

Robust and Efficient Methods for Bayesian Finite Population Inference

by

Xi Xia

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2015

Doctoral Committee:

Professor Michael R. Elliott, Chair
Professor Richard D. Gonzalez
Professor Timothy D. Johnson
Professor Roderick J. Little
Research Professor Richard L. Valliant

ACKNOWLEDGEMENTS

It is a long journey that finally comes to an end. And it would not have been possible without the kind support and help of other individuals and organizations. I would like to extend my sincere thanks to all of them.

I am highly indebted to my advisor, Prof. Elliott, for his constant supervision, providing necessary information for this dissertation, and his patience that allows me to walk through the whole process.

I would express my gratitude to all my committee members for their kind cooperation and suggestions which help me finish this dissertation.

Also my thanks goes to all people that shared their comments, or helped me out with their abilities.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	ii
LIST OF FIGURES	vi
LIST OF TABLES	vii
LIST OF APPENDICES	xii
ABSTRACT	xiii
CHAPTER	
I. Introduction	1
II. Advancements in “Weight Pooling” Approaches to Reduce Mean Square Error in Weighted Estimators	7
2.1 Introduction	7
2.2 Weight Pooling Method	10
2.2.1 Weight Trimming	10
2.2.2 Weight Prediction	11
2.2.3 Weight Pooling	12
2.3 Simulation Study	18
2.3.1 Association between Probability of Selection and Regression Slope	19
2.3.2 No Association between Probability of Selection and Regression Slope	24
2.4 Application: Estimating Associations between Blood Levels of Dioxin and Age and Gender	26
2.5 Discussion	29
III. Weight Smoothing for Generalized Linear Models Using a Laplace Prior	34

3.1	Introduction	34
3.2	Weight Smoothing Methodology	37
3.2.1	Finite Bayesian Population Inference	37
3.2.2	Weight Prediction	38
3.2.3	Weight Smoothing	38
3.2.4	Weight Smoothing for Linear and Generalized Linear Regression Models	40
3.2.5	Laplace Prior for Weight Smoothing	42
3.3	Simulation Study	45
3.3.1	Hierarchical Weight Smoothing Model for Ordinary Linear Regression	46
3.3.2	Hierarchical Weight Smoothing Model for Logistic Regression	50
3.4	Application	54
3.4.1	Application on Dioxin Data from NHANES	54
3.4.2	Application on Partner for Child Passenger Safety Data	56
3.5	Discussion	58

IV. Weighted Dirichlet Process Mixture Model to Accommodate Complex Sample Designs for Linear and Quantile Regression 60

4.1	Introduction	60
4.2	Background Methodology	64
4.2.1	Bayesian Finite Population Inference	64
4.2.2	Quantile Regression	65
4.2.3	Finite Gaussian Mixture Model	67
4.2.4	Dirichlet Process Mixture Model	67
4.3	Weighted Dirichlet Process Mixture Model	69
4.4	Simulation Study	71
4.4.1	Weighted Dirichlet Mixture Model for Ordinary Linear Regression	71
4.4.2	Weighted Dirichlet Mixture Model for Quantile Regression	75
4.4.3	Weighted Dirichlet Mixture Model for Quantile Regression with Binary Covariate	80
4.5	Application on Dioxin data from NHANES	82
4.5.1	Linear Regression Setting	83
4.5.2	Quantile Regression Setting	86
4.6	Discussion	92

V. Conclusion 94

APPENDICES	97
A.1 Metropolis Algorithm for Advanced Weight Pooling Method .	98
B.1 Full Conditional Distribution for Linear Model	101
B.2 Full Conditional Distribution for Logistic Model	104
C.1 Gibbs Sampler for Linear Regression	107
C.2 Gibbs Sampler for Quantile Regression	110
BIBLIOGRAPHY	113

LIST OF FIGURES

Figure

4.1	Scatter plot of generated population vs covariate. Population are marked grey in background and sample in blue cross. From left to right, $\sigma^2 = 10^{1.5}, 10^{3.5}, 10^{5.5}$; from top to bottom, setting $\beta_a, \beta_b, \beta_c$.	72
4.2	Plot of quantile regression simulation settings. Population as grey in background and sample in blue cross. From left to right: t distribution, exponential distribution and bimodal distribution; From top to bottom: Scatter plot, Histogram of Y at X=0, 5, and 10.	77
4.3	Figure of biases of estimated population slope for median from all three methods. Biasness of point estimates from each simulation are plotted sequentially. Blue circles mark the unweighted quantile regression, black crosses mark the fully weighted quantile regression, and red circles mark WDPM estimates.	82
4.4	Quantile regression of log TCDD on age. Sample presented as background dots, color represents weighting in log form, and lines from top to bottom are linear slope for first quartile, median and third quartile, black for WDPM, blue for fully-weighted	87

LIST OF TABLES

Table

2.1	Bias, RMSE, and coverage of nominal 95% confidence/credible interval for linear slope when true model is linear and probability of selection is associated with slope: unweighted, fully weighted, standard weight pooling estimator (without and with fractional Bayes Factor priors), linear spline weight pooling estimator (without and with fractional Bayes Factor priors), and generalized design-based estimator.	21
2.2	Bias, RMSE, and coverage of nominal 95% confidence/credible interval for linear slope when true model is convex and probability of selection is associated with slope: unweighted, fully weighted, standard weight pooling estimator (without and with fractional Bayes Factor priors), linear spline weight pooling estimator (without and with fractional Bayes Factor priors), and generalized design-based estimator.	22
2.3	Bias, RMSE, and coverage of nominal 95% confidence/credible interval for linear slope when true model is concave and probability of selection is associated with slope: unweighted, fully weighted, standard weight pooling estimator (without and with fractional Bayes Factor priors), linear spline weight pooling estimator (without and with fractional Bayes Factor priors), and generalized design-based estimator.	24
2.4	Bias, RMSE, and coverage of nominal 95% confidence/credible interval for linear slope when true model is convex and probability of selection is not associated with slope: unweighted, fully weighted, standard weight pooling estimator (without and with fractional Bayes Factor priors), linear spline weight pooling estimator (without and with fractional Bayes Factor priors), and generalized design-based estimator.	25

2.5	Bias, RMSE, and coverage of nominal 95% confidence/credible interval for linear slope when true model is concave and probability of selection is not associated with slope: unweighted, fully weighted, standard weight pooling estimator (without and with fractional Bayes Factor priors), linear spline weight pooling estimator (without and with fractional Bayes Factor priors), and generalized design-based estimator.	26
2.6	Bias and RMSE for linear slope estimated for age: unweighted, fully weighted, standard weight pooling estimator, and linear spline weight pooling estimator (without and with fractional Bayes Factor priors).Units present in parenthesis.	28
2.7	Bias and RMSE for linear slope estimated for gender: unweighted, fully weighted, standard weight pooling estimator, and linear spline weight pooling estimator (without and with fractional Bayes Factor priors).Units present in parenthesis.	28
2.8	Bias and RMSE for linear slope estimated for age and gender: unweighted, fully weighted, standard weight pooling estimator, and linear spline weight pooling estimator (without and with fractional Bayes Factor priors).Units present in parenthesis.	29
2.9	Bias and RMSE for linear slope estimated for age, gender and interaction: unweighted, fully weighted, standard weight pooling estimator, and linear spline weight pooling estimator (without and with fractional Bayes Factor priors).Units present in parenthesis.	30
3.1	Comparison of various estimators of slope B_1 under β_a linear spline setting. Bias and RMSE under populations with residual variance $10^{1.5}$, $10^{3.5}$ and $10^{5.5}$ from following model: unweighted, fully weighted, hierarchical weight smoothing, exchangeable random effect and weight prediction by y, degree 5 polynomial of y, linear combination of x and y, and degree 5 polynomial of x,y.	48
3.2	Comparison of various estimators of slope B_1 under β_b linear spline setting. Bias and RMSE under populations with residual variance $10^{1.5}$, $10^{3.5}$ and $10^{5.5}$ from following model: unweighted, fully weighted, hierarchical weight smoothing, exchangeable random effect and weight prediction by y, degree 5 polynomial of y, linear combination of x and y, and degree 5 polynomial of x,y.	48

3.3	Comparison of various estimators of slope B_1 under β_c linear spline setting. Bias and RMSE under populations with residual variance $10^{1.5}$, $10^{3.5}$ and $10^{5.5}$ from following model: unweighted, fully weighted, hierarchical weight smoothing, exchangeable random effect and weight prediction by y, degree 5 polynomial of y, linear combination of x and y, and degree 5 polynomial of x,y.	49
3.4	Comparison under model misspecification. Bias and RMSE under populations with underlying model quadratic coefficient 0, .45 and .80 from following model: unweighted, fully weighted, hierarchical weight smoothing, exchangeable random effect and weight prediction by y, degree 5 polynomial of y, linear combination of x and y, and degree 5 polynomial of x,y.	52
3.5	Comparison under informative sampling. Bias and RMSE under populations with correlation between Z and Y equal to .05, .50 and .95 from following model: unweighted, fully weighted, hierarchical weight smoothing, exchangeable random effect and weight prediction by y, degree 5 polynomial of y, linear combination of x and y, and degree 5 polynomial of x,y.	53
3.6	Regression of log TCDD on Age. Bias and RMSE for linear slope estimated for age: unweighted, fully weighted and hierarchical weight smoothing.	55
3.7	Regression of log TCDD on Gender. Bias and RMSE for linear slope estimated for gender: unweighted, fully weighted and hierarchical weight smoothing.	55
3.8	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.	55
3.9	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.	56
3.10	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.	58

