

Weight Smoothing for Generalized Linear Models Using a Laplace Prior

Xi Xia¹ and Michael R. Elliott^{1,2}

When analyzing data sampled with unequal inclusion probabilities, correlations between the probability of selection and the sampled data can induce bias if the inclusion probabilities are ignored in the analysis. Weights equal to the inverse of the probability of inclusion are commonly used to correct possible bias. When weights are uncorrelated with the descriptive or model estimators of interest, highly disproportional sample designs resulting in large weights can introduce unnecessary variability, leading to an overall larger mean square error compared to unweighted methods.

We describe an approach we term ‘weight smoothing’ that models the interactions between the weights and the estimators as random effects, reducing the root mean square error (RMSE) by shrinking interactions toward zero when such shrinkage is allowed by the data. This article adapts a flexible Laplace prior distribution for the hierarchical Bayesian model to gain a more robust bias-variance tradeoff than previous approaches using normal priors. Simulation and application suggest that under a linear model setting, weight-smoothing models with Laplace priors yield robust results when weighting is necessary, and provide considerable reduction in RMSE otherwise. In logistic regression models, estimates using weight-smoothing models with Laplace priors are robust, but with less gain in efficiency than in linear regression settings.

Key words: Weight trimming; winsorization; Bayesian finite population inference; Hierarchical models.

1. Introduction

Studies based on data sampled with unequal inclusion probabilities typically apply case weights equal to the inverse of the probability of inclusion to reduce or remove bias in estimators of descriptive population quantities, such as means or totals (Horvitz and Thompson 1952). This “fully weighted” approach can be extended to estimate analytical quantities that focus on association between risk factors and outcomes, such as population slopes in linear and generalized linear models, by applying sampling weights to score equations, and solving for the resulting “pseudo-maximum likelihood” estimators (PMLEs) (Binder 1983; Pfeffermann 1993). Unweighted and weighted estimators generally correspond when the underlying model (either implicit or explicit) is correctly specified and the sampling scheme is noninformative. When the model is misspecified or the sampling

¹ Dept. of Biostatistics, University of Michigan School of Public Health, 1415 Washington Heights, Ann Arbor, MI 48109, USA. Email: xiaxi@umich.edu.

² Survey Methodology Program, Institute for Social Research, University of Michigan, 426 Thompson St., Ann Arbor, MI 48109, USA. Email: mreliott@umich.edu.

Acknowledgments: We thank the Editor, the Associate Editor, and three reviewers for their careful review that helped to improve the quality of the article. This work was supported in part by Grant Number R01CA129101 from the National Cancer Institute.

scheme is informative, weighted estimators typically reduce bias, often (although not always) at the cost of increased variance. As model assumptions improve and/or sampling better approximates noninformativeness, the increase in variance from weighted analysis could overwhelm the reduction in bias, leading to an overall larger mean square error (MSE) than would be the case if the weights were ignored or at least controlled in some fashion.

In many, if not most cases, fully weighted estimators are used without concerns about such tradeoffs. When variability in weights is of concern, weight trimming, or “winsorization,” is used to control the variation in weights by capping the weights at some value w_0 , and redistributing the values above w_0 among the rest (Alexander et al. 1997; Kish 1992; Potter 1990). Various criteria have been used to determine the cap value based on data. Some examples include the National Assessment of Education Progress (NAEP) method by Potter (1988), which set the cutoff point equal to $\sqrt{c \sum_{i \in S} w_i^2 / n}$, where c was chosen in an ad-hoc manner. Cox and McGrath (1981) approached it by estimating the cutoff point value which optimizes the empirical MSE estimated by $\widehat{MSE}(\hat{\theta}_t) = (\hat{\theta}_t - \hat{\theta}_w)^2 - \widehat{Var}(\hat{\theta}_t) + 2\sqrt{\widehat{Var}(\hat{\theta}_t)\widehat{Var}(\hat{\theta}_w)}$, where $\hat{\theta}_w$ is the fully weighted estimator, and $\hat{\theta}_t$, $t = 1, \dots, T$, is the weight-trimmed estimator, with t denoting various trimming levels ordered from lowest to highest. These levels are ad hoc, except for $w_0 = \bar{w}$ for $t = 1$, which fixes all weights to their mean value and yields the unweighted estimator, and $w_0 = \max w_i$ for $t = T$, which yields the fully weighted (untrimmed) estimator. Chowdhury et al. (2007) suggested treating the weights as coming from a skewed cumulative distribution (e.g., an exponential distribution), and using the upper one percent of the fitted distribution as a cut point for weight trimming. Beaumont (2008) proposed a generalized design-based method, replacing the actual weights with weights predicted using a function of response and design variables. Details of these design-based approaches are summarized in Henry and Valliant’s (2012) review.

An alternative to standard design-based weighted estimation is a model-based approach that accommodates disproportional probability-of-selection design in a finite population Bayesian inference setting. By creating dummy variables stratified by equal or approximately equal case weights, a fully weighted data analysis is obtained by building a model with indicators for the weight strata together with interaction terms between the weight stratum indicators and model parameters of interest, then obtaining inference about the population quantity of interest from its posterior predictive distribution. Elliott and Little (2000) established two model-based approaches for weight-trimming: model averaging, or “weight pooling”, and hierarchical modeling, or “weight smoothing”. A weight pooling model collapses strata with similar weights together with their associated interaction terms, mimicking a data-driven weight-trimming process. Weight smoothing treats the underlying weight strata as random effects, and achieves a balance between fully weighted and unweighted estimates using a shrinkage estimator: thus the weight strata are smoothed if data provide little evidence of difference between strata, and are separated if data suggest that interactions with strata are present. Under a Bayesian framework, a two-level model is implemented, assigning a multivariate normal prior for the random effects, with inference obtained from the posterior predictive distribution of the population parameter of interest. Elliott (2007) extended the application of weight-smoothing models to linear and generalized linear models, and discussed different structures for the random-effect priors, namely exchangeable, autoregressive, linear and nonparametric random

slopes. Both of these papers found that there was an efficient/robustness tradeoff with respect to structure in the prior mean and variance, with simple exchangeable models providing highly efficient estimates when the weights provided little bias correction, but being susceptible to “oversmoothing” and yielding biased estimators when weights really were necessary to provide substantial bias correction. Highly structured models such as smoothing splines provided robust bias correction, but had less dramatic gains in efficiency and were more complex to implement. Hence we are motivated to find alternative models that will induce weight smoothing under simple mean and covariance matrix settings, while improving the bias-variance tradeoff. A logical choice would be a Laplace prior, which can be viewed as providing a Bayesian version of a LASSO regression model (Park and Casella 2008). This heavier-tailed prior might be expected to provide little or no shrinkage when bias correction is required, but still allow approximation to an unweighted estimator when the data suggest weak relationships with probability of selection.

In this article we extend the weight-smoothing approach by use of Laplace priors for the random-effect weight strata and interaction terms instead of multivariate normal priors, in order to achieve more robustness against “oversmoothing” in settings where weights are required to accommodate model misspecification or nonignorable sampling. We evaluate the performance of our proposed model in a simulation study, under both model misspecification and informative sampling, for both continuous and dichotomous outcomes, and compare it with competing methods. The article is organized as follows. In Section 2 we review the theory of model smoothing together with recently proposed model-assisted methods, and develop our model with Laplace priors. Section 3 provides simulation studies, and compares bias, coverage and MSE of the proposed method with competing methods. Section 4 demonstrates the method’s performance for both linear and logistic scenarios by applications to dioxin data from the National Health and Nutrition Examination Survey (NHANES) and Partners for Child Passenger Safety dataset. Section 5 provides a summary discussion.

2. Weight-Smoothing Methodology

2.1. Finite Bayesian Population Inference

For finite Bayesian population inference, we model the population data $Y: Y \sim f(Y| \theta, Z)$, where Z are the variables associated with the sample design (probabilities of selection, cluster indicators, stratum variables). Note that the parametric model f can either be highly parametric with a low dimension θ (e.g., a normal model with common mean and variance), or have a more semiparametric or nonparametric flavor with a high-dimension θ (such as a spline or Dirichlet process model). Inference about some population quantity of interest $Q(Y)$ is based on the posterior predictive distribution of

$$p(Y_{nob}|Y_{obs}, I, Z) = \frac{\int \int p(I|Y, Z, \theta, \phi)p(Y_{nob}|Y_{obs}, Z, \theta, \phi)p(Y_{obs}|Z, \theta)p(\theta, \phi)d\theta d\phi}{\int \int \int p(I|Y, Z, \theta, \phi)p(Y_{nob}|Y_{obs}, Z, \theta, \phi)p(Y_{obs}|Z, \theta)p(\theta, \phi)d\theta d\phi dY_{nobs}} \quad (1)$$

where Y_{nob} consists of the $N - n$ unobserved cases in the population, θ models Y (possibly conditional on Z) and ϕ models the inclusion indicator I (equal to 1 if the unit is sampled

and observed and 0 otherwise). Thus $p(I|Y, Z, \theta, \phi)$ refers to the distribution of the sample mechanism given the design variables and the data of interest, $p(Y_{nob}|Y_{obs}, Z, \theta, \phi)$ gives the distribution of the unobserved elements in the sample given the observed elements in the sample and the (fully observed) design variables, and $p(Y_{obs}|Z, \theta)$ models the observed data given any design variables. Assuming that ϕ and θ have independent priors, the sampling mechanism is said to be “noninformative” if the distribution of I is independent of $Y|Z$, or “ignorable” if the distribution of I only depends on $Y_{obs}|Z$. When the sampling design is ignorable, $p(I|Y, Z, \theta, \phi) = p(I|Y_{obs}, Z, \phi)$, and thus (1) reduces to

$$p(Y_{nob}|Y_{obs}, Z) = \frac{\int p(Y_{nob}|Y_{obs}, Z, \theta)p(Y_{obs}|Z, \theta)p(\theta)d\theta}{\int \int p(Y_{nob}|Y_{obs}, Z, \theta)p(Y_{obs}|Z, \theta)p(\theta)d\theta dY_{obs}},$$

allowing inference about $Q(Y)$ to be made without explicitly modeling the sampling inclusion parameter I (Ericson 1969; Holt and Smith 1979; Little 1993; Rubin 1987; Skinner et al. 1989). Note that if inference about quantities $Q(Y|X)$ involving covariates X is desired (e.g., regression slopes), noninformative or ignorable sample designs can be relaxed to have the distribution of I depend on X .

