Indian Journal of Statistics, Series B*, 61, 166–186.
———. (2007), "Small-Area Estimation under Informative Probability Sampling of Areas and within the Selected Areas," *Journal of the American Statistical Association*, 102, 1427–1439.
Prasad, N., and Rao, J. (1990), "The Estimation of Mean Squared Error of Small-Area Estimators," *Journal of the American Statistical Association*, 85, 163–171.
Rabe-Hesketh, S., and Skrondal, A. (2006), "Multilevel Modelling of Complex Survey Data," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169, 805–827.
Rao, J. N. K., and Molina, I. (2015), *Small Area Estimation*, Hoboken, New Jersey: Wiley.
Rao, J., Verret, F., and Hidiroglou, M. A. (2013), "A Weighted Composite Likelihood Approach to Inference for Two-Level Models from Survey Data," *Survey Methodology*, 39, 263–282.
Ribatet, M., Cooley, D., and Davison, A. C. (2012), "Bayesian Inference from Composite Likelihoods, with an Application to Spatial Extremes," *Statistica Sinica*, 22, 813–845.
Rubin, D. B. (1976), "Inference and Missing Data," *Biometrika*, 63, 581–592. 3

Savitsky, T. D., and Toth, D. (2016), "Bayesian Estimation under Informative Sampling," *Electronic Journal of Statistics*, 10, 1677–1708.

Scott, A., and Smith, T. M. F. (1969), "Estimation in Multi-Stage Surveys," *Journal of the American Statistical Association*, 64, 830–840.

Si, Y., Pillai, N. S., and Gelman, A. (2015), "Bayesian Nonparametric Weighted Sampling Inference," *Bayesian Analysis*, 10, 605–625.

Skinner, C. J. (1989), "Domain Means, Regression and Multivariate Analysis," in *Analysis of Complex Surveys*, eds. C. J. Skinner, D. Holt, and T. M. F. Smith, Chichester: Wiley, pp. 80–84 (chap. 2).

Slud, E. V. (2020), "Model-Assisted Estimation of Mixed-Effect Model Parameters in Complex Surveys," Technical Reports, Center for Statistical Research and Methodology, Research and Methodology Directorate, U.S. Census Bureau. Available at https://www.census.gov/content/dam/Census/library/working-papers/2020/adrm/RRS2020-05.pdf.

Stukel, D. M., and Rao, J. (1999), "On Small-Area Estimation under Two-Fold Nested Error Regression Models," *Journal of Statistical Planning and Inference*, 78, 131–147.

Sugden, R. A., and Smith, T. M. F. (1984), "Ignorable and Informative Designs in Survey Sampling Inference," *Biometrika*, 71, 495–506.

Vandendijck, Y., Faes, C., Kirby, R. S., Lawson, A., and Hens, N. (2016), "Model-Based Inference for Small Area Estimation with Sampling Weights," *Spatial Statistics*, 18, 455–473.

Varin, C., Reid, N., and Firth, D. (2011), "An Overview of Composite Likelihood Methods," *Statistica Sinica*, 21, 5–42.

Verret, F., Rao, J. N. K., and Hidiroglou, M. A. (2015), "Model-Based Small Area Estimation under Informative Sampling," *Survey Methodology*, 41, 333–347.

Wang, Z., Kim, J. K., and Yang, S. (2018), "Approximate Bayesian Inference under Informative Sampling," *Biometrika*, 105, 91–102.

Williams, M. R., and Savitsky, T. D. (2020), "Bayesian Estimation under Informative Sampling with Unattenuated Dependence," *Bayesian Analysis*, 15, 57–77.

Yi, G., Rao, J. N. K., and Li, H. (2016), "A Weighted Composite Likelihood Approach for Analysis of Survey Data under Two-Level Models," *Statistica Sinica*, 26, 569–587.

You, Y., and Rao, J. (2002), "A Pseudo-Empirical Best Linear Unbiased Prediction Approach to Small Area Estimation Using Survey Weights," *Canadian Journal of Statistics*, 30, 431–439.

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," *American Journal of Epidemiology*, 179, 1025–1033.

Zheng, H., and Little, R. J. (2003), "Penalized Spline Model-Based Estimation of the Finite Populations Total from Probability-Proportional-to-Size Samples," *Journal of Official Statistics*, 19, 99.

Zimmerman, T., and Münnich, R. T. (2018), "Small Area Estimation with a Lognormal Mixed Model under Informative Sampling," *Journal of Official Statistics*, 34, 523–542.