4.1	Comparison of various estimators of slope B_1 under β_a linear spline setting. Bias and relative RMSE under populations with residual variance $10^{1.5}$, $10^{3.5}$ and $10^{5.5}$ from following model: Unweighted, Fully Weighted, Weight Trimming and weighted Dirichlet Process Mixture Model.	74
4.2	Comparison of various estimators of slope B_1 under β_b linear spline setting. Bias and relative RMSE under populations with residual variance $10^{1.5}$, $10^{3.5}$ and $10^{5.5}$ from following model: Unweighted, Fully Weighted, Weight Trimming and weighted Dirichlet Process Mixture Model.	74
4.3	Comparison of various estimators of slope B_1 under β_c linear spline setting. Bias and relative RMSE under populations with residual variance $10^{1.5}$, $10^{3.5}$ and $10^{5.5}$ from following model: Unweighted, Fully Weighted, Weight Trimming and weighted Dirichlet Process Mixture Model.	74
4.4	Comparison across various estimators of slope B_1 under non central T distribution. Bias and relative RMSE of estimates for 1st quartile, median and 3rd quartile of the outcome from following model: Unweighted, Fully Weighted, Weight Trimming and weighted Dirichlet Process Mixture Model.	78
4.5	Comparison across various estimators of slope B_1 under Gamma distribution. Bias and relative RMSE of estimates for 1st quartile, median and 3rd quartile of the outcome from following model: Unweighted, Fully Weighted, Weight Trimming and weighted Dirichlet Process Mixture Model.	78
4.6	Comparison across various estimators of slope B_1 under Bimodal distribution. Bias and relative RMSE of estimates for 1st quartile, median and 3rd quartile of the outcome from following model: Unweighted, Fully Weighted, Weight Trimming and weighted Dirichlet Process Mixture Model.	79
4.7	Comparison across various estimators of slope B_1 under Bimodal distribution with Binary covariates. Bias and RMSE of estimates for 1st quartile, median and 3rd quartile of the outcome from following model: Unweighted, Fully Weighted, Weight trimming by 3 SD and weighted Dirichlet Process Mixture Model.	81
4.8	Regression of log TCDD on Age. Bias, RMSE and 95% CI for linear slope estimated for age in unweighted, fully weighted, weight trimming by 3 SD and weighted Dirichlet Process Mixture models. . . .	84

4.9	Regression of log TCDD on Gender. Bias, RMSE and 95% CI for linear slope estimated for gender in unweighted, fully weighted, weight trimming by 3 SD and weighted Dirichlet Process Mixture models. .	84
4.10	Regression of log TCDD on Age and Gender. Bias, RMSE and 95% CI for linear slope estimated for age and gender in unweighted, fully weighted, weight trimming by 3 SD and weighted Dirichlet Process Mixture models.	85
4.11	Regression of log TCDD on Age, Gender and Interaction. Bias, RMSE and 95% CI for linear slope estimated for age, gender and interaction in unweighted, fully weighted, weight trimming by 3 SD and weighted Dirichlet Process Mixture models.	85
4.12	Quantile regression of log TCDD on age. Bias, RMSE and 95%CI for linear slope estimated for age in unweighted, fully weighted, weight trimming by 3 SD and weighted Dirichlet Process Mixture models. .	87
4.13	Quantile regression of log TCDD on gender. Bias, RMSE and 95% CI for linear slope estimated for gender in: unweighted, fully weighted, weight trimming by 3 SD and weighted Dirichlet Process Mixture model.	88
4.14	Quantile regression of log TCDD on age and gender. Bias, RMSE and 95% CI for linear slope estimated for age and gender in unweighted, fully weighted, weight trimming by 3 SD and weighted Dirichlet Process Mixture models.	89
4.15	Quantile regression of log TCDD on age, gender and interaction. Bias, RMSE and 95% CI for linear slope estimated for age, gender and interaction in unweighted, fully weighted, weight trimming by 3 SD and weighted Dirichlet Process Mixture model.	91

LIST OF APPENDICES

Appendix

A. Metropolis Algorithm For Advanced Weight Pooling Method 98

B. Gibbs Sampler For Weight Smoothing Model 101

C. Gibbs Sampler for Weighted Dirichlet Process Mixture Model with Survey Weights 107

ABSTRACT

Robust and Efficient Methods for Bayesian Finite Population Inference

by

Xi Xia

Chair: Professor Michael R. Elliott

Bayesian model-based approaches provide data-driven estimates of population quantity of interest from complex survey data to achieve balance between bias correction and efficiency. We focus on the issue of accommodating sample weights equal to the inverse of the probabilities of inclusion. In settings with highly variable weights, weight “trimming” is often employed in an ad-hoc manner to decrease variance, while possibly increasing bias. We consider three model-based methods to provide principled bias-variance tradeoffs.

Weighted estimators can be developed in a model-based framework by including interactions between the quantity of interest (e.g., mean, regression parameters) and the weights; weight pooling builds a variable selection model where these interactions are dropped for differing values of the weights; and estimation proceeds using the posterior distribution of model averages. The extension considers a weight pooling linear spline model that uses a linear spline to capture regression coefficient patterns for all strata, and collapses together the strata with minor differences. Our model achieves robustness when weights are needed to guard against model misspecification, and efficiency when weight-coefficient interactions could be ignored. We also model

the interactions between the weights and the estimators of interest as random effects, reducing the overall RMSE by shrinking interactions toward zero when such shrinkage is supported by data (Elliott and Little 2000, Elliott 2007). We adapt a flexible Laplace prior distribution to gain robustness against model misspecification. We find that weight smoothing models with Laplace priors approximate unweighted estimates when weighting is not necessary, and could greatly reduce the RMSE if strong pattern exists in data in linear model setting. Under logistic regression with same sample size, the estimates are still robust, but with less gain in efficiency. Finally, we adapt a Dirichlet process mixture (DPM) model that is capable of approximating highly-skewed and multimodal distributions, often with small number of components. The extended weighted DPM version allows the DP prior to be a mixture of DP random basis measures that is a function of covariates, extends applications to regression, and creates a natural link to survey weights. We also investigate its application to inference for quantile regression, providing a new approach for quantile regression incorporating complex survey design. Simulation results suggest great reduction in RMSE from weighted DPM method under most of the scenarios.

CHAPTER I

Introduction

When analyzing data with complex survey design, biasness in estimating the quantity of interest (population mean, slope, etc.) is often introduced by unequal probabilities of inclusion, and fully-weighted estimators with weights equal to the inverse of probabilities of inclusion are the common countermeasure to correct biasness. Obtained by applying weights in score equations and forming pseudo-maximum likelihood estimators (PMLEs), the fully-weighted estimator is design consistent for MLEs, either as defined in *Cochran* (1977) – the estimator equals the population quantity being estimated when the sample consists of the whole population – or as defined in *Sarndal* (1980) – the estimator converges to the target quantity where samples of increasing size are selected in an identical fashion from infinite replicates of the population. However, this welcome property can come at the cost of increase in the variance of estimator. When quantity of interest is not related to weights, or extreme values of weight exist, the gain in bias-reduction may not compensate the loss due to increasing variance, leading to an overall larger mean square error (MSE).

To reduce the inflated variance in that situation, various methods have been proposed. One general approach, weight trimming, or winsorization, limits the influence from extreme weights by capping the weights at some cut point w_0 , and redistributing the trimmed values equally among the rest (*Alexander et al.*, 1997; *Kish*, 1992;

Potter, 1990). This leads to various extensions in deciding the cut point w_0 . Some examples include the NAEP method by *Potter* (1988), which set the cutoff point equal $\sqrt{c \sum_{i \in s} w_i^2/n}$, where c is chosen in an ad-hoc manner; the empirical MSE method by *Cox and McGrath* (1981) which estimates the cutoff point value by optimizing the empirical MSE estimated from $M\hat{S}E(\hat{\theta}_t) = (\hat{\theta}_t - \hat{\theta}_w)^2 - \hat{V}ar(\hat{\theta}_t) + 2\sqrt{\hat{V}ar(\hat{\theta}_t)\hat{V}ar(\hat{\theta}_w)}$, 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, from 1 as the unweighted estimator to T as the fully-weighted estimator; and the skewed CDF method by *Chowdhury et al.* (2007), which assumes a skewed cumulative distribution (e.g., an exponential distribution) on the weights, and uses the upper 1% of the fitted distribution as a cut point for weight trimming. Details of these approaches are summarized in *Henry and Valliant* (2012) review.