2.2. Weight Modeling

Beaumont (2008) proposed an alternative model-assisted method, tamping down the extreme values in weights by replacing weights with values from a prediction model of weights regressed on response and design variables. Denote $I = (I_1, \dots, I_N)^T$ as the vector of sample inclusion indicators, that is, $I_i = 1$ as i th unit sampled and $I_i = 0$ otherwise, $Y = (Y_1, \dots, Y_N)^T$ the vector of survey-response variables, and $Z = (Z_1, \dots, Z_N)^T$ the vector of design variables. Assuming a noninformative sampling design, thus $P(I|Z, Y) = P(I|Z)$, the predicted weights are obtained by $\tilde{w}_i = E_M(w_i|I_i = 1, y_i)$, where M refers to the expectation of w_i under a given model. Beaumont discussed two models, the linear form $E_M(w_i|I, Y) = H_i^T \beta + v_i^{1/2} \epsilon_i$, and the exponential form, $E_M(w_i|I, Y) = 1 + \exp(H_i^T \beta + v_i^{1/2} \epsilon_i)$, where H_i and $v_i > 0$ are known functions of y_i . (The exponential form prevents the predicted weights from being negative.) He presented two examples of $H_i^T \beta$, one-degree polynomial and five-degree polynomial of y_i . The predicted weights are obtained by fitting the (unweighted) model on the sampled data, then the reweighted estimator of the survey-response variable of interest is obtained using the predicted weights. Extensions to regression settings can consider models of the form $E_M(w_i|I, Y, X) = J_i^T \beta + u_i^{1/2} \epsilon_i$, where J_i and u_i are functions of Y_i and X_i (possibly including interactions).

2.3. Weight Smoothing

In general, weight smoothing stratifies the data by inclusion probability, and applies a hierarchical model treating strata means as random effects, thus achieving trimming via shrinkage. Considering the population mean as the quantity of interest, a general weight-smoothing model is as follows:

$$Y_{hi} \stackrel{iid}{\sim} N(\mu_h, \sigma^2)$$

$$\mu \sim N_H(\phi, G)$$

where $\mu = (\mu_1, \dots, \mu_H)$, $\phi = (\phi_1, \dots, \phi_H)$, and $h = 1, \dots, H$ indexes different “weight strata” defined, for example, by same or similar inclusion probabilities. In the case of stratified or poststratified sample designs, h indexes the actual strata. In more general designs, subjects can be formed into strata with equal or similar weights. We assume ϕ , G , and σ^2 all have weak or noninformative priors. Typically these strata are ordered from the smallest weight (highest probability of selection) to the largest weight (lowest probability of selection), but this is not required if a more natural ordering is available, for example, if the weight strata represent a disproportionately stratified sample by age. Based on this model, the posterior mean of the population mean is derived as:

$$E(\bar{Y}|y) = \sum_{h=1}^H [n_h \bar{y}_h + (N_h - n_h) \hat{\mu}_h] / N$$

where N_h and n_h are the population and sample sizes in stratum h , respectively, and $\hat{\mu}_h = E(\mu_h|y)$. Various assumptions can be made for the prior distribution of μ , such as

Exchangeable random effect (XRE): $\phi_h = \phi_0$ for all h , $G = \tau^2 I_H$

Autoregressive (AR1): $\phi_h = \phi_0$ for all h , $G = \tau^2 A$, $A_{jk} = \rho^{|j-k|}$, $j, k = 1, \dots, H$

Linear (LIN): $\phi_h = \phi_0 + \phi' * h$, $G = \tau^2 I_H$

Nonparametric (NPAR): $\phi_h = g(h)$, $G = 0$ where g is an unspecified, twice-differentiable function.

See Elliott and Little (2000) for a detailed review.

The weight-smoothing mechanism can be easily intuited in the simplest case of the exchangeable random-effect (XRE) model (Holt and Smith 1979; Ghosh and Meeden 1986; Little 1991; Lazzaroni and Little 1998), where $\phi_h = \mu$ for all h , and $G = \tau^2 I_H$. The estimation of $\hat{\mu}_h$ is now a shrinkage estimator as $\hat{\mu}_h = w_h \bar{y}_h + (1 - w_h) \bar{y}$, for $w_h = \tau^2 n_h / (\tau^2 n_h + \sigma^2)$ and $\bar{y} = (\sum_h n_h / (n_h \tau^2 + \sigma^2))^{-1} \sum_h n_h / (n_h \tau^2 + \sigma^2) \bar{y}_h$. As $\tau^2 \rightarrow \infty$, $w_h \rightarrow 1$, and $E(\bar{Y}|y) = \sum_{h=1}^H [n_h \bar{y}_h + (N_h - n_h) \bar{y}] / N = \sum_{h=1}^H (N_h / N) \bar{y}_h$, the fully weighted estimator. On the other hand, as $\tau^2 \rightarrow 0$, $w_h \rightarrow 0$, and the estimation shrinks toward the unweighted mean: since $\bar{y} = \sum_h \frac{n_h \bar{y}_h / \sigma^2}{\sum_h n_h / \sigma^2} = \bar{y}$ if $\tau^2 = 0$, $E(\bar{Y}|y) = \sum_{h=1}^H [n_h \bar{y}_h + (N_h - n_h) \bar{y}] / N = (n/N) \bar{y} + \bar{y} (1 - n/N) = \bar{y}$ if $\tau^2 = 0$. Since τ^2 is itself estimated from the data, and is a measure of the information available to distinguish how the population means within a weight strata differ, the weight-smoothing model achieves a “data-driven” compromise between the weighted estimator, which is design consistent but may be highly inefficient, and the unweighted estimator, which is fully efficient when the assumption of independence between inclusion probability and mean of Y holds, but is likely biased otherwise.

2.4. Weight Smoothing for Linear and Generalized Linear Regression Models

Generalized linear regression models (McCullagh and Nelder 1989) postulate a likelihood for y_i of the form

$$f(y_i | \theta_i, \sigma) = \exp \left[\frac{y_i \theta_i - b(\theta_i)}{a_i(\sigma)} + c(y_i, \sigma) \right]$$

where $a_i(\sigma)$ is a known function of (nuisance) scale parameter σ , and the mean of y_i given by $\mu_i = b'(\theta_i)$ is based on a linear combination of fixed covariates x_i through some link function g such that $E(y_i | \theta_i) = \mu_i$, and $g(\mu_i) = g(b'(\theta_i)) = \eta_i = x_i^T \beta$. In the meantime,

$\text{Var}(y_i|\theta_i) = a_i(\sigma)V(\mu_i)$, where $V(\mu_i) = b''(\theta_i)$; thus the variance is usually a function of the mean, with the exception of normal distribution, for which $b''(\theta_i) = 1$. The link is considered canonical if $\theta_i = \eta_i$, with the simplifying result that $V(\mu_i) = 1/g'(\mu_i)$. Some examples include Gaussian (linear) regression, where $a_i(\sigma) = \sigma^2$ and the canonical link is $g(\mu_i) = \mu_i$; logistic regression, where $a_i(\sigma) = n_i^{-1}$ and the canonical link is $g(\mu_i) = \log(\mu_i/(1 - \mu_i))$, and Poisson regression, where $a_i(\sigma) = 1$ and the canonical link is $g(\mu_i) = \log(\mu_i)$.

When considering weighted estimators, we index by the inclusion stratum h , thus $g(E[y_{hi}|\beta_h]) = x_{hi}^T \beta_h$. For weight-smoothing models, the hierarchical structure is considered as

$$(\beta_1^T, \dots, \beta_H^T)^T | \beta^*, G \sim N_{HP}(\beta^*, G)$$

where β^* is an unknown vector of mean values for the regression coefficients and G is an unknown covariance matrix. Our interest is to estimate the target-population quantity $B = (B_1, \dots, B_p)^T$, which is the slope that solves the population score equation $U_N(B) = 0$ where

$$U_N(\beta) = \sum_{i=1}^N \frac{\partial}{\partial \beta} \log f(y_i; \beta) = \sum_{h=1}^H \sum_{i=1}^{N_h} \frac{y_{hi} - g^{-1}(\mu_i(\beta))x_{hi}}{V(\mu_{hi}(\beta))g'(\mu_{hi}(\beta))}$$

Notice that the quantity B that satisfies $U(B) = 0$ is always a meaningful population quantity even if the model is misspecified, since it is a linear approximation of x_i to η_i . A first-order approximation of $E(B|y, X)$ is given based on \hat{B} where

$$\sum_{h=1}^H W_h \sum_{i=1}^{n_h} \frac{(\hat{y}_{hi} - g^{-1}(\mu_i(\hat{B})))x_{hi}}{V(\mu_{hi}(\hat{B}))g'(\mu_{hi}(\hat{B}))} = 0$$

where $W_h = N_h/n_h$, $\hat{y}_{hi} = g^{-1}(x_{hi}^T \hat{\beta}_h)$, and $\hat{\beta}_h = E(\beta_h|y, X)$. For linear regression, where $V(\mu_i) = \sigma^2$ and $g'(\mu_i) = 1$,

$$\hat{B} = E(B|y, X) = \left[\sum_h W_h \sum_{i=1}^{n_h} x_{hi} x_{hi}' \right]^{-1} \left[\sum_h W_h \left(\sum_{i=1}^{n_h} x_{hi} x_{hi}' \right) \hat{\beta}_h \right].$$

In case of logistic regression, $V(\mu_i) = \mu_i(1 - \mu_i)$ and $g'(\mu_i) = \mu_i^{-1}(1 - \mu_i)^{-1}$, and $E(B|y, X)$ is obtained by solving the weighted score equation for population regression parameter B

$$\sum_{h=1}^H W_h \sum_{i=1}^{n_h} x_{hi} (\text{expit}(x_{hi}' B) - \text{expit}(x_{hi}' \hat{\beta}_h)) = 0$$

where $\text{expit}(\cdot) = \exp(\cdot)/(1 + \exp(\cdot))$.

2.5. Laplace Prior for Weight Smoothing

Instead of using a multivariate normal distribution as the prior of β_s , we propose using a multivariate Laplace distribution. Unlike the normal distribution prior which restricts the variation between random-effect term and prior mean in an L^2 -norm manner, Laplace

measures by the L^1 distance. This should allow more severe downweighting of interactions between the regression parameters and the probability of selection for which there is only weak evidence in the data, while preserving those that we need for bias correction.