An Alternative approach we focus on here is the Bayesian model-based approach, in particular finite population Bayesian inference that provides data-driven estimates that balance bias and efficiency. The Bayesian approach posits a model, parametric or nonparametric, as a prior distribution for the fixed and unknown finite population that depends on unknown parameters. Inference about the quantity of interest as a function of the population is made by summarizing the posterior predictive distribution of unobserved observations in the population. It is an attractive feature which Bayesian model-based approach does not rely on asymptotic arguments.

To account for survey designs with disproportional selection probabilities under a Bayesian framework, one can create dummy indicators that stratify samples by equal or approximately equal probabilities of selection. And a fully-weighted data analysis is equivalent to estimating a model with interactions between stratum indicators and target model parameters of interest. However, as number of strata increases, the numbers of interaction terms increase accordingly, reduce efficiency of the interaction model, or even make it impossible to estimate all parameters under extreme circum-

stances. Thus, various models are proposed that are simpler than the full interaction model, but still maintain enough flexibility to accommodate different population patterns precisely and yield better inferences. In the following chapters, we study these models and expand their applications.

In chapter II, we extend the “weight pooling” models of *Elliott and Little* (2000); *Elliott* (2007, 2008) that provides analog of the standard weight trimming method, adapted to a more principled Bayesian model framework. Note that a fully-weighted data analysis can be viewed as inference from the posterior predictive distribution of a population quantity under a model in which interaction terms are presented between the weight stratum indicators and the underlying model parameters of interest. Weight pooling drops unnecessary interactions based on different values of weights in a data-driven manner. Inference is based on averaging all possible weight pooling models based on their posterior probabilities of being “correct”. The early version in *Elliott and Little*(2000) merged together only strata with largest weights, assuming the underlying data in these combined strata are exchangeable. Later, by considering the possibility of pooling all contiguous strata, *Elliott* (2008, 2009) introduced higher degree of robustness into the model, and protected against “over-pooling” that occurred in the earlier version.

To further extend the model’s capability in handling complex association between strata and quantity of interest, we construct a linear spline model with potential knots to capture regression coefficient patterns for all strata, and collapse together strata with minor differences. Also, with large number of knots considered, we apply a Metropolis step to move around the potential model space, in contrast to previous approaches that required computing posterior probabilities of all possible pooling models, and greatly reduce the burden in computing. Furthermore, we apply a Fractional Bayes Factor prior (*O’Hagan*, 1995) to boost the model’s performance. We assess our new method in a simulation study that aims at estimating population slope

under various data patterns, competing against unweighted estimate, fully weighted estimate and different weight trimming methods. Bias, RMSE and coverage rate are recorded, and the results suggest that our method maintains consistent better performance regarding RMSE.

In chapter III, we study the “weight smoothing” approach that estimates population quantity of interest through mixed models that consider interactions between the weights and the quantities of interest as random effects, shrinking them towards zero when data provide little evidence of interactions, but keeping strata separated when data suggest interacting (*Elliott, 2007; Elliott and Little, 2000; Ghosh and Meeden, 1986; Lazzeroni and Little, 1998; Little, 1991, 1993; Rizzo, 1992*). *Elliott (2008, 2009)* extended the application to linear and generalized linear models, and discussed four different settings for the random effect priors, namely exchangeable, autoregressive, linear and nonparametric random slopes, and evaluated their performances.

We adapt a more flexible Laplace prior distribution instead of multivariate normal distribution for the hierarchical Bayesian model in order to achieve more robustness against “oversmoothing” in settings where weights are required to accommodate model misspecification or non-ignorable sampling. Given the prevailing performance of Laplace prior in sparse model selection, we expect the hierarchical model to maintain high performance, even under simplistic mean and covariance matrix settings such as exchangeable random priors, therefore reduce the overall complexity of the model. Again, we test the performance of our proposed model in simulation studies, under both model misspecification and informative sampling settings, for both numerical and dichotomous outcomes, and compare it with competing methods. Despite some minor decrease in coverage rates, our method performs consistently better than the competing methods, reducing RMSE by up to 50% in certain settings.

In Chapter IV, we introduce the Dirichlet Process Mixture model (*Dunson et al., 2007*) in complex survey data analysis. Depicting interactions between regression pa-

rameters and probabilities of inclusion by a combination of fixed number of component distributions, the Gaussian mixture model is known for approximating highly-skewed and multimodal distributions, often with fairly small number of components. DPM model relaxes the assumption of pre-determined number of mixture components. The extended Weighted Dirichlet Process Mixture Model (WDPM model) further allows the DP prior to be a mixture of DP random basis measures that is a function of covariates, granting more flexibility to extend applications to mean of quantile regression models.

In this manuscript, we investigate the application of the WDPM model in analyzing complex survey design data with small sample sizes, targeting data-driven inference that captures a wide variety of normal and non-normal distributions, in a fashion that is sensitive to unequal probability of selection aspects of the sample design, but also offers increased efficiency when data permit. In addition to the linear regression setting, we also consider applications to quantile regression. Due to the fact that the WDPM models are highly flexible and can generate predictive distributions that are accurate in tails of the distribution, they are a natural choice to consider for model-based methods to obtain population quantile regression estimates. To evaluate the performance of the WDPM models, we run a series of simulations across various mis-specified models and non-normal distributions, under both linear model and quantile regression settings. The results show considerable improvement compared to fully weighted methods across all settings tested.

Besides simulations, we also test all three models' performances with Dioxin Dataset from NHANES study, focusing on identifying the relationship between log transformed 2,3,7,8 tetrachlorodibenzo-p-dioxin (TCDD), a toxic substance that accumulates in blood, and demographic factors like gender and age. The fully-weighted estimates are assumed to be the true values, and bias, RMSE are calculated accordingly. To evaluate the extended application in generalized linear regression from

chapter III, we apply the method on Partner of Child Passenger Safety Dataset from State farm Insurance database, featuring the status of cars involved in accident, and the injury status of their children passengers. The outcome variable is a binary injury index, and predictors are various car status. We still observe some improvement from our proposed methods, but less clearly compared to well-controlled simulation studies.

CHAPTER II

Advancements in “Weight Pooling” Approaches to Reduce Mean Square Error in Weighted Estimators

2.1 Introduction

When analyzing data for sample designs with unequal probabilities of inclusion, standard design-based approaches typically use “fully-weighted” estimates of population means, totals, regression slopes, etc., where the weights are equal to the inverse of the probabilities of inclusion (*Horvitz and Thompson, 1952*). The fully-weighted estimators of model parameters obtained by using sampling weights in score equations are sometimes termed “pseudo-maximum likelihood” estimators (PMLEs) (*Binder, 1983; Pfeffermann, 1993*) because they are design consistent for MLEs that would solve the score equations under the data model $f(Y | \theta)$ if we had observed data for the entire population. While design consistency, either in the sense of Cochran (1977) – the estimator equals the population quantity being estimated when the sample consists of the whole population – or in the asymptotic sense of Sarndal (1980), where asymptotics are formed from samples selected in an identical fashion from $t \rightarrow \infty$ replicates of the population, is an attractive property, bias reduction typically comes at the cost of increased variance. This increase can overwhelm the reduction in bias,

so that the mean square error (MSE) actually increases under weighted analysis.

An alternative to standard design-based procedures is to use model-based procedures, in particular finite population Bayesian inference. Design-based inference in survey data analysis treats the survey outcome variables $Y = \{Y_1, Y_2, \dots, Y_N\}$ for N subjects in the population as fixed unknown constants; the random process is the sample design which identifies the n sample subjects in the population. An estimate of the population quantity $Q = Q(Y)$ is constructed based on the sample $q = q(Y_{obs})$, a function of the sample $Y_{obs} = \{y_1, y_2, \dots, y_n\}$. In contrast, Bayesian approaches posit a model for the population data Y as a function of parameters θ : $Y \sim f(Y | \theta, Z)$, where Z designates the design variables. This is treated as prior distribution for the fixed and unknown finite population Y depending on some unknown parameter θ . This prior distribution can be parametric or nonparametric. Inference about $Q(Y)$ is made on the posterior predictive distribution of $p(Y_{nob} | Y_{obs}, I, Z)$, where Y_{nob} consists of the $N - n$ unobserved quantities in the population Y :

$$p(Y_{nob} | Y_{obs}, I, Z) = \frac{\int \int p(Y_{nob} | Y_{obs}, Z, \theta, \phi) p(I | Y, Z, \theta, \phi) p(Y_{obs} | Z, \theta) p(\theta, \phi) d\theta d\phi}{\int \int \int p(Y_{nob} | Y_{obs}, Z, \theta, \phi) p(I | Y, Z, \theta, \phi) p(Y_{obs} | Z, \theta) p(\theta, \phi) d\theta d\phi dY_{nob}} \quad (2.1)$$

where ϕ models the inclusion indicator I . If we assume that ϕ and θ are priori independent and if the distribution of I is independent of $Y|Z$, the sampling design is said to be “unconfounded” or “noninformative”; if the distribution of I depends only on $Y_{obs}|Z$, and $p(\theta, \phi) = p(\theta)p(\phi)$, then the sampling mechanism is said to be “ignorable” (*Rubin*, 1987). Under ignorable sampling designs, $p(I | Y, Z, \theta, \phi) = p(I | Y_{obs}, Z, \phi)$, and thus (2.1) reduces to

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

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). Under the Bayesian approach, the posterior distribution – the conditional distribution of parameters of interest given data – is used for all purposes. This is an attractive feature of a Bayesian method in that it offers conditional inference given data and does not rely on asymptotic arguments.