The general form of multivariate Laplace distribution is given by [Eltoft et al. \(2006\)](#):

$$p_Y(y) = \frac{1}{(2\pi)^{d/2}} \frac{2}{\lambda} \frac{K_{(d/2)-1} \left(\sqrt{\frac{2}{\lambda}} q(y) \right)}{\left(\sqrt{\frac{2}{\lambda}} q(y) \right)^{(d/2)-1}}$$

where y is a d -dimensional random variable $y = (y_1, \dots, y_d)$; $K_m(x)$ denotes the modified Bessel function of the second kind and order m , evaluated at x ; $q(y) = (y - \mu)^T \Gamma^{-1} (y - \mu)$; $\Gamma = \{\gamma_{jk}\}$, $j, k = 1, \dots, d$ is a $d \times d$ matrix defining the internal covariance structure of the variable Y , $\mu = (\mu_1, \dots, \mu_d)$ is the vector of means, and λ an overall scale parameter. However, this format is inconvenient for application. The alternative approach is to represent the Laplace distribution as a scale mixture of normals with an exponential mixing density:

$$\begin{aligned} \beta_h | \beta_h^*, D_\tau, \sigma^2 &\sim MVN(\beta_h^*, \sigma^2 D_{\tau h}) \\ \beta_h^* | \sigma_0^2 &\sim MVN(0, \sigma_0^2 I_p) \\ D_{\tau h} &= \text{diag}(\tau_{h1}^2, \dots, \tau_{hp}^2) \\ \sigma^2, \tau_1^2, \dots, \tau_{Hp}^2 &\sim 1/\sigma^2 \prod_{j=1}^{Hp} \frac{\lambda^2}{2} e^{-\lambda^2 \tau_j^2 / 2} \\ \lambda^2 &\sim \text{Gamma}(\gamma, \delta) \end{aligned}$$

where γ and δ are known constants. The first level of the model depends on the distribution assumption of the generalized linear model used. In this article, we take linear regression and logistic regression as examples, and provide the full hierarchical Bayesian model and related Gibbs Sampler algorithm.

For linear regression, Y conditional on all other parameters follows a normal distribution. Assuming that the residual variance σ^2 is independent from the latent mixing variables τ_i , the hierarchical model is as follows:

$$\begin{aligned} y_{hi} | x_{hi}, \beta_h, \sigma^2 &\stackrel{\text{ind}}{\sim} N(x_{hi}^T \beta_h, \sigma^2) \\ \beta_h | \beta_h^*, D_\tau, \sigma^2 &\stackrel{\text{ind}}{\sim} MVN(\beta_h^*, \sigma^2 D_{\tau h}) \\ \beta_h^* | \sigma_0^2 &\stackrel{\text{ind}}{\sim} MVN(0, \sigma_0^2 I_p) \\ D_{\tau h} &= \text{diag}(\tau_{h1}^2, \dots, \tau_{hp}^2) \\ \sigma^2, \tau_1^2, \dots, \tau_{Hp}^2 &\sim 1/\sigma^2 \prod_{j=1}^{Hp} \frac{\lambda^2}{2} e^{-\lambda^2 \tau_j^2 / 2} \\ \lambda^2 &\sim \text{Gamma}(\gamma, \delta) \end{aligned}$$

where again γ and δ are known constants. Following the deduction in [Park and Casella \(2008\)](#), the analytical forms of all fully conditional distributions of β , σ^2 and so on exist, and the posterior predictive distribution could be obtained through a Gibbs Sampler as below. A detailed derivation is available online at www.doi.org/10.1515/jos-2016-0026, Appendix A.

$$\begin{aligned} \beta_h | rest &\sim MVN(A^{-1}(X_h^T Y_h + D_{\tau h}^{-1} \beta_h^*), \sigma^2 A^{-1}), A = X_h^T X_h + D_{\tau h}^{-1} \\ \beta_h^* | rest &\sim MVN\left((\sigma^2 D_{\tau h})^{-1} \left((\sigma^2 D_{\tau h})^{-1} + (\sigma_0^2 I)^{-1}\right)^{-1} \beta_h, \left((\sigma^2 D_{\tau h})^{-1} + (\sigma_0^2 I)^{-1}\right)^{-1}\right) \\ \sigma^2 | rest &\sim InvGamma\left((n + Hp)/2, \frac{1}{2} \left[\sum_{h=1}^H (Y_h - X_h \beta_h)^T (Y_h - X_h \beta_h) \right. \right. \\ &\quad \left. \left. + \sum_{h=1}^H (\beta_h - \beta_h^*)^T (D_{\tau h})^{-1} (\beta_h - \beta_h^*) \right] \right) \\ 1/\tau_{hi}^2 | rest &\sim InvGaussian\left(\sqrt{\frac{\lambda^2 \sigma^2}{(\beta_h - \beta_h^*)^2}}, \lambda^2\right) \\ \lambda^2 &\sim Gamma\left(Hp + \gamma, \frac{1}{2} \sum_{h=1}^H \sum_{i=1}^p \tau_{hi}^2 + \delta\right) \end{aligned}$$

At each step in the Gibbs Sampler chain, a draw of the linear regression population slope B is obtained from the draw of β_h as $B = [\sum_h W_h \sum_{i=1}^{n_h} x_{hi} x'_{hi}]^{-1} [\sum_h W_h (\sum_{i=1}^{n_h} x_{hi} x'_{hi}) \beta_h]$.

For logistic regression, the model is similar to that for linear regression, except that Y follows a binomial distribution, and estimation of σ^2 is no longer necessary:

$$\begin{aligned} y_{hi} | x_{hi}, \beta_h &\sim \prod_{h=1}^H \prod_{i=1}^{n_h} \left(\frac{\exp(x_{hi} \beta_h)}{1 + \exp(x_{hi} \beta_h)} \right)^{y_{hi}} \left(\frac{1}{1 + \exp(x_{hi} \beta_h)} \right)^{1-y_{hi}} \\ \beta_h | \beta_h^*, D_{\tau} &\sim MVN(\beta_h^*, D_{\tau h}) \\ \beta_h^* | \sigma_0^2 &\sim MVN(0, \sigma_0^2 I_p) \\ D_{\tau h} &= \text{diag}(\tau_{h1}^2, \dots, \tau_{hp}^2) \\ \tau_1^2, \dots, \tau_{Hp}^2 &\sim \prod_{j=1}^{Hp} \frac{\lambda^2}{2} e^{-\lambda^2 \tau_j^2 / 2} \\ \lambda^2 &\sim Gamma(\gamma, \delta) \end{aligned}$$

where γ and δ are known constants. When the first level is not normally distributed, the fully conditional distribution of β does not belong to any known distribution, and thus direct sampling is impossible. Instead we apply a Metropolis method, and the proposed β_h is drawn from $N_p(\beta'_h, c_{\beta} D_{\beta})$, for $D_{\beta} = (V_{\beta h}^{-1} + D_{\tau h}^{-1})^{-1}$, where β'_h is the ML estimate of the logistic regression of y on z from strata h , and $V_{\beta h}$ the associated covariance matrix

obtained from the expected information matrix evaluated at β_h . The proposed β_h is accepted with probability $r = \max[1, \{f_\beta(\beta_{prop})\}/\{f_\beta(\beta)\}]$, where f_β is the posterior distribution of β proportional to $p(\beta_h) \prod_{i=1}^{n_h} f(y_{hi}|\beta_h)$. All other parameters follow the Gibbs Sampler algorithm, and are directly drawn from their fully conditional distributions as below (full derivation is available online at www.doi.org/10.1515/jos-2016-0026, Appendix B):

$$\beta_h^*|rest \sim MVN\left((D_{\tau h})^{-1}\left((D_{\tau h})^{-1} + (\sigma_0^2 I)^{-1}\right)^{-1} \beta_h, (D_{\tau h})^{-1} + (\sigma_0^2 I)^{-1}\right)$$

$$1/\tau_{hi}^2|rest \sim InvGaussian\left(\sqrt{\frac{\lambda^2}{(\beta_h - \beta_h^*)^2}}, \lambda^2\right)$$

$$\lambda^2 \sim Gamma\left(Hp + \gamma, \frac{1}{2} \sum_{h=1}^H \sum_{i=1}^p \tau_{hi}^2 + \delta\right)$$

At each step in the Gibbs Sampler chain, a draw of the logistic regression population slope B is obtained from the draw of β_h by solving the weighted score equation $\sum_{h=1}^H W_h \sum_{i=1}^{n_h} x_{hi} (expit(x'_{hi} B) - expit(x'_{hi} \beta_h)) = 0$. In practice, B can be obtained by replacing the observed y_{hi} with the predicted values $g(x'_{hi} \beta_h)$ for each draw of $\hat{\beta}_h$ and obtaining the weighted pseudo-MLE for the logistic regression model.

3. Simulation Study

To evaluate the performance of weight-smoothing models using Laplace priors, we create several scenarios for ordinary linear regression and logistic regression, generating separate populations with normally distributed outcomes and dichotomized outcomes accordingly. We also consider scenarios where heteroscedasticity or multiple covariates occur. The target of interest is the population slope in a regression model. In addition to our Laplace prior estimator, we include an unweighted estimator, a fully weighted estimator, a normal-prior (exchangeable) estimator (Elliott and Little 2000; Elliott 2007), and several variations of the estimator proposed by Beaumont (2008) for comparison. For each scenario and estimator, we compute bias, square root of mean square error (RMSE) and coverage of 95% confidence or credible intervals as follows:

$$bias = S^{-1} \sum_{s=1}^s (\hat{B}_s - B)$$

$$RMSE = \sqrt{S^{-1} \sum_{s=1}^s (\hat{B}_s - B)^2}$$

$$coverage = S^{-1} \sum_{s=1}^s I(\hat{B}_s^L \leq B \leq \hat{B}_s^U)$$

where s indexes the independent samples drawn for each simulation, \hat{B}_s is the point estimator of the regression coefficient of interest, \hat{B}_s^L and \hat{B}_s^U correspond to the lower and upper bounds of the 95% confidence or credible interval, and B to the regression coefficient computed using the population data (i.e., the inference target of interest).

Relative root mean square error (RRMSE) is reported as the ratio of the estimator's RMSE to the fully weighted estimator's RMSE.

3.1. Hierarchical Weight Smoothing Model for Ordinary Linear Regression

We first generate a population of size $N = 20,000$. The predictor X is uniformly distributed between zero and ten, and is equally divided into 20 strata with intervals of 0.5 each. The response variable Y is generated as a spline function of X noted below, with knots located between strata. Three sets of coefficients of the spline are applied separately to represent the various patterns of $Y|X$ from straight slope to accelerating and decelerating curves.

$$Y_i|X_i, \beta, \sigma^2 \sim N\left(\beta_0 + \sum_{h=1}^{20} \beta_h(x_i - h/2)_+, \sigma^2\right)$$

$$X_i \sim \text{UNI}(0, 10), i = 1, \dots, N = 20,000$$

We assume $\beta_0 = 0$, and consider three sets of the spline coefficients

$$\beta_a = (2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$$

$$\beta_b = (0, 0, 0, 0, 0, 0, .25, .25, .25, .25, .5, .5, .5, .5, 1, 1, 1, 1, 1, 1)$$

$$\beta_c = (11, -1, -1, -1, -1, -1, -1, -0.5, -0.5, -0.5, -0.5, -0.25,$$

$$-0.25, -0.25, -0.25, 0, 0, 0, 0, 0).$$

We let the population variance σ^2 vary across $10^{1.5}$, $10^{3.5}$ and $10^{5.5}$ to create varying levels of background noise compared to changes in slope. Figure 1 shows each of the $3 \times 3 = 9$ populations generated for the linear regression.