To accommodate disproportional probability-of-selection designs in a Bayesian framework, case weights can be transformed into dummy variables that stratify by equal or approximately equal probabilities of inclusion, ordered by the inverse of the probability of inclusion (*Holt and Smith*, 1979). Then, a fully-weighted data analysis estimates the posterior predictive distribution of a population quantity under a model in which interaction terms are present between weight stratum indicators and underlying model parameters of interest. Elliott and Little (2000) and Elliott (2007, 2008, 2009) developed model-based estimators for weight trimming using two broad approaches: Bayesian hierarchical modeling, or “weight smoothing,” and Bayesian variable selection modeling, or “weight pooling.” “Weight smoothing” models treat the underlying weight strata as random effects, and induce weight trimming by smoothing strata for which the data provide little evidence of difference, and separating strata that the data suggest should be separated (*Elliott*, 2007; *Elliott and Little*, 2000; *Ghosh and Meeden*, 1986; *Lazzeroni and Little*, 1998; *Little*, 1991, 1993; *Rizzo*, 1992). “Weight pooling” models collapse together inclusion strata. Collapsing only the largest valued strata mimics weight trimming by assuming the underlying data from these combined strata are exchangeable. By averaging over all possible of these “weight pooling” models, we can compute an estimator of the population parameter of interest whose bias-variance tradeoff is data-driven. By allowing for all contiguous inclusion strata to be considered for pooling, Elliott (2008, 2009) induced a high degree of robustness into this model, protecting against “over-pooling” that simpler models suffered from Elliott and Little (2000).

Here we extend the weight-pooling method from Elliott (2008, 2009) in two ways. First, we borrow from the weight-smoothing approach, constructing a linear spline model with many potential knots to grant more flexibility to handle non-linear associations. Second, in contrast to previous approaches that require computing directly the posterior probabilities of all possible models, we use a Metropolis step to move around the potential model space, allowing a much larger number of weight strata combinations to be considered. We test the properties of our proposed approach through simulation, comparing it with competing methods recently proposed for data-driven weight trimming. We also apply the method to determine the relationship between dixon blood level and age and gender using data from the 2003-2004 National Health and Nutrition Examination Survey (NHANES).

This Chapter is organized as follows. In Section 2 we review traditional weight trimming methods as well as recently proposed model-assisted and model-based methods, and develop our proposed linear spline model. Section 3 presents the results of the simulation studies, comparing bias, coverage, and MSE of the proposed method and the competing methods. In Section 4 we evaluate the method through application to the NHANES survey. In section 5 we discuss possible directions for further study.

2.2 Weight Pooling Method

2.2.1 Weight Trimming

When weights are overly variable, they are commonly trimmed or “winsorized” so that weights larger than some value w_0 are fixed as w_0 , with the values above w_0 distributed among the rest (*Alexander et al.*, 1997; *Kish*, 1992; *Potter*, 1990). Some approaches have been developed that focus on determining the cap value w_0 based on

data. These include the NAEP method by Potter (1988), which determines the cutoff point as $\sqrt{c \sum_{i \in s} w_i^2 / n}$, where c is empirically chosen according to $nw_i^2 / \sum_{i \in s} w_i^2$. Cox and McGrath (1981) proposed an empirical MSE approach that relies on optimizing the empirical MSE of $M\hat{S}E(\bar{y}_t) = (\bar{y}_t - \bar{y}_w)^2 + \hat{V}ar(\bar{y}_t) + 2\sqrt{\hat{V}ar(\bar{y}_t)\hat{V}ar(\bar{y}_w)}$, $t = 1 \dots T$, where t denotes various trimming level, from 1 as unweighted estimator to T as fully-weighted estimator. Chowdbury et al. (2007) approached the problem by assuming the weights follow a skewed cumulative distribution, such as exponential distribution, and determining the cut point by the upper 1% of the fitted distribution. For more details of these approaches, see review of design-based weight trimming methods in Henry (2012).

2.2.2 Weight Prediction

Beaumont (2008) proposed a design-based, model-assisted method, considering a prediction model of weights using a polynomial form of a response variable and design variables. Assuming a prediction model for weights, and allowing response variable and design variables in the model, the predicted weights from the fitted model tamp down extreme values. To be more specific, denote $I = (I_1, \dots, I_N)^T$ as the vector of sample inclusion indicators, i.e. $I_i = 1$ as i th unit sampled and $I_i = 0$ otherwise, $Y = (Y_1, \dots, Y_N)^T$ the vector of survey response variable, and $Z = (Z_1, \dots, Z_N)^T$ the vector of design variables. Also it is assumed that the probability of sampling is noninformative, thus $P(I|Z, Y) = P(I|Z)$. The smoothed weights are obtained from $\tilde{w}_i = E_M(w_i|I_i, z_i, y_i)$, where $E_M(\cdot)$ denotes expectation taken with respect to the model for the weight. Beaumont also suggested dropping dependence on design variables, so that $\tilde{w}_i = E_M(w_i|I_i, y_i)$. Two estimators are proposed: a linear model, $E_M(w_i|I, Y) = H_i^T \beta + v_i^{1/2} \epsilon_i$, and an exponential model, $E_M(w_i|I, Y) = 1 + \exp(H_i^T \beta + v_i^{1/2} \epsilon_i)$, where H_i and $v_i > 0$ are known functions of y_i . The latter model prevents the predicted weights from being negative. Two examples

of H functions are given as well, respectively linear combination of y_i and a degree five polynomial of y_i . Once the predicted weights are obtained by fitting the model to the sampled data, the re-weighted estimator of interest is created.

By including the design variables in the model, Beaumont’s suggested method incorporates more information and yields improved prediction of weights for the weighted mean and population estimators. Yet the method does not actually focus on assessing the degree of uncertainty associated with the relationship between the probability of selection and the sample statistic of interest. Large degrees of uncertainty suggest that maintaining such relationships, at least in an unattenuated form, may add variance in excess of any bias correction. Hence a stronger link between weights and the covariate effects could be considered, and a more efficient form of prediction model obtained, which is the approach we pursue in this manuscript.

2.2.3 Weight Pooling

The traditional weight trimming method effectively reduces variance by constraining weights to certain capped value w_0 (often chosen to be 3 or 6 times than mean weight $\bar{w} = n^{-1} \sum_i w_i$), and redistributing the extra weights among uncapped ones by multiplying them by a normalizing constant $\gamma = (N_+ - \sum \kappa_i w_o) / \sum (1 - \kappa_i) w_i$, where κ_i is an indicator variable for whether or not $w_i \geq w_0$. To put this in a modeling context, consider a disproportionally stratified design, and focus on the population mean as the value of interest. The weighted estimator is then defined as $\bar{y}_w = \frac{\sum_h w_h \sum_i^{n_h} y_{hi}}{\sum_h n_h w_h} = \sum_h (N_h/N) \bar{y}_h$, where $w_h = N_h/n_h$ is the inverse of the probability of selection in stratum h , and n_h and N_h correspond to the sample and population size in stratum h respectively. It is straightforward to show that \bar{y}_w is the posterior mean of Gaussian model that assumes different means for each stratum and a constant

variance, with non-informative priors (Elliott and Little 2000):

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

$$\mu_h, \mu_l, \log \sigma \propto \text{const.}$$

The trimmed weight mean estimator is given by

$$\begin{aligned} \bar{y}_{wt} &= \sum_{h=1}^{l-1} \gamma \frac{N_h}{N_+} \bar{y}_h + \sum_{h=l}^H \frac{w_0 n_h}{N_+} \bar{y}_h = \\ &\gamma \sum_{h=1}^{l-1} \frac{N_h}{N_+} \bar{y}_h + \frac{w_0 \sum_{h=l}^H n_h}{N_+} \bar{y}^{(l)} \end{aligned}$$

where $\gamma = \frac{N_+ - w_0 \sum_{h=l}^H n_h}{\sum_{h=1}^{l-1} N_h}$ and $\bar{y}^{(l)} = (1 / \sum_{h=l}^H n_h) \sum_{h=l}^H n_h \bar{y}_h$. Choosing $w_0 = \frac{\sum_{h=l}^H N_h}{\sum_{h=l}^H n_h}$ yields $\gamma = 1$ and $\bar{y}_{wt} = \sum_{h=1}^{l-1} \frac{N_h}{N_+} \bar{y}_h + \frac{(\sum_{h=l}^H N_h)}{N_+} \bar{y}^{(l)}$; the resulting trimmed weight mean corresponds to the estimate for a model that assumes distinct stratum means for the smaller weight strata and a common mean for the larger weight strata above a cut point l (Elliott and Little 2000):

$$Y_{hi}|\mu_h \sim N(\mu_h, \sigma^2) \quad h < l$$

$$Y_{hi}|\mu_l \sim N(\mu_l, \sigma^2) \quad h \geq l$$

$$\mu_h, \mu_l, \log \sigma \propto \text{const}$$

Elliott and Little (2000) proposed a model-based approach, where the cut point l was no longer a known constant, but a data-driven parameter. The hierarchical model is

written as following:

$$Y_{hi}|\mu_h \sim N(\mu_h, \sigma^2) \quad h < l$$

$$Y_{hi}|\mu_l \sim N(\mu_l, \sigma^2) \quad h \geq l$$

$$P(L = l) = 1/H$$

$$P(\sigma^2|L = l) = \sigma^{-(l+1/2)}$$

$$P(\beta|\sigma^2, L = l) = 2\pi^{-l}$$

where $\mu_1 = \beta_0, \mu_l = \beta_0 + \beta_{l-1}$, and L indicates the selected pooling model. Via Bayesian model averaging, a posterior distribution of L is determined, and the estimated mean is obtained by summarizing estimations from all models. By averaging based on the probability of certain potential pooling model been the true model, the method avoids the possibility of relying on one single misspecified model, and gains robustness.