From the population, samples of size $n = 1,000$ are repeatedly drawn without replacement, according to inclusion probabilities proportional to $\pi_h = (1 + h/30) * h$ for the h th stratum, which results in a ratio of 35 times between the maximum and minimum probabilities. We also ensure that the sample size of each stratum is greater than three for computation convenience. Z is created as $Z = I \otimes X$, where $I = c(I_1, \dots, I_h)$ is an indicator vector marking that the current observation belongs to the i th stratum. It is also centered within each column with respect to each stratum, and used as predictor in the simulations. Thus β_a corresponds to a linear model (no model misspecification); β_b to a setting where the nonlinearity is greatest where the probability of selection is the highest, and β_c to a setting where the nonlinearity is greatest where the probability of selection is the lowest.

Our inferential target is $B = \left(\sum_{i=1}^N \tilde{X}_i \tilde{X}_i'\right)^{-1} \sum_{i=1}^N \tilde{X}_i Y_i$ for $\tilde{X}_i = (1 \ X_i)'$, the least-squares linear approximation of Y to X . Under β_b and β_c , weights correct bias from model misspecification. Under β_a , the model is correctly specified, suggesting that the unweighted estimator is most efficient. Also note that under β_b , the curvature is largest where the data are most densely sampled, while the reverse is true under β_c , suggesting that varying degrees of trimming will be required to optimize the bias-variance tradeoff.

For the hyperprior parameters, σ_0^2 is arbitrarily defined as 1,000 to approximate a noninformative prior; the prior for λ follows a gamma hyperprior with parameter $\gamma = 1$ and $\delta = 1.78$, as suggested by Park and Casella (2008). All other parameters in simulation are initialized at zero, except for the variance estimator σ^2 , which is initialized at one.

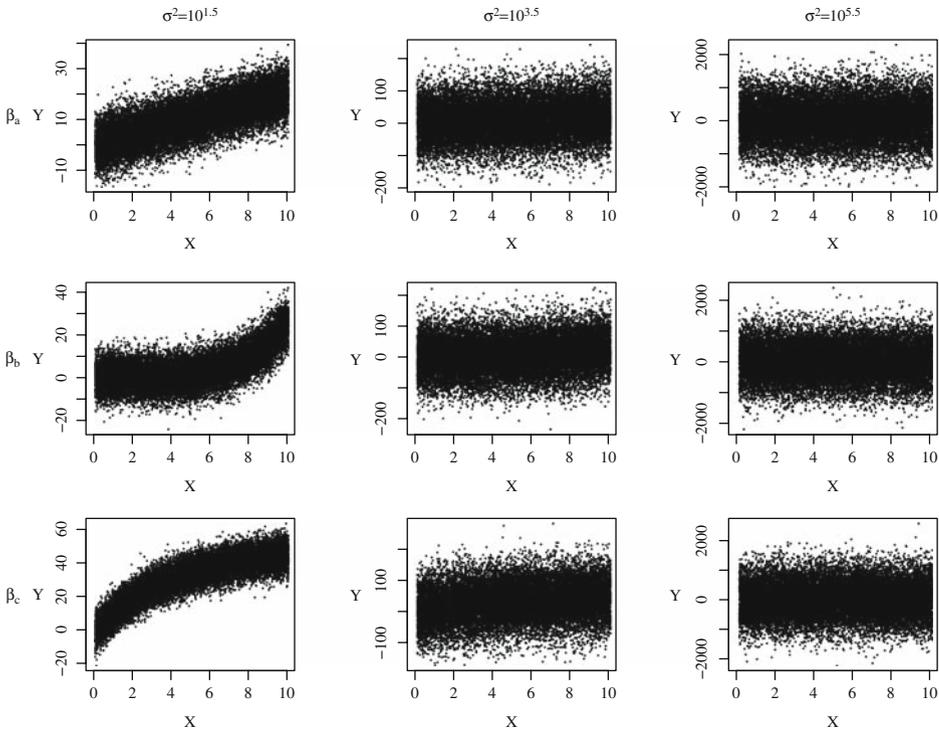


Fig. 1. Scatter plot of population for the linear regression simulation

A Gibbs Sampler method is applied, that is, all parameters are sequentially drawn from the full conditional distribution for each iteration. Then, to obtain the estimate from the posterior predictive distribution, the unobserved Y are generated based on sampled parameters from each iteration, and the target population slope B is obtained by fully weighted regression on observed and predicted Y . The process iterates 10,000 times, with a burn-in of 2,000. Diagnostic plots are generated to ensure the algorithm’s convergence via visual inspection. Overall, 200 samples are generated from each population to provide the empirical distribution for the repeated measures properties.

We compare the properties of our Laplace model (HWS) with major competitors, including the unweighted model (UNWT), fully weighted model (FWT), weight-smoothing model with normal prior and exchangeable random-slope assumption (XRS), and two variations of the estimators proposed by Beaumont (2008): predicted weights on y and x (PREDYX) and predicted weights on degree 5 polynomial of y , together with x (PREDYX5). Bias and nominal 95% coverage are recorded directly, while RMSE is rescaled relative to the fully weighted estimator. Results are provided in Tables 1, 2, and 3.

Under β_a , where the model is correctly specified, all methods yield unbiased results, and the unweighted estimator maintains the best efficiency, with an approximate 30% decrease in RMSE compared to the fully weighted estimator. The original weight-smoothing method under XRS tends to provide unstable results, inflating the variance when the population signal is strong, but achieving similar RMSE as the unweighted estimator when the population signal is weak relative to the noise. Our model, under the same XRS

Table 1. Comparisons of bias, relative root mean square error (RRMSE) and coverage of nominal 95% confidence interval/credible interval (CI) of various estimators for the linear regression coefficient B under β_a using populations with different residual variances corresponding to various models under consideration

Estimator	$\sigma^2 = 10^{1.5}$				$\sigma^2 = 10^{3.5}$				$\sigma^2 = 10^{5.5}$			
	Bias	RRMSE	95% CI coverage		Bias	RRMSE	95% CI coverage		Bias	RRMSE	95% CI coverage	
UNWT	-0.022	0.684	0.95		-0.079	0.637	0.97		0.156	0.644	0.95	
FWT	-0.011	1	0.94		0.093	1	0.95		-0.222	1	0.93	
HWS	-0.011	1.025	0.96		0.082	0.859	0.98		-0.906	0.741	0.99	
XRS	-0.012	1.278	0.99		-0.062	0.678	0.96		-0.166	0.651	0.96	
PREDYX	-0.011	0.791	0.97		0.083	0.783	0.95		-0.375	0.756	0.94	
PREDYX5	-0.011	0.952	0.94		0.078	0.944	0.94		-1.089	0.960	0.94	

Table 2. Comparisons of bias, relative root mean square error (RRMSE) and coverage of nominal 95% confidence interval/credible interval (CI) of various estimators for the linear regression coefficient B under β_B using populations with different residual variances corresponding to various models under consideration

Estimator	$\sigma^2 = 10^{1.5}$			$\sigma^2 = 10^{3.5}$			$\sigma^2 = 10^{5.5}$		
	Bias	RRMSE	95% CI coverage	Bias	RRMSE	95% CI coverage	Bias	RRMSE	95% CI coverage
UNWT	0.939	5.316	0	0.876	0.807	0.80	2.481	0.666	0.96
FWT	-0.010	1	0.96	0.093	1	0.95	2.736	1	0.93
HWS	-0.011	0.887	0.97	0.050	0.864	0.98	1.047	0.747	1
XRS	-0.063	1.012	0.95	0.869	0.873	0.94	1.134	0.729	0.94
PREDYX	0.114	1.375	0.52	0.385	0.735	0.91	0.746	0.726	0.93
PREDYX5	0.025	1.000	1	0.022	0.955	0.95	1.515	0.974	0.95

Table 3. Comparisons of bias, relative root mean square error (RRMSE) and coverage of nominal 95% confidence interval/credible interval (CI) of various estimators for the linear regression coefficient B under β_c using populations with different residual variances corresponding to various models under consideration

Estimator	$\sigma^2 = 10^{1.5}$				$\sigma^2 = 10^{3.5}$				$\sigma^2 = 10^{5.5}$			
	Bias	RRMSE	95% CI coverage		Bias	RRMSE	95% CI coverage		Bias	RRMSE	95% CI coverage	
UNWT	-0.925	4.517	0		-0.976	0.848	0.76		-0.738	0.643	0.95	
FWT	-0.011	1	1		0.094	1	0.95		-0.225	1	0.93	
HWS	-0.010	0.779	0.95		0.171	0.875	0.98		-0.876	0.743	0.99	
XRS	-0.264	1.527	0.26		-0.953	1.053	0.64		-0.729	0.662	0.93	
PREDYX	-0.236	1.301	0.65		-0.382	0.991	0.85		-0.616	0.759	0.92	
PREDYX5	-0.010	1.018	1		-0.043	0.965	0.96		-0.552	0.974	0.95	

assumption but with a Laplace prior, gives more stable results that resemble the fully weighted estimator when variance is low, but increase in efficiency as the population variance increases. Both the XRS and HWS estimators have correct to somewhat conservative coverage when the linear model is correctly specified. The Beaumont estimator PREDYX has improved RMSE compared to the fully weighted estimator, since it models the weight as a linear function of X ; PREDYX5 yields estimates similar to the fully weighted estimator, as variance reduction was minimal due to the increased variability in the model weights.

For scenarios under β_b and β_c , the unweighted estimator of B is biased, and the fully weighted estimator strongly prevails over the unweighted estimator with respect to both RMSE and true coverage for small to moderate levels of residual variances. The weight-smoothing method under XRS remains biased at moderate levels of variance for β_b and β_c , and also at small levels of variance for β_c , raising RMSE relative to FWT and destroying nominal coverage, suggesting that the exchangeable random-slope structure is not able to capture the relation in mean and variance among strata. The weight-smoothing estimator with Laplace prior has limited bias similar to that of the fully weighted estimator, but very substantially reduced RMSE, with correct to conservative coverage. The PREDYX estimator is insufficiently structured to reduce bias in the small-to-medium residual-variance settings; PREDYX5 mimics the fully weighted estimator and thus has little savings in relative RMSE under any of the scenarios.