Elliott (2008, 2009) extended the method to a linear and generalized model regression setting of Y_i on fixed covariates X_i . By allowing interactions between weights and the population slope, the model could mimic either a fully-weighted estimator or an unweighted estimator, depend on whether the full interaction model or minimal model (all strata share the same slope) been selected. Any model between the two represents a degree of trimming determined by data. And by allowing any adjacent strata, rather than strata on the high end only, to collapse, it introduces more flexibility in modeling, thus potentially increasing the robustness. The model is given

by:

$$Y_{hi}|x_{hi}, \beta_l, \sigma^2, L = l \sim N(Z_{li}\beta_l, \sigma^2)$$

$$\beta_l|\sigma^2, L = l \sim N(\beta_0, \sigma^2\Sigma_0)$$

$$\sigma^2|L = l \sim Inv - \chi^2(a, s^2)$$

$$P(L = l) = 2^{-(H-1)}$$

where $Z_{li} = D_{hl} \otimes x_{hi}$ and D_{hl} is a vector of dummy variables indicates the l th pooling pattern.

2.2.3.1 Weight Pooling Linear Spline Model

The weight pooling method in Elliott (2008) identifies data patterns by either estimating separately the population statistic of interest in each weight stratum, or pooling together the strata to estimate the population statistic. Yet if the statistic follows an approximate linear pattern across these strata, the model may either fail to identify the pattern, or spend unnecessary degrees of freedom estimating it. To obtain further flexibility to balance between simple and complex underlying data structures, we propose a linear spline version of weight pooling, assuming the regression parameters themselves follows a linear spline trajectory. This proposed method replaces the changepoint model above, which treats all regression parameters as being equal within pooled strata, with a linear spline model that assumes that the regression parameters themselves follow a linear spline trajectory, with knots $\{\tau_k\}$ under the l th knot pattern (Zheng and Little, 2003, 2004, 2005):

$$E(y_i | \mathbf{x}_i, \boldsymbol{\beta}^l, L = l) = \sum_{j=0}^p \left[\beta_{j0}^l + \sum_{k=1}^{H^*} \beta_{jk}^l (h - \tau_k)_+ \right] x_{ji} \quad (2.2)$$

where $x_{0i} \equiv 1$ for all i (intercept term), $(x)_+ = x$ if $x > 0$, 0 otherwise, and H^* is the number of knots in the l th knot pattern. A generalized linear model (GLM) can

be obtained by replacing with $g(E(y_i | \mathbf{x}_i, \boldsymbol{\beta}^l, L = l))$, where $g(\cdot)$ is the GLM link function. Thus the unweighted analysis is obtained if $L = 1$ so that $\tau_1 = H$:

$$\beta_{j_0}^l + \sum_{k=1}^{H^*} \beta_{j_k}^l (h - \tau_k)_+ = \beta_{j_0}^1 + \beta_{j_1}^1 (h - H)_+ = \beta_{j_0}^1, \quad i \in h = 1, \dots, H$$

Conversely, the fully weighted analysis is obtained when $L = 2^{H-1}$ so that $\tau_1 = 1, \dots, \tau_{H-1} = H - 1$ and $H^* = H - 1$:

$$\beta_{j_0}^l + \sum_{k=1}^{H-1} \beta_{j_k}^l (h - \tau_k)_+ = \begin{cases} \beta_{j_0}^{2^{H-1}}, & i \in h = 1 \\ \beta_{j_0}^{2^{H-1}} + \beta_{j_1}^{2^{H-1}}, & i \in h = 2 \\ \vdots \\ \beta_{j_0}^{2^{H-1}} + (H - 1)\beta_{j_1}^{2^{H-1}} + \dots + \beta_{j_{H-1}}^{2^{H-1}}, & i \in h = H \end{cases}$$

as all interactions between the regression parameters and the weight strata are included. Intermediate pooling models provide mean structure under a flexible linear model, so that, if $L = l_1(H - 2) + l_2$ (changepoints at $l_1, l_2, l_1 < l_2$), then $H^* = 2$ and:

$$\beta_{j_0}^l + \sum_{k=1}^2 \beta_{j_k}^l (h - \tau_k)_+ = \begin{cases} \beta_{j_0}^{l^*}, & i \in h = 1, \dots, l_1 - 1 \\ \beta_{j_0}^{l^*} + \beta_{j_1}^{l^*} (h - l_1) = \\ (\beta_{j_0}^{l^*} - \beta_{j_1}^{l^*} l_1) + \beta_{j_1}^{l^*} h, & i \in h = l_1, \dots, l_2 - 1 \\ \beta_{j_0}^{l^*} + \beta_{j_1}^{l^*} (h - l_1) + \beta_{j_2}^{l^*} (h - l_2) = \\ (\beta_{j_0}^{l^*} - \beta_{j_1}^{l^*} l_1 - \beta_{j_2}^{l^*} l_2) + (\beta_{j_1}^{l^*} + \beta_{j_2}^{l^*}) h, & i \in h = l_2, \dots, H \end{cases}$$

for $l^* = l_1(H - 2) + l_2$.

We anticipate that this model should reduce the number of changepoints needed to pick up local non-linearities in the interactions between regression slopes and weights, while providing a more rapid tradeoff toward variance reduction when bias correction is unimportant but still retaining robustness. Note that the set of knots $\{\tau_1, \dots, \tau_{H^*}\}$ maps 1-1 with the model index $L = l$, so the prior on the knots corresponds to $p(L)$. Also note that, if we denote $Z_i = x_i(1, (h - \tau_1)_+, \dots, (h - \tau_H)_+)'$, and $\beta_l = (\beta_0, \dots, \beta_H)$, the expected transformed mean of $Y_{hi} | x_{hi}, \beta_l, \sigma^2, L = l$ can be written as $Z_i \beta$, which

resembles the format in Elliott (2008, 2009) for computation.

Assuming a generalized linear model with link function $g(\mu_i)$ and variance function $V(\mu_i)$ for $E(Y_i) = \mu_i$, our population quantity of interest $\mathbf{B} = (B_1, \dots, B_p)^T$ is the slope that solves the population score equation:

$$U_N(\mathbf{B}) = \sum_{i=1}^N \frac{(y_i - g^{-1}(\mu_i(\mathbf{B})))\mathbf{x}_i}{V(\mu_i(\mathbf{B}))g'(\mu_i(\mathbf{B}))}.$$

The posterior predictive distribution of \mathbf{B} is then given by

$$p(\mathbf{B} | y) = \sum_l \int p(\mathbf{B} | y, \beta_l, \phi, L = l) p(\beta_l, \phi | y, L = l) P(L = l | y) d\beta_l d\phi$$

Simulations from $p(\mathbf{B} | y, X)$ can be obtained by drawing first from $P(L = l | y)$, then from $p(\beta_l, \phi | y, L = l)$, and solving

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

where $W_h = N_h/n_h$ for the population size N_h , sample size n_h in the h th inclusion stratum, and $\mathbf{Z}_{li} = D_{hl} \otimes \mathbf{x}_{hi}$ where D_{hl} is a vector of dummy variables that pool the appropriate conterminous inclusion strata based on the l th pooling pattern. Since the computation of the kernel of $P(L|y, X)$ is applicable under conjugate prior, if H is of moderate size, the factor between kernel and actual distribution could be achieved by summing up all kernels, and a direct draw from posterior probability $P(L|y, X)$ is possible. Alternatively, the ratio of $P(L|y, X)$ from two pooling patterns is always accessible since their distributions share the same constant factor. This leads to a Metropolis step to approximate the marginal posterior distribution $P(L = l|y, X)$, and direct draws of other parameters accordingly. The latter approach is computationally plausible under large H . We provide details of the Markov Chain Monte Carlo algorithm for the Gaussian linear regression setting in Appendix A.