3.2. Hierarchical Weight Smoothing Model for Logistic Regression

Following Elliott (2007), we set up populations in two settings: model misspecification and informative sampling. For model misspecification, the population is equally divided into 20 strata, and the predictor X is uniformly distributed within each stratum on an interval ranging from $0.5(h - 1)$ to $0.5h$. The binary response variable is generated as follows:

$$P(Y_i = 1|X_i) \sim BER(\text{expit}(1.5 - .75X_i + C * X_i^2)),$$

$$X_{hi} \sim UNI(0.5*(h - 1), 0.5*h), h = 1, \dots, 20, i = 1, \dots, 1000$$

Our inferential target is $B = (B_0 \ B_1)'$, the value of $\beta = (\beta_0 \ \beta_1)'$ that solves the score equation $U(\beta) = \sum_{i=1}^N \tilde{X}_i(Y_i - \text{expit}(\tilde{X}_i'\beta))$, corresponding to the best linear approximation to X_i and $\log\left(\frac{E(Y_i|X_i)}{1-E(Y_i|X_i)}\right)$. For C , we consider values of 0, .027, .045, .061, .080, corresponding to no model misspecification at $c = 0$ to increasing levels of model misspecification. The selection probability for each observation remains the same within each stratum, and increases linearly along strata, with a ratio between maximum and minimum probabilities equal to 20.

For the informative sampling setting, we follow the same formula of

$$P(Y_i = 1|X_i) \sim BER(\text{expit}(1.5 - .75X_i + C * X_i^2)),$$

$$X_{hi} \sim UNI(0.5*(h - 1), 0.5*h), h = 1, \dots, 20, i = 1, \dots, 1000$$

but fix $C = 0$, so the model is correctly specified. We also create a vector of binary value Z_i^* such that $Cor(Y_i, Z_i^*) = r$, and let r range from 0.05 to 0.95 to represent different levels

of correlation with Y . Then we let $Z_i = Z_i^* U_i + (1 - Z_i^*) X_i$, where $U_i \sim U(0, 10)$ independent of X_i , and the selection probability is proportional to Z_i . Thus whether the selection probability is related to X or not is determined by the value of Z^* , which is correlated with Y to some level. The process results in a ratio of roughly 30 between the maximum weight and minimum weight, with the correlation between selection probability and Y varying from 0 to 30% as the correlation between Z^* and Y increases from .05 to .95. Twenty strata of equal size are created by pooling observations with similar selection probabilities together.

From this population, samples with $n = 1,000$ are selected without replacement, with the selection probability stated above. We create weight strata using the values of h . A total of 200 samples are generated to create the empirical distribution for inference. A single MCMC chain is built for each data set, and for each iteration in the algorithm, all parameters are sequentially drawn from the full conditional distribution, except for β , which is drawn via a Metropolis step (proposed from a normal distribution centered at MLE with inverse expected information as covariance matrix, and accepted according to likelihood ratio times prior distribution). Then the predicted Y is calculated based on drawn parameters, and the target population slope is obtained by fully weighted logistic regression. The initial values of parameters are assigned the same as linear regression setting, and the process iterates 10,000 times, with a burn-in of 2,000.

We compare the properties of our Laplace model (HWS) with the same major competitors as in the linear regression setting. Bias and nominal 95% coverage are recorded directly, while RMSE is rescaled according to the fully weighted estimator. Results are provided in [Table 4](#) and [5](#).

While comparing different models under the model misspecification settings, the unweighted model has increased bias as the population model is less correctly specified, resulting in a change from an efficient estimate to a poor estimate (RMSE ratio from 69.7% to 281.9% of FWTs as C increases) and poor coverage as misspecification increases. The exchangeable random-slope model estimator is not robust, with bias similar to the unweighted model, and larger RMSE than the fully weighted estimator, although coverage is conservative. The hierarchical weight smoothing model with Laplace prior provides a more robust estimator, with minimal bias, and RMSE reduced by up to 14% compared to the FWT estimator, with true coverage similar to that of the fully weighted estimator. The weight-prediction model PREDYX performs similarly to the unweighted estimator, gaining efficiency when the model is correctly specified, and suffering as misspecification increases. PREDYX5, which predicts weights with a degree-five polynomial of x , essentially mimics the fully weighted estimator.

Under informative sampling, the unweighted estimator has only slightly larger RMSE than the fully weighted estimator, but is substantially biased with poor coverage. The exchangeable random-effect model has a similar degree of bias compared to the unweighted estimator, but has increased variability that, while providing conservative coverage, yields substantially increased RMSE over the fully weighted estimator. The hierarchical weight-smoothing model with Laplace prior again provides a more robust estimator, with minimal bias, and RMSE reduced by up to twelve percent compared to the FWT estimator, although true coverage suffers to a moderate degree except when the

Table 4. Comparisons of bias, relative root mean square error (RRMSE) and coverage of nominal 95% confidence interval/credible interval (CI) of various estimators for the logistic regression coefficient using populations with different degrees of model misspecification corresponding to various models under consideration

Estimator	c = 0			c = .45			c = .80		
	Bias	RRMSE	95% CI coverage	Bias	RRMSE	95% CI coverage	Bias	RRMSE	95% CI coverage
UNWT	0.014	0.697	0.96	0.063	1.151	0.47	0.132	2.819	0
FWT	0.006	1	0.84	-0.015	1	0.82	-0.014	1	0.94
HWS	0.011	0.915	0.84	-0.014	0.909	0.82	-0.013	0.860	0.88
XRS	0.038	1.125	1	0.078	1.696	0.94	0.042	1.644	0.98
PREDYX	0.003	0.791	0.96	0.021	0.903	0.92	0.038	1.123	0.79
PREDYX5	-0.004	0.962	0.93	-0.005	0.965	0.94	-0.001	0.967	0.98

Table 5. Comparisons of bias, relative root mean square error (RRMSE) and coverage of nominal 95% confidence interval/credible interval (CI) of various estimators for the logistic regression coefficient using populations with different degrees of informative sampling corresponding to various models under consideration

Estimator	$r = .05$				$r = .50$				$r = .95$			
	Bias	RRMSE	95% CI coverage		Bias	RRMSE	95% CI coverage		Bias	RRMSE	95% CI coverage	
UNWT	0.057	0.990	0.76		0.069	1.155	0.52		0.053	0.914	0.64	
FWT	0.023	1	0.94		0.009	1	0.88		0.001	1	1	
HWS	0.022	0.906	0.82		0.009	0.914	0.84		0.001	0.875	0.96	
XRS	0.067	1.417	1		0.071	1.463	0.98		0.059	1.272	1	
PREDYX	0.021	0.832	0.94		0.023	0.859	0.91		0.031	0.880	0.93	
PREDYX5	0.004	0.977	0.94		0.002	0.969	0.95		0.007	0.997	0.97	

sampling is highly informative. PREDYX improves RMSE by up to 17% while having only slight undercoverage. PREDYX5 again mimics the fully weighted model.

3.3. Hierarchical Weight-Smoothing Model for Heteroscedasticity Scenario

In this section we evaluate the performance of the hierarchical weight-smoothing model compared to unweighted and fully weighted models under the heteroscedasticity setting. Here the main purpose of weighting is not to correct bias, but to adjust for the violation in the homoscedasticity assumption, and to yield correct inference on target quantities. We expect the data-driven method to capture the heteroscedasticity pattern in data well, and lead to proper inference.

First, we create a population of size $N = 20,000$. The interval between zero and ten is evenly divided into ten strata with a length of one, and the predictor X is uniformly distributed within each stratum. The response variable Y is then generated from a normal distribution, with mean equal to twice of X , and variance as an increasing function of X :

$$X_i \sim UNI(0, 10), i = 1, \dots, N = 20,000$$

$$P_i = (1 + [X_i]/30) * [X_i]/2$$

$$Y_i | X_i, \sigma^2 \sim N(2 * X_i, P_i * \sigma^2)$$

where the population variance σ^2 is set to 10^1 , 10^3 and 10^5 to adjust for different scales of population variance.

We repeatedly draw samples from the population without replacement by inclusion probabilities proportional to P , to link the corresponding weights to the heteroscedasticity pattern. The inferential target remains the population slope; all other settings for parameters and simulations remain the same. Altogether 200 samples are drawn, and the HWT model is fit using 10,000 iterations, with 2,000 as burn-in. To evaluate the results, we compare bias, relative RMSE and true coverage of the nominal 95% confidence interval or credible interval across the unweighted model, fully weighted model, and hierarchical weight-smoothing model in [Table 6](#).

The results suggest that violation in homoscedasticity undermines the performance of the unweighted estimator when the population variance is small. But as the population variance increases, this effect is quickly overwhelmed by concerns about efficiency, where the unweighted estimator has about a 35% reduction in RMSE compared to the fully weighted estimator. The hierarchical weight-smoothing model performs well in the heteroscedastic setting, where it correctly retains the weight interactions at low variances, yet “tunes them out” when efficiency is the dominating component.

3.4. Hierarchical Weight-Smoothing Model with Multiple Covariates

In the last simulation study we focus on the hierarchical weight-smoothing model’s performance when multiple covariates exist in the model. The challenge lies in that some covariates may be related to the sampling scheme, and thus could benefit from weighting, but some may be independently distributed, and could lose efficiency in weighting. Simply applying the fully weighted or the unweighted model will sacrifice either variance for

Table 6. Comparisons of bias, relative root mean square error (RRMSE) and coverage of nominal 95% confidence interval/credible interval (CI) of various estimators

Estimator	$\sigma^2 = 10$			$\sigma^2 = 10^3$			$\sigma^2 = 10^5$		
	Bias	RRMSE	95% CI coverage	Bias	RRMSE	95% CI coverage	Bias	RRMSE	95% CI coverage
UNWT	-0.030	1.432	0.95	0.009	0.608	0.95	-0.766	0.679	0.96
FWT	-0.002	1	0.97	-0.012	1	0.93	-1.253	1	0.96
HWS	-0.004	1	1	-0.020	0.896	0.96	-1.112	0.653	0.99

covariate associations unrelated to the sampling scheme, or bias for covariate associations related to the sampling scheme; we expect our data-driven method could reach a balance in between these options.

To replicate such a situation, we look into three different scenarios: independent covariates X_1 and X_2 with model misspecification on X_1 , and weight related to X_1 ; independent covariates X_1 and X_2 with model misspecification on X_1 , and weight related to X_2 ; and correlated covariates X_1 and X_2 , with model misspecification on X_1 , and weight related to X_1 .

For the first scenario, we generate a population of size $N = 20,000$, with two covariates X_1 and X_2 independently drawn from uniform distributions on an interval between zero and ten. The outcome Y is generated from a normal distribution with mean equal to a spline function of X_1 plus X_2 , thus our linear approximation to X_1 leads to model misspecification.

$$Y_i|X_{1i}, X_{2i}, \beta, \sigma^2 \sim N(\beta_0 + x_{2i} + \sum_{h=1}^{10} \beta_h(x_{1i} - h)_+, \sigma^2)$$

$$X_{1i} \sim UNI(0, 10), i = 1, \dots, N = 20,000$$

$$X_{2i} \sim UNI(0, 10), i = 1, \dots, N = 20,000$$

$$\beta = (0, 0, 0.5, 0.5, 0.5, 0.5, 1, 1, 1, 2)$$

where the variance σ^2 is set to 10^1 , 10^3 , and 10^5 for variation in background noise.

The inclusion probabilities are proportional to $\pi_i = (1 + [X_{1i}]/30) * [X_{1i}]/2$, and samples of size $n = 1,000$ are repeatedly drawn. Weights equal to inverse inclusion probabilities naturally create ten strata. For computation convenience, we also create Z as $Z = I \otimes (X_1, X_2)$, where $I = c(I_1, \dots, I_h)$ indicates to which stratum an observation belongs. We assess the model performance through its inference on the population slopes of Y to X_1 and X_2 : $\tilde{B} = (B_0, B_1, B_2) = \left(\sum_{i=1}^N \tilde{X}_i \tilde{X}_i' \right)^{-1} \sum_{i=1}^N \tilde{X}_i Y_i$ for $\tilde{X}_i = (1 \ X_{1i} \ X_{2i})'$.

When initializing the process, we retain all previous hyperprior parameters settings. The same Gibbs Sampler method is applied, and the predicted Y is generated to yield inference on target population slopes B_1 and B_2 . The process iterates 10,000 times, with a burn-in of 2,000. We conduct 200 simulations to provide the empirical distribution to estimate repeated measurement properties. Table 7 compares bias, relative RMSE and true coverage of the nominal 95% confidence interval or credible interval across the hierarchical weight-smoothing method, unweighted model and fully weighted model.

The results in Table 7 suggest that when the population variance is small, the unweighted model result suffers from model misspecification and has a substantially larger bias compared to the fully weighted estimator. Yet for the covariate on which the model is correctly specified, the unweighted estimator gains efficiency and yields an RMSE about half the size of the RMSE of the weighted estimator. As background variance increases, the influence of bias decreases and the unweighted estimator prevails over the fully weighted estimator in estimating both X_1 and X_2 . The hierarchical weight-smoothing model closely resembles the fully weighted estimator with minor improvements. That is, it

Table 7. Comparisons of bias, relative root mean square error (RRMSE) and coverage of nominal 95% confidence interval/credible interval (CI) of various estimators for the multivariate linear regression coefficients B_1 , B_2 when weight is related to X_1 using populations with different residual variances corresponding to various models under consideration

Estimator	$\sigma^2 = 10$			$\sigma^2 = 10^3$			$\sigma^2 = 10^5$		
	Bias	RRMSE	95% CI coverage	Bias	RRMSE	95% CI coverage	Bias	RRMSE	95% CI coverage
B_1									
UNWT	0.976	6.592	0	0.947	1.293	0.55	2.747	0.697	0.99
FWT	-0.074	1	0.93	-0.385	1	0.97	1.519	1	0.93
HWS	-0.076	1.007	0.84	-0.377	0.999	0.96	1.442	0.992	0.99
B_2									
UNWT	-0.013	0.440	0.94	0.075	0.392	0.97	-1.754	0.490	0.93
FWT	-0.012	1	0.89	0.153	1	0.95	-0.656	1	0.89
HWS	-0.014	0.991	0.88	0.160	0.993	0.98	-0.841	0.990	0.91

correctly applies weight when model misspecification occurs, but fails to tune the result when efficiency is more important.

For the second scenario, we follow the same setting as the previous scenario, except that the inclusion probabilities are now proportional to $\pi_i = (1 + [X_{2i}]/30) * [X_{2i}]/2$. Since the weights are related to the covariate that is correctly specified in the model, and independent from the covariate that requires adjustment, we expect no bias correction from the fully weighted estimator, and better efficiency from the unweighted estimator.

We keep the same parameter initialization and simulation setting. The simulation consists of 200 samples, 10,000 iterations within each sample, including 2,000 burn-in. The bias, relative RMSE and true coverage of the nominal 95% confidence interval or credible interval are reported in Table 8.

As expected, the result from the unweighted model is more efficient than the fully weighted model, leading to a 30% to 70% reduction in RMSE. Since weighting is unnecessary for either X_1 or X_2 according to the population setup, the weight-smoothing model is able to limit the side effect of weighting, and achieves on average a 30% reduction in RMSE compared to the weighted model. Also note that the weight-smoothing model result suffers a moderate drop in the true coverage of a nominal 95% credible interval when the population variance is small relative to the misspecification.

For the last scenario, we study the model behavior when X_1 and X_2 are correlated. For this purpose, we create X_1 by the former approach, that is, from a uniform distribution on interval between zero and ten, but define X_2 as having a uniform distribution centered at X_1 to yield a correlation of about 0.45 between X_1 and X_2 . The rest of the settings stay the same:

$$Y_i | X_{1i}, X_{2i}, \beta, \sigma^2 \sim N\left(\beta_0 + x_{2i} + \sum_{h=1}^{10} \beta_h (x_{1i} - h)_+, \sigma^2\right)$$

$$X_{1i} \sim UNI(0, 10), i = 1, \dots, N = 20,000$$

$$X_{2i} \sim UNI(0, 10) + X_{1i}, i = 1, \dots, N = 20,000$$

$$\beta = (0, 0, 0.5, 0.5, 0.5, 0.5, 1, 1, 1, 2)$$

Similarly, three settings of variance are considered, and the inclusion probabilities are related to X_1 according to the formula $\pi_i = (1 + [X_{1i}]/30) * [X_{1i}]/2$. The simulation again consists of 200 samples and 10,000 iterations within each sample, including 2,000 burn-in. The results are presented in Table 9.

Although X_1 and X_2 are correlated in this scenario, the results are very close to the first scenario. The weighted method is useful when model misspecification exists and the population variance is small, but loses to the unweighted method due to lack of efficiency when the model is correctly specified, or the variance is large compared to potential biasedness. The hierarchical weight-smoothing model fails to balance the two situations, closely resembling the fully weighted method.

Combining all three scenarios, we conclude that the weight-smoothing model with Laplace prior has large gains in efficiencies when weighting is not necessary for any of the covariates, at the cost of a moderate drop in the true coverage rate when the residual variance is small. In other settings, its performance is similar to the fully weighted estimator.

Table 8. Comparisons of bias, relative root mean square error (RRMSE) and coverage of nominal 95% confidence interval/credible interval (CI) of various estimators for the multivariate linear regression coefficients B_1 , B_2 when weight is related to X_2 using populations with different residual variances corresponding to various models under consideration

Estimator	$\sigma^2 = 10$			$\sigma^2 = 10^3$			$\sigma^2 = 10^5$		
	Bias	RRMSE	95% CI coverage	Bias	RRMSE	95% CI coverage	Bias	RRMSE	95% CI coverage
B_1									
UNWT	-0.009	0.478	0.95	0.005	0.286	1	0.205	0.420	1
FWT	0.018	1	0.86	0.096	1	0.92	1.856	1	0.92
HWS	0.021	0.912	0.77	0.067	0.565	0.96	1.741	0.794	0.95
B_2									
UNWT	0.022	0.533	0.98	-0.254	0.629	0.97	0.833	0.748	0.95
FWT	0.044	1	0.95	0.045	1	0.95	-1.517	1	0.98
HWS	0.029	0.683	0.74	-0.091	0.615	0.91	1.230	0.698	0.97

Table 9. Comparisons of bias, relative root mean square error (RRMSE) and coverage of nominal 95% confidence intervals/credible interval (CI) of various estimators for the multivariate linear regression coefficients B_1 , B_2 when weight is related to X_1 and X_2 are correlated, using populations with different residual variances corresponding to various models under consideration

Estimator	$\sigma^2 = 10$			$\sigma^2 = 10^3$			$\sigma^2 = 10^5$		
	Bias	RRMSE	95% CI coverage	Bias	RRMSE	95% CI coverage	Bias	RRMSE	95% CI coverage
B_1									
UNWT	0.942	6.934	0	1.005	1.024	0.57	0.749	0.527	0.94
FWT	-0.071	1	0.96	0.067	1	0.96	-0.638	1	0.88
HWS	-0.099	1.118	0.85	0.033	0.999	0.95	-0.756	0.995	0.96
B_2									
UNWT	0.015	0.393	0.93	0.033	0.499	1	0.982	0.533	1
FWT	0.006	1	0.95	0.031	1	0.95	1.695	1	0.97
HWS	0.023	1.016	0.77	0.050	1.005	0.92	1.478	0.966	0.99

4. Application

4.1. Application on Dioxin data from NHANES

To demonstrate the performance of our method in the linear regression setting, we consider its application on the dioxin dataset from the National Health and Nutrition Examination Survey (NHANES). During the 2003–2004 survey, 1,250 representative adult subjects were selected in a probability sample of the US, and had their blood biomarkers measured, including 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD), a compound usually formed through incomplete combustion such as incineration, paper and plastics manufacturing, and smoking. Other demographic variables including age and gender are available from the survey. The sampled data are stratified into 25 strata, with each consisting of two Masked Variance Units (MVU's) to account for geographic clustering in the sample design without compromising confidentiality. Survey weights are provided as well. Due to technical limits, 674 readings are below limit of detection, and are imputed through multiple imputation using the model described in [Chen et al. \(2010\)](#), resulting in five replicate data sets. Both survey structure and imputation are incorporated in analysis using a jackknife method and Rubin's formula ([Rubin 1987](#)).