2.2.3.2 Fractional Bayes Factor

During the Metropolis step of drawing $P(L|y, X)$, we must compute the Bayes Factors(BF), comparing weight pooling model l with model l' :

$$BF(y, X) = \frac{p(L = l|y, X)}{p(L = l'|y, X)} = \frac{p(y|L = l, X)p(L = l)}{p(y|L = l', X)p(L = l')} = \frac{\int f(y|\theta_l)p(\theta_l)d\theta_l p(L = l)}{\int f(y|\theta_{l'})p(\theta_{l'})d\theta_{l'} p(L = l')}.$$

Under weakly informative priors, the BF is usually quite sensitive to the choice of $p(\theta_l)$ (*Kass and Raftery, 1995*). To counter this, we use the fractional Bayes Factor approach developed by O'Hagan (1995). The approach sets aside a fraction b of data for a data-based prior for θ_l , and the relevant fractional Bayes Factor(FBF) is defined as $BF_b(y, X) = q_l(f, y, X)P(L = l)/q_{l'}(f, y, X)P(L = l')$, where

$$q_l(f, y, X) = \frac{\int p(\theta_l)f(y|\theta_l)d\theta_l}{\int p(\theta_l)f(y|\theta_l)^b d\theta_l}.$$

O'Hagan suggested $n^{-1} \log n$ and $n^{-1/2}$ as increasingly “robust” choices of b . The former one is smaller, which would make the model sensitive to data structure, while the latter with larger value leads to more robust model selection against possible outliers (data generated under a model not in the classes considered).

2.3 Simulation Study

We explore the linear spline weight pooling model in linear regression setting via a set of simulation studies. We consider two basic settings: one in which the probability of selection is associated with the linear approximation between the outcome Y and the predictor of interest X , and one in which the probability of selection is independent of this linear approximation.

2.3.1 Association between Probability of Selection and Regression Slope

Here we generate the population under a model similar to Elliott (2008), but extended to 20 strata:

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

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

The sample design is a disproportionally stratified model, with 20 strata where the probability of selection within each stratum is given by:

$$P(I_i|H_i) = \pi_i \propto (1 + H_i/15)H_i$$

$$H_i = \lceil 2X_i \rceil / 2.$$

Note that the probability of selection is related to regression covariate.

The probabilities of inclusion are constructed that the ratio between the maximum and minimum of weights is around 35, and the sizes of potential strata are greater than 3 to be estimable. A total number of 1000 elements are sampled without replacement for each simulation. Our target estimator is the best linear approximation of the linear slope B_2 relating Y to X , given by

$$\begin{pmatrix} B_1 \\ B_2 \end{pmatrix} = \left(\sum_{i=1}^N X_i X_i^T \right)^{-1} \sum_{i=1}^N X_i y_i$$

where $X_i = (1 \ x_i)$, The number of possible weight pooling models is given by $2^{19} = 524,288$, far too many to be averaged over directly as in Elliott (2008).

For simulation settings we consider three β patterns:

$$1. \beta_a = (0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)'$$

$$2. \beta_b = (0, 0, 0, 0, 0, 0, .5, .5, .5, .5, 1, 1, 1, 1, 2, 2, 2, 2, 4, 4, 4)'$$

$$3. \beta_c = (0, 22, -4, -4, -2, -2, -2, -2, -1, -1, -1, -1, -.5, -.5, -.5, -.5, 0, 0, 0, 0, 0)'$$

which represent respectively a pure linear model, an increasing curve, and a flattening curve. Under β_a where ordinary linear model is correctly specified, the weights are not needed for bias correction, so the unweighted estimator should be most efficient. Under β_b , the unweighted estimator is biased, but the most non-linear part of the model space is the most heavily sampled, somewhat dampening the variance inflation for fully-weighted estimator. Under β_c , correction of bias is also necessary, but the most non-linear part of the model space is the least sampled, inflating the variance of the fully-weighted estimator. Population variance σ^2 varied as 10^l , $l = 1.5, 3.5, 5.5$, emphasizing different degrees of bias-variance tradeoff. 200 simulations are generated under each β and σ^2 combination.

We apply data-based priors for the regression parameters, centering them at the unweighted value with a very inflated variance: $\beta_0 = \hat{\beta} = (X^T X)^{-1} X^T y$, $\Sigma_0 = cn \text{Var}(\hat{\beta})$ for $\text{Var}(\hat{\beta}) = \hat{\tau}^2 (X^T X)^{-1}$, $\hat{\tau}^2 = (n - p)^{-1} (y - X \hat{\beta})^T (y - X \hat{\beta})$, where $c = 1000$. For the the variance prior, we assume $a = s = 10^{-8}$. We also consider Fractional Bayes Factor priors with training fraction of $\log n/n$ and $n^{-1/2}$.

To evaluate the estimation of population slope, we consider a comparison among an unweighted model (UNWT), a fully weighted model (FWT), the Elliott (2008) weight pooling method (PWT) and the Fractional Bayesian Factor versions (PWTFBF1 and PWTFBF2 corresponding to $b = n^{-1} \log n$ and $b = n^{-1/2}$ respectively), the linear spline version of weight pooling developed in this manuscript (PWTLs), and their respective Fractional Bayesian Factor versions. We also consider the general-

Estimator	Bias			RMSE relative to FWT			95% Coverage		
	Model $\sigma^2(\log_{10})$			Model $\sigma^2(\log_{10})$			Model $\sigma^2(\log_{10})$		
	1.5	3.5	5.5	1.5	3.5	5.5	1.5	3.5	5.5
UNWT	-0.027	0.081	-0.201	0.753	0.701	0.694	0.93	0.96	0.96
FWT	-0.018	0.137	0.254	1	1	1	0.96	0.93	0.97
PWT	-0.018	0.135	0.243	1.037	1.012	1.004	0.98	0.95	0.98
PWTFBF1	-0.015	0.129	-0.037	0.853	0.811	0.810	1	0.98	0.98
PWTFBF2	-0.015	0.128	0.035	0.946	0.912	0.911	0.99	1	0.99
PWTLS	-0.013	0.163	0.031	0.914	0.942	0.917	0.99	0.98	0.98
PWTLSFBF1	-0.016	0.131	-0.372	0.815	0.770	0.793	1	1	1
PWTLSFBF2	-0.014	0.172	-0.165	0.815	0.817	0.849	1	1	1
WT1Y	-0.011	0.131	-0.248	0.753	0.708	0.706	0.96	0.98	0.94
WT5Y	0.092	0.391	-0.256	1.374	0.833	0.707	0.71	0.89	0.94
WT5YX	0.071	0.165	-0.153	1.196	0.810	0.795	0.74	0.91	0.92
WT5Y5X	-0.018	0.136	-0.241	0.935	0.911	0.928	0.96	0.96	0.98

Table 2.1: Bias, RMSE, and coverage of nominal 95% confidence/credible interval for linear slope when true model is linear and probability of selection is associated with slope: unweighted, fully weighted, standard weight pooling estimator (without and with fractional Bayes Factor priors), linear spline weight pooling estimator (without and with fractional Bayes Factor priors), and generalized design-based estimator.

ized design-based estimator from Beaumont (2008) as a competing method, where the choice of function H varies from y (WT1Y), degree five polynomial of y (WT5Y), degree 5 polynomial of y plus x (WT5YX), to degree 5 polynomial of y and x separately (WT5Y5X). Under each setting, the estimators are compared with respect to bias, RMSE, and nominal 95% coverage rate, where RMSE is presented as relative to the fully weighted estimator. Results are provided in Tables 2.1, 2.2, and 2.3.

Under β_a , the linear model is correctly specified, all estimators are unbiased, and as expected the unweighted estimator has the best performance, maintaining approximately 70% RMSEs comparing to fully-weighted estimator at all values of σ^2 considered, together with approximately correct coverage rates. The standard weight pooling estimator has results that closely parallel to the fully-weighted estimator; it is substantially improved by applying FBF, reducing RMSEs relative to the fully-weighted one by approximately 20% for FBF1 and 10% for FBF2 while maintaining

Estimator	Bias			RMSE relative to FWT			95% Coverage		
	Model $\sigma^2(\log_{10})$			Model $\sigma^2(\log_{10})$			Model $\sigma^2(\log_{10})$		
	1.5	3.5	5.5	1.5	3.5	5.5	1.5	3.5	5.5
UNWT	2.514	2.483	2.216	10.474	2.785	0.748	0	0	0.91
FWT	-0.011	0.151	0.243	1	1	1	0.98	0.95	0.98
PWT	-0.010	0.152	0.304	0.391	0.972	1.002	1	0.94	0.99
PWTFBF1	0.005	0.284	0.875	0.385	0.934	0.871	0.94	0.94	0.95
PWTFBF2	-0.002	0.173	0.319	0.387	0.935	0.934	0.97	0.96	0.99
PWTLS	-0.010	0.144	0.544	0.401	0.991	0.962	1	0.96	0.97
PWTFBF1	-0.002	0.231	0.483	0.362	0.887	0.866	0.95	0.98	0.98
PWTFBF2	-0.009	0.238	0.552	0.421	0.893	0.857	0.93	1	1
WT1Y	1.489	2.137	2.094	5.907	2.614	0.734	0	0	0.89
WT5Y	1.453	2.323	2.105	5.923	2.598	0.742	0	0	0.89
WT5YX	0.497	0.681	0.536	2.495	1.185	0.803	0.07	0.54	0.91
WT5Y5X	0.204	0.405	0.037	1.282	0.980	0.945	0.99	0.95	0.98

Table 2.2: Bias, RMSE, and coverage of nominal 95% confidence/credible interval for linear slope when true model is convex and probability of selection is associated with slope: unweighted, fully weighted, standard weight pooling estimator (without and with fractional Bayes Factor priors), linear spline weight pooling estimator (without and with fractional Bayes Factor priors), and generalized design-based estimator.

approximately correct or somewhat conservative coverage rates. The linear spline model of weight pooling, and its FBF versions, maintain a small but consistent improvement, about 2% to 5%, over original weight pooling methods. Among different models of generalized design-based estimators, the simple linear model WT1Y essentially matches the unweighted estimator, while the more complex estimators are unstable even relative to the fully weighted estimator and have poor coverage when $\sigma^2 = 10^{1.5}$; for larger values of σ^2 the estimators generally perform similar to the weight pooling estimators with respect to RMSE, with approximately nominal coverage, except for the most complex generalized design-based estimator WT5Y5X, which has only slight RMSE improvements over the fully-weighted estimator.

For scenario β_b , the RMSEs of unweighted estimator are greatly inflated due to bias. The exchangeable weight pooling estimator has bias similar to the fully weighted estimator except for the FBF1 version in the large variance setting; however the

reduced variability yields dramatic reductions of over 60% in RMSE in the small variance setting, and up to 15% in the median and large variance setting, while retaining approximately nominal coverage. The linear spline model method and its FBF versions act similar to original weight pooling method when σ^2 is small, but prevail by up to 10% relative to the exchangeable weight pooling estimator in RMSE when σ^2 increases, indicating that, through proper model selection, linear spline model is more helpful under scenarios in which the data pattern has large variation. The coverage of linear spline model estimator is generally nominal to conservative, with the exception of the FPF2 version in the small variance setting, where the data-based prior may be underestimating variance. The generalized design-based estimators perform well in the large variance setting but poor with respect to both RMSE and coverage in the median and small-variance settings, where the indirect weight adjustment is insufficient to overcome bias. The exception is the most complex WT5Y5X weight model; however, it has little gain in terms of RMSE over the fully-weighted estimator.

The RMSEs of unweighted estimator are greatly inflated due to bias in the β_c scenario, although to a somewhat lesser degree due to the increased variability of the fully-weighted estimators than in the β_b scenario. The results for the weight pooling estimators are similar in terms of bias and RMSE to the β_b scenario, with dramatic reductions in RMSE in the low variance setting and smaller reductions in the medium and large variance settings, and with the linear spline model offering modest improvements over the exchangeable estimator. Coverages of 95% credible intervals for the weight pooling estimators are approximately nominal in the medium and large variance settings, dropping to around 80% in the small variance setting. The generalized design-based estimators again perform rather poorly with respect to both RMSE and coverage in the median and small variance settings, although their biases are less pronounced than in the β_b setting; the most complex WT5Y5X weight model again has similar performance to the fully-weighted estimator.

Estimator	Bias			RMSE relative to FWT			95% Coverage		
	Model $\sigma^2(\log_{10})$			Model $\sigma^2(\log_{10})$			Model $\sigma^2(\log_{10})$		
	1.5	3.5	5.5	1.5	3.5	5.5	1.5	3.5	5.5
UNWT	-1.897	-1.771	-2.058	6.417	2.059	0.752	0	0	0.85
FWT	-0.030	0.117	0.249	1	1	1	1	1	1
PWT	-0.030	0.103	0.227	0.354	0.971	1.001	0.84	0.94	0.99
PWTFBF1	-0.034	-0.039	-0.993	0.355	0.934	0.852	0.84	0.96	0.98
PWTFBF2	-0.033	0.023	-0.287	0.353	0.927	0.925	0.84	0.96	0.97
PWTLS	-0.029	0.104	-0.249	0.352	0.987	0.946	0.84	0.95	0.98
PWTFBF1	-0.032	-0.072	-1.413	0.350	0.914	0.833	0.80	0.91	0.94
PWTFBF2	-0.032	0.091	-1.352	0.354	0.893	0.854	0.81	0.94	0.99
WT1Y	-0.235	-1.847	-2.011	1.411	1.936	0.748	0.85	0.17	0.83
WT5Y	-0.291	-1.153	-2.131	1.339	1.660	0.754	0.95	0.51	0.84
WT5YX	-0.158	-0.538	-0.815	1.125	1.042	0.799	0.97	0.77	0.95
WT5Y5X	-0.193	-0.061	-0.430	1.168	0.993	0.941	0.98	0.95	0.98

Table 2.3: Bias, RMSE, and coverage of nominal 95% confidence/credible interval for linear slope when true model is concave and probability of selection is associated with slope: unweighted, fully weighted, standard weight pooling estimator (without and with fractional Bayes Factor priors), linear spline weight pooling estimator (without and with fractional Bayes Factor priors), and generalized design-based estimator.

2.3.2 No Association between Probability of Selection and Regression Slope

Similar to Section 2.3.1, a spline model of 20 knots is generated:

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

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

$$H_i = [2 * UNI(0, 10)]/2$$

$$P(I_i|H_i) = \pi_i \propto (1 + H_i/15)H_i.$$

In contrast to Section 3.1, the probability of selection is independent from the regression covariate. The ratio between the maximum and minimum of weights is around 35, and the sizes of potential strata are greater than 3 to be estimable.

Estimator	Average Biasness			RMSE relative to FWT			True Coverage		
	Model $\sigma^2(\log_{10})$			Model $\sigma^2(\log_{10})$			Model $\sigma^2(\log_{10})$		
	1.5	3.5	5.5	1.5	3.5	5.5	1.5	3.5	5.5
UNWT	-0.045	0.253	-0.073	0.710	0.732	0.715	0.91	0.94	0.92
FWT	-0.009	0.147	0.351	1	1	1	0.90	0.91	0.92
PWT	-0.012	0.247	-0.040	0.638	0.712	0.733	0.24	0.91	0.92
PWTFBF1	-0.011	0.271	-0.091	0.633	0.726	0.720	0.24	0.89	0.91
PWTFBF2	-0.011	0.269	-0.090	0.633	0.727	0.725	0.21	0.88	0.90
PWTLS	-0.029	0.215	-0.182	0.629	0.723	0.712	0.89	0.94	0.93
PWTLSFBF1	0.004	0.088	-0.277	0.714	0.792	0.769	0.86	0.91	0.93
PWTLSFBF2	0.006	0.114	-0.278	0.749	0.794	0.774	0.85	0.90	0.93
WT1Y	-0.027	0.246	-0.071	0.726	0.731	0.715	0.88	0.94	0.94
WT5Y	-0.015	0.259	-0.076	0.713	0.731	0.718	0.91	0.91	0.92
WT5YX	-0.020	0.264	-0.079	0.725	0.739	0.719	0.88	0.91	0.91
WT5Y5X	-0.020	0.257	-0.077	0.731	0.728	0.723	0.88	0.91	0.90

Table 2.4: Bias, RMSE, and coverage of nominal 95% confidence/credible interval for linear slope when true model is convex and probability of selection is not associated with slope: unweighted, fully weighted, standard weight pooling estimator (without and with fractional Bayes Factor priors), linear spline weight pooling estimator (without and with fractional Bayes Factor priors), and generalized design-based estimator.

Here we focus on the misspecified model setting only:

$$1. \beta_b = (0, 0, 0, 0, 0, 0, .5, .5, .5, .5, 1, 1, 1, 1, 2, 2, 2, 2, 4, 4, 4)'$$

$$2. \beta_c = (0, 22, -4, -4, -2, -2, -2, -2, -1, -1, -1, -1, -.5, -.5, -.5, -.5, 0, 0, 0, 0, 0)'.$$

Thus the relationship between Y and X is curved, regardless of strata. The linear assumption is violated, yet the population slope B remains a meaningful statistics and is our value of interest. Again a total number of 1000 elements are sampled without replacement according to inclusion probabilities for each simulation, and a total 200 simulations are conducted. Results are presented in Tables 2.4 and 2.5.