To determine the connection between log of TCDD level and individual demographic information, four linear regression models are fitted as log TCDD on age, log TCDD on gender, log TCDD on age and gender, and log TCDD on age, gender, and interaction. The hierarchical model is built as described before, with same initial value of parameters as those in the simulation. For each model setting, the unweighted (UNWT), fully weighted (FWT), and the hierarchical weight-smoothing (HWS) estimators are obtained (the exchangeable random-slope model fails to converge and is removed from the result). To estimate mean square error, the fully weighted version is treated as unbiased. Note that the fully weighted estimator is unbiased only in expectation, leading to an unbiased estimated square bias of regression coefficient $\hat{\beta}$ given by $\max(0, (\hat{\beta} - \hat{\beta}_w)^2 - \hat{V}_{01})$, where $\hat{V}_{01} = \widehat{\text{Var}}(\hat{\beta}) + \widehat{\text{Var}}(\hat{\beta}_w) - 2\widehat{\text{Cov}}(\hat{\beta}, \hat{\beta}_w)$ ([Little et al. 1997](#)). To fully account for the design features, all variance/covariance estimates are calculated via jackknife as $\widehat{\text{Var}}(\hat{\beta}_w) = \sum_h \frac{k_h-1}{k_h} \sum_{i=1}^{k_h} (\hat{\beta}_{w_{(hi)}} - \hat{\beta}_w)^2$, $\hat{\beta}_{w_{(hi)}} = (X'W_{(hi)}X)^{-1}XW_{(hi)}y$, where $\hat{\beta}_{w_{(hi)}}$ denotes the weighted β estimator from sample excluding i_{th} MVU in h_{th} stratum, and $W_{(hi)}$ is a diagonal matrix consisting of case weight w_j for all elements $j \notin i$, $w_j \frac{k_h}{k_h-1}$ for all elements $j \in h, j \notin i$, and 0 for elements $j \in h, j \in i$. $\widehat{\text{Var}}(\hat{\beta})$ and $\widehat{\text{Cov}}(\hat{\beta}_w, \hat{\beta})$ are calculated accordingly, and estimates from five imputed replicate datasets are combined using Rubin's formula ([Rubin 1987](#)). All Gibbs Sampler estimates are based on 10,000 iterations after discarding 2,000 draws as burn-in. The resulting bias and RMSE estimates are summarized in [Tables 10 through 13](#).

For the first two models of log TCDD on age and gender separately, the estimation of the single predictor using an unweighted model appears to be biased compared to the fully weighted model, resulting in an estimated bias of about 40% and 70% of RMSE. However, the weighted model also fails to provide an efficient estimate for effect on age, supported by a RMSE of 3.888, larger than the RMSE of 3.265 from the unweighted model. Meanwhile, the hierarchical weight-smoothing model shows its ability to improve efficiency, both reducing the bias comparing to the unweighted model, and maintaining a

Table 10. Regression of log TCDD on Age. Bias and RMSE for linear slope estimated for age: unweighted, fully weighted and hierarchical weight smoothing

Estimator	Est. _{.95%CI}	Bias(10^{-3})	RMSE(10^{-3})
UNWT	.0331 _{.0284,.0378}	- 1.262	3.265
FWT	.0343 _{.0266,.0420}	0	3.888
HWS	.0343 _{.0319,.0367}	- 0.086	1.214

RMSE similar to or smaller than the fully weighted model depending on the severity of variance inflation.

As more predictors enter the model, the estimated bias rapidly decreases in scale, leading to a scenario in which both bias and inflation in variance could dominate the overall RMSE, and neither the unweighted model nor the fully weighted model prevails in estimating all predictors. Hence the hierarchical weight-smoothing model cannot reduce bias further, yet it succeeds in reducing variance, resulting in overall smaller RMSE comparing to both the unweighted and fully weighted estimator (although the narrowness of the interval suggests that its coverage may be compromised to some degree, as in Subsection 3.4 of the simulation study).

4.2. Application on Partners for Child Passenger Safety Data

In this section, we use a Partners for Child Passenger Safety dataset to demonstrate our method’s performance under a logistic regression setting. Unit observations in the dataset are damaged vehicles disproportionately sampled from State Farm Insurance claims records between December 1998 and December 2005, when at least one child occupant less than 16 years of age was a passenger in a model year 1990 or newer State Farm-insured vehicle with a damage claims report. The focus of the study is children’s consequential injuries, defined by either facial lacerations or other injuries rated two or more on the Abbreviated Injury Scale (AIS) ([Association for the Advancement of Automotive Medicine 1990](#)). Due to the rare occurrence of the injury among all claims, to improve accuracy of the corresponding estimation of this rare outcome, the overall population was divided into three strata based on injury status – vehicles with at least one child occupant screened positive for injury at the time of the crash, vehicles with all child occupants reported receiving medical treatment but screened negative for injury, and vehicles with no occupants receiving medical treatment – and crossed with two strata defined by whether the vehicle was driveable or not. Since the stratification was associated

Table 11. Regression of log TCDD on Gender. Bias and RMSE for linear slope estimated for gender: unweighted, fully weighted and hierarchical weight smoothing

Estimator	Est. _{.95%CI}	Bias(10^{-2})	RMSE(10^{-1})
UNWT	.154 _{.002,.306}	- 8.219	1.248
FWT	.236 _{.110,.362}	0	0.637
HWS	.242 _{.122,.362}	0.589	0.607

Table 12. Regression of log TCDD on age and gender. Bias and RMSE for linear slope estimated for age and gender: unweighted, fully weighted and hierarchical weight smoothing

Estimator	Age			Gender		
	Est. _{.95%CI}	Bias(10^{-4})	RMSE (10^{-3})	Est. _{.95%CI}	Bias(10^{-2})	RMSE (10^{-2})
UNWT	.0336 _{.0287,.0385}	-9.067	3.296	.256 _{.108,.404}	-0.159	9.017
FWT	.0345 _{.0268,.0422}	0	3.895	.254 _{.132,.377}	0	6.161
HWS	.0344 _{.0320,.0368}	-0.841	1.227	.265 _{.153,.377}	1.058	5.659

with risk of injury (ascertained by follow-up survey), and cannot be fully explained by other auxiliary variables, the sampling design is informative, with weights varying from one to 50, and nine percent of weights lying outside three times their standard deviation.

As determined by [Winston et al. \(2002\)](#), children rear seated in compacted extended cab pickups are at greater risk of consequential injuries than children rear seated in other vehicles. To strengthen the conclusion, two models are applied, the unadjusted logistic model of injury status on car type (compacted extended cab pickups or others), and adjusted logistic model adapting control variables including child age (years), use of restraint (Y/N), intrusion into the passenger cabin in accident (Y/N), tow away after accident (Y/N), direction of impact (front/side/rear/other), and weight of the vehicle (pounds). The logistic hierarchical weight-smoothing model is set up as stated in the previous section, then the Gibbs sampler is executed for 10,000 iterations with 2,000 burn-in, and odds ratios are compared with the unweighted and fully weighted model. As this is a disproportionally stratified sample design, standard variance estimators are used for the unweighted and fully weighted estimators, while the posterior predictive distribution of the HWT model is used to compute point estimates and 95% credible intervals for the HWT estimator.

The estimated odds ratios for compacted extended cab pickups indicator did not vary much from the unadjusted model to the fully adjusted model, while unweighted regression and fully weighted regression lead to quite different results, from an OR of 3.534 to 11.317 for the unadjusted model, and from 3.448 to 13.890 when all other control variables are included (see [Table 14](#)). The hierarchical weight-smoothing model estimates lie in between the unweighted and weighted estimates, although much closer to the fully weighted model. It is also worth noting that with similar point estimates, the HWS model provides a considerable reduction in estimated standard deviation, leading to a narrower 95% confidence interval compared to the fully weighted model, a characteristic also presented in the previous simulation study.

5. Discussion

Model-based approaches to “trimming” survey weights attempt to formally balance bias and variance, resulting in an estimate usually lying between those from the unweighted model and fully weighted model. The weight-smoothing model using a Laplace prior shows the potential to provide a more efficient estimate than either the unweighted model

Table 13. Regression of log TCDD on age and gender, and interaction between age and gender. Bias and RMSE for linear slope estimated for age, gender and interaction: unweighted, fully weighted and hierarchical weight smoothing

Estimator	Age			Gender			Interaction		
	Est. _{.95%CI}	Bias(10^{-4})	RMSE(10^{-3})	Est. _{.95%CI}	Bias(10^{-2})	RMSE(10^{-1})	Est. _{.95%CI}	Bias(10^{-3})	RMSE(10^{-3})
UNWT	.0284 _{.0227,.0341}	-5.063	3.758	-.195 _{-.508,.118}	2.882	1.591	.0096 _{.0031,.0161}	-0.880	3.285
FWT	.0289 _{.0236,.0342}	0	2.661	-.223 _{-.865,.419}	0	3.259	.0105 _{-.0040,.0250}	0	7.335
HWS	.0280 _{.0241,.0319}	-9.142	2.048	-.289 _{-.541,-.037}	-6.530	1.282	.0122 _{.0069,.0174}	1.646	2.667

Table 14. Odds ratio and relevant 95% confidence interval for estimated effect on injury from compacted extended cab pickups: unweighted, fully weighted, hierarchical weight smoothing, and exchangeable random effect

Estimator	Odds Ratio (OR)	
	Unadjusted	Adjusted
UNWT	3.534 _(2.003,6.234)	3.448 _(1.850,6.430)
FWT	11.317 _(2.737,46.784)	13.890 _(3.176,60.760)
HWS	10.559 _(3.731,29.876)	13.268 _(7.919,22.232)

or the fully weighted model, using an approach that is nearly as simple to implement in a regression setting as an exchangeable model, with equivalent or improved increases in efficiency but better robustness properties. Large increases in efficiency occur when bias is present due to model misspecification, and population variance is small so the weight-smoothing model is able to model the underlying data structure precisely, yielding an estimate with greatly reduced mean square error. However, this aggressive estimation comes at some cost of robustness, that is, the reduced variance could lead to lower than nominal coverage rates. As presented in the simulation, the HWS model suffers a moderate drop in the coverage rate when population variance is small, although it is usually competitive with the coverage of the fully weighted estimator. In future, it would be worth exploring the model's mechanism in reducing the overall RMSE, and the limit of the scenarios under which it still maintains reasonable coverage.

The distribution of weights with a high degree of variability can itself vary considerably, from relatively uniform or heavy-tailed distributions, as we have seen in the simulations or the NHANES examples, to a small number of extreme outlying weights, to intermediate cases, as was the case in the Partners for Passenger Safety example. Since the method works by smoothing a given weight-stratum estimate inversely to its stability relative to other weight-stratum estimates, we would anticipate that there would not be consistent differences between these different types of weight distributions. If either the heavy-tailed or outlier weight strata are well estimated and sufficiently different from other weight strata, this interaction will be preserved and the fully weighted estimator will be approximated; otherwise the interaction will be shrunk and the estimator will move away from the fully weighted estimator.

We also note that the Laplace weight-smoothing model is largely agnostic to the construction of the weights. Instead, it focuses on whether there are enough data to support main effects and interactions between the weights and the parameter of interest (note that the main effects themselves can be viewed as interactions with the intercept in the case of estimating a population mean). While simple to implement, there may be settings where one wants to smooth some components of the weights (e.g., selection and nonresponse) while retaining others (e.g., post stratification or calibration). Such "partial smoothing" model-based approaches remain a topic for future research.

Comparing the results of the Laplace prior weight-smoothing models with the model-based estimators of [Beaumont \(2008\)](#), we find that the Laplace estimators offer the promise of relatively simple estimators that can approximate fully weighted estimators

when weights are required for bias correction, but improve over weighted estimators in terms of variability while maintaining an approximately correct nominal coverage of credible intervals. In contrast, in some settings the Beaumont estimators can “over smooth” weights when bias correction is needed and yield unstable estimators when the weight prediction is weak. The predicted weights in the weight-modeling approach of Beaumont incorporate information from design variables, thus yielding better predictions for weighted mean and population total estimates than unweighted estimators. However, in some settings even a degree-five polynomial may fail to correctly approximate the relationship between the inverse of the probability of selection and the sample statistic of interest. Perhaps even more importantly, highly structured models for weight prediction such as high-degree polynomials may result in unstable estimates of weights, adding unnecessary variance rather than dampening it, although model-selection methods may reduce such impacts. Methods such as those proposed by Pfeffermann (2011) and Kim and Skinner (2013), who proposed a form of “stabilized” weight models as $\tilde{w}_i = E_M(w_i|I, Y, X)/E_M(w_i|I, Y)$, may be of use in informative sampling settings (in noninformative sampling settings, $\tilde{w}_i = w_i$ if the model is correctly specified). Ultimately we find attempts to model weights rather than data misguided, as this focuses on design factors on which we should be conditioning, rather than assessing uncertainties in the data that may be fertile ground for mean square error reduction while preserving approximate nominal coverage: that is, calibrated Bayes estimators (Little 2011).

As a final note, there is an issue of whether a census estimate of a parameter of a misspecified model is a sensible inferential target. Our perspective is that statistical models are rarely perfect, and that complex sample designs can sometimes magnify the degree of these failures. We recognize, however, that there is controversy in this area. For example, Rothman et al. (2013) make a case that truly scientific endeavors attempt to make causal statements that should be independent of sample selection. Keiding and Louis (2015) replied to this with an argument perhaps close to the one we make here, which is that the transportability (in the formal sense of Pearl and Bareinboim 2014) of model results may still require attention to the effects of sample design.

6. References

- Alexander, C.H., S. Dahl, and L. Weidman. 1997. “Making Estimates from the American Community Survey.” Paper presented at the 1997 Joint Statistical Meetings, August 10–14, 1997, Anaheim, CA. Available at: https://www.census.gov/content/dam/Census/library/working-papers/1997/acs/1997_Alexander_01.pdf (accessed March 2016).
- Association for the Advancement of Automotive Medicine. 1990. *The Abbreviated Injury Scale*, 1990 Revision. Des Plaines, IL: Association for the Advancement of Automotive Medicine.
- Beaumont, J.P. 2008. “A New Approach to Weighting and Inference in Sample Surveys.” *Biometrika* 95: 539–553. Doi: <http://dx.doi.org/10.1093/biomet/asn028>.
- Binder, D.A. 1983. “On the Variances of Asymptotically Normal Estimators from Complex Surveys.” *International Statistical Review* 51: 279–292. Doi: <http://dx.doi.org/10.2307/1402588>.

- Chen, Q., D.H. Garabrant, E. Hedgeman, R.J.A. Little, M.R. Elliott, B. Gillespie, B. Hong, S.Y. Lee, J.M. Lepkowski, A. Franzblau, P. Adriaens, A.H. Demond, and D.G. Patterson. 2010. "Estimation of Background Serum 2,3,7,8-TCDD Concentrations Using Quantile Regression in the UMDES and NHANES Populations." *Epidemiology* 21: S51–S57. Doi: <http://dx.doi.org/10.1097/EDE.0b013e3181ce9550>.
- Chowdhury, S., M. Khare, and K. Wolter. 2007. "Weight Trimming in the National Immunization Survey." In Proceedings of the Joint Statistical Meetings, Section on Survey Research Methods, American Statistical Association, July 29–August 2, 2007, 2651–2658. Available at: <http://www.amstat.org/sections/SRMS/Proceedings/y2007/Files/JSM2007-000077.pdf> (accessed March 2016).
- Cox, B.G. and D.S. McGrath. 1981. "An Examination of the Effect of Sample Weight Truncation on the Mean Square Error of survey Estimates." Paper presented at the 1981 Biometric Society ENAR meeting, Richmond, VA.
- Elliott, M.R. 2007. "Bayesian Weight Trimming for Generalized Linear Regression Models." *Survey Methodology* 33: 23–34.
- Elliott, M.R. and R.J.A. Little. 2000. "Model-Based Approaches to Weight Trimming." *Journal of Official Statistics* 16: 191–210.
- Eltoft, T., T. Kim, and T.W. Lee. 2006. "On the Multivariate Laplace Distribution." *Signal Processing Letters, IEEE* 13: 300–303. Doi: <http://dx.doi.org/10.1109/LSP.2006.870353>.
- Ericson, W.A. 1969. "Subjective Bayesian Modeling in Sampling Finite Populations." *Journal of the Royal Statistical Society Series B* 31: 195–234. Available at: <http://www.jstor.org/stable/2984206>.
- Ghosh, M. and G. Meeden. 1986. "Empirical Bayes Estimation of Means from Stratified Samples." *Journal of the American Statistical Association* 81: 1058–1062.
- Henry, K. and R.V. Valliant. 2012. "Methods for Adjusting Survey Weights when Estimating a Total." In Proceedings of the 2012 Federal Committee on Statistical Methodology Research Conference. Available at: http://fcsm.sites.usa.gov/files/2014/05/Henry_2012FCSM_V-A.pdf (accessed March 2016).
- Holt, D. and T.M.F. Smith. 1979. "Post Stratification." *Journal of the Royal Statistical Society Series A* 142: 33–46. Doi: <http://dx.doi.org/10.2307/2344652>.
- Horvitz, D.G. and D.J. Thompson. 1952. "A Generalization of Sampling Without Replacement From a Finite Universe." *Journal of the American Statistical Association* 47: 663–685. Doi: <http://dx.doi.org/10.1080/01621459.1952.10483446>.
- Keiding, N. and T.A. Louis. 2015. "Perils and Potentials of Self-Selected Entry to Epidemiological Studies and Surveys." *Journal of the Royal Statistical Society, Series A* 179: 319–376. Doi: <http://dx.doi.org/10.1111/rssa.12136>.
- Kim, J.K. and C.J. Skinner. 2013. "Weighting in Survey Analysis under Informative Sampling." *Biometrika* 100: 385–398. Doi: <http://dx.doi.org/10.1093/biomet/ass085>.
- Kish, L. 1992. "Weighting for Unequal π ." *Journal of Official Statistics* 8: 183–200.
- Lazzeroni, L.C. and R.J.A. Little. 1998. "Random-Effects Models for Smoothing Post-Stratification Weights." *Journal of Official Statistics* 14: 61–78.
- Little, R.J.A. 1991. "Inference with Survey Weights." *Journal of Official Statistics* 7: 405–424.

- Little, R.J.A. 1993. "Poststratification: A Modeler's Perspective." *Journal of the American Statistical Association* 88: 1001–1012. Doi: <http://dx.doi.org/10.1080/01621459.1993.10476368>.
- Little, R.J.A. 2011. "Calibrated Bayes, for Statistics in General, and Missing Data in Particular." *Statistical Science* 26: 162–174. Doi: <http://dx.doi.org/10.1214/10-STS318>.
- Little, R.J.A., S. Lewitzky, S. Heeringa, J. Lepkowski, and R.C. Kessler. 1997. "Assessment of Weighting Methodology for the National Comorbidity Survey." *American Journal of Epidemiology* 146: 439–449.
- McCullagh, P. and J.A. Nelder. 1989. *Generalized Linear Models*, 2nd ed. Boca Raton, FL: CRC Press.
- Park, T. and G. Casella. 2008. "The Bayesian Lasso." *Journal of the American Statistical Association* 103: 681–686. Doi: <http://dx.doi.org/10.1198/016214508000000337>.
- Pearl, J. and E. Bareinboim. 2014. "External Validity: from Do-calculus to Transportability Across Populations." *Statistical Science* 29: 579–595. Doi: <http://dx.doi.org/10.1214/14-STS486>.
- Pfeffermann, D. 1993. "The Role of Sampling Weights when Modeling Survey Data." *International Statistical Review* 61: 317–337. Doi: <http://dx.doi.org/10.2307/1403631>.
- Pfeffermann, D. 2011. "Modelling of Complex Survey Data: Why Model? Why is it a Problem? How Can We Approach It." *Survey Methodology* 37: 115–136.
- Potter, F.A. 1988. "Survey of Procedures to Control Extreme Sampling Weights." In Proceedings of the Section on Survey Research Methods, American Statistical Association. 453–458. Available at: http://www.amstat.org/sections/srms/Proceedings/papers/1988_083.pdf (accessed March 2016).
- Potter, F. 1990. "A Study of Procedures to Identify and Trim Extreme Sample Weights." In Proceedings of the Survey Research Methods Section, American Statistical Association. 225–230. Available at: http://www.amstat.org/sections/srms/Proceedings/papers/1990_034.pdf (accessed March 2016).
- Rothman, K.J., E.E. Gallacher, and E.E. Hatch. 2013. "Why Representativeness Should be Avoided." *International Journal of Epidemiology* 42: 1012–1014. Doi: <http://dx.doi.org/10.1093/ije/dys223>.
- Rubin, D.B. 1987. *Multiple Imputation for Non-Response in Surveys*. New York: Wiley.
- Skinner, C.J., D. Holt, and T.M.F. Smith. 1989. *Analysis of Complex Surveys*. New York: Wiley.
- Winston, F.K., M.K. Kallan, M.R. Elliott, R.A. Menon, and D.R. Durbin. 2002. "Risk of Injury to Child Passengers in Compact Extended Pickup Trucks." *Journal of the American Medical Association* 287: 1147–1152. Doi: <http://dx.doi.org/10.1001/jama.287.9.1147>.

Received June 2014

Revised September 2015

Accepted November 2015