From Table 2.4 and 2.5 we observe that, as a result of no association between weight and the $X - Y$ association, none of the unweighted, model based, or model assisted estimators presents large bias, and all of these estimators have substantial

Estimator	Average Biasness			RMSE relative to FWT			True Coverage		
	Model $\sigma^2(\log_{10})$			Model $\sigma^2(\log_{10})$			Model $\sigma^2(\log_{10})$		
	1.5	3.5	5.5	1.5	3.5	5.5	1.5	3.5	5.5
UNWT	-0.014	0.171	-0.045	0.651	0.693	0.720	0.96	0.92	0.93
FWT	-0.020	0.048	0.372	1	1	1	0.95	0.89	0.93
PWT	-0.024	0.134	-0.039	0.602	0.678	0.733	0.33	0.90	0.90
PWTFBF1	-0.024	0.162	-0.068	0.602	0.678	0.720	0.32	0.91	0.92
PWTFBF2	-0.025	0.163	-0.070	0.602	0.680	0.723	0.30	0.93	0.91
PWTLS	-0.020	0.125	-0.241	0.573	0.703	0.701	0.91	0.92	0.93
PWTLSFBF1	-0.019	0.127	0.027	0.650	0.771	0.758	0.90	0.89	0.90
PWTLSFBF2	-0.020	0.178	0.033	0.651	0.775	0.789	0.90	0.87	0.91
WT1Y	-0.016	0.174	-0.047	0.689	0.695	0.720	0.94	0.93	0.93
WT5Y	-0.003	0.193	-0.046	0.701	0.720	0.724	0.94	0.90	0.93
WT5YX	-0.003	0.189	-0.044	0.694	0.714	0.724	0.94	0.92	0.93
WT5Y5X	-0.012	0.189	-0.048	0.694	0.715	0.727	0.94	0.91	0.93

Table 2.5: Bias, RMSE, and coverage of nominal 95% confidence/credible interval for linear slope when true model is concave and probability of selection is not associated with slope: unweighted, fully weighted, standard weight pooling estimator (without and with fractional Bayes Factor priors), linear spline weight pooling estimator (without and with fractional Bayes Factor priors), and generalized design-based estimator.

RMSE improvements over the the fully-weighted estimator. The exchangeable weight pooling methods provide estimators with RMSEs close to, or smaller than, the un-weighted estimator, but have extremely low coverage rates when σ^2 is small. The linear spline model and its FBF versions provide more robust results, which have slightly larger RMSEs than those from original weight pooling, but maintain only somewhat less than nominal coverage rates. The generalized design-based estimators provide results resembling the un-weighted estimator under all situations.

2.4 Application: Estimating Associations between Blood Levels of Dioxin and Age and Gender

To evaluate the proposed linear spline weight pooling model in application, we apply our method on the dioxin dataset from the National Health and Nutrition Ex-

amination Survey (NHANES). The outcome variable of interest is the amount of blood dioxin, an organic compound formed through incomplete combustion, and primarily generated by trash incineration, paper and plastics manufacturing, and smoking. The goal is to determine relationship between log scale of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in the blood and age and gender using data from 1,250 representative adult subjects interviewed during the 2003-2004 NHANES survey. The survey design consists of 25 strata, with 2 Masked Variance Units (MVU's) within each stratum to accommodate the (confidential) primary sampling units, in addition to sampling weights to account for unequal probabilities of selection, and nonresponse and post-stratification adjustment. Using the method developed in Chen et al. (2010), 674 below limit-of-detection cases are imputed, and five multiply-imputed data sets are created. Rubin's formula (*Rubin*, 1987) is used to summarize information from all imputed data sets.

Four different models are fit: TCDD on age, TCDD on gender, TCDD on age and gender, and TCDD on age, gender and their interaction. All prior distributions of parameters are defined as in the simulation. Since the population slope is unknown, we assume that the fully-weighted model provides an unbiased estimator, and calculate the relative bias accordingly. As pointed out in Kish (1992), the fully weighted estimator $\hat{\beta}_w$ is unbiased only in expectation, and the true estimated square bias of a regression coefficient $\hat{\beta}$ is given by $max((\hat{\beta} - \hat{\beta}_w)^2 - \hat{V}_{01}, 0)$, where $\hat{V}_{01} = Var(\hat{\beta}) + Var(\hat{\beta}_w) - 2Cov(\hat{\beta}, \hat{\beta}_w)$. To fully account for the design feature, all variance/covariance estimates are calculated via jackknife, that is, $Var(\hat{\beta}_w) = \sum_h \frac{k_h-1}{k_h} \sum_{i=1}^{k_h} (\hat{\beta}_{w_{(hi)}} - \hat{\beta}_w)^2$, $\hat{\beta}_{w_{(hi)}} = (X'W_{(hi)}X)^{-1}XW_{(hi)}y$, 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 $j \notin h, j \notin i$, $\frac{k_h}{k_h-1}w_j$ for all elements $j \in h, j \notin i$, and 0 for elements $j \in h, j \in i$. $Var(\hat{\beta})$ and $Cov(\hat{\beta}_w, \hat{\beta})$ are calculated accordingly, and estimates from five imputed replicate datasets are com-

Model	Est($\times 10^{-2}$)	Bias($\times 10^{-3}$)	Var($\times 10^{-6}$)	RMSE($\times 10^{-3}$)
UNWT	3.296	-1.352	1.046	1.858
WT	3.431	0	1.456	1.207
PWT	3.310	-1.215	1.014	1.237
PWTLS	3.377	-0.540	1.104	1.050
PWTLSFBF1	3.422	-0.090	1.208	1.099
PWTLSFBF2	3.413	-0.186	1.208	1.099

Table 2.6: Bias and RMSE for linear slope estimated for age: unweighted, fully weighted, standard weight pooling estimator, and linear spline weight pooling estimator (without and with fractional Bayes Factor priors). Units present in parenthesis.

Model	Est($\times 10^{-1}$)	Bias($\times 10^{-2}$)	Var($\times 10^{-3}$)	RMSE($\times 10^{-1}$)
UNWT	1.585	-8.800	1.802	1.003
WT	2.465	0	2.163	0.465
PWT	2.486	0.203	3.537	0.595
PWTLS	2.425	-0.406	3.439	0.586
PWTLSFBF1	2.416	-0.049	3.736	0.611
PWTLSFBF2	2.435	-0.304	3.548	0.596

Table 2.7: Bias and RMSE for linear slope estimated for gender: unweighted, fully weighted, standard weight pooling estimator, and linear spline weight pooling estimator (without and with fractional Bayes Factor priors). Units present in parenthesis.

bined with Rubin’s formula. The resulting estimated bias and RMSE are summarized in Tables 2.6-2.9.

For the regression model of log TCDD on gender, the unweighted model is biased. The original weight pooling method is also biased, but the reduction in variance yields slightly improved RMSEs comparing to unweighted model. The linear spline model, plus two FBF versions, all greatly reduce the bias relative to the original weight pooling method, and lead to substantially improved RMSEs over the weighted model.

In the second model estimating the age effect on log TCDD, the unweighted model suffers from severe bias for not considering the weights. All weight pooling models manage the trade-off between bias and variance inflation, achieving overall RMSE somewhat higher than the weighted model.

Model	Age			
	Est($\times 10^{-2}$)	Bias($\times 10^{-4}$)	Var($\times 10^{-6}$)	RMSE($\times 10^{-3}$)
UNWT	3.347	-9.757	1.033	1.603
WT	3.445	0	1.478	1.216
PWT	3.350	-9.449	0.729	0.854
PWTLS	3.426	-1.922	0.986	0.993
PWTLSFBF1	3.438	-0.663	1.138	1.067
PWTLSFBF2	3.430	-1.492	1.170	1.082
Model	Gender			
	Est($\times 10^{-1}$)	Bias($\times 10^{-2}$)	Var($\times 10^{-3}$)	RMSE(10^{-2})
UNWT	2.648	-0.158	1.043	3.297
WT	2.663	0	2.147	4.634
PWT	2.651	-0.121	1.628	4.035
PWTLS	2.535	-0.128	2.623	5.121
PWTLSFBF1	2.667	0.346	2.818	5.309
PWTLSFBF2	2.663	-0.046	3.034	5.508

Table 2.8: Bias and RMSE for linear slope estimated for age and gender: unweighted, fully weighted, standard weight pooling estimator, and linear spline weight pooling estimator (without and with fractional Bayes Factor priors). Units present in parenthesis.

For the third model with both age and gender, the original weight pooling method has the smallest RMSE, although the linear spline weight pooling estimator has improved bias reduction while besting the fully weighted estimator RMSE for age, and lagging it somewhat for gender.

And for the last model, we introduce the age and gender interaction in the model. The results are somewhat similar to the third model, with the linear spline model performing better than the fully weighted model with respect to RMSE for the main effects but slightly worse for the interaction, and better than the original weight pooling model with respect to bias.

2.5 Discussion

Weighted estimators whose weights are derived from the inverse of selection probabilities inflate the variance of estimators; weights derived from non-response adjust-

Age		
Model	Bias($\times 10^{-4}$)	RMSE($\times 10^{-3}$)
UNWT	-3.330	1.889
WT	0	1.909
PWT	-3.099	1.386
PWTLS	-4.010	1.663
PWTLSFBF1	-4.298	1.618
PWTLSFBF2	-2.991	1.727
Gender		
Model	Bias($\times 10^{-2}$)	RMSE($\times 10^{-1}$)
UNWT	4.613	1.074
WT	0	1.122
PWT	4.727	0.911
PWTLS	-0.955	1.097
PWTLSFBF1	-1.090	1.069
PWTLSFBF2	-0.795	1.108
Interaction		
Model	Bias($\times 10^{-3}$)	RMSE($\times 10^{-3}$)
UNWT	-1.371	2.269
WT	0	2.039
PWT	-1.392	1.876
PWTLS	0.074	2.224
PWTLSFBF1	0.096	2.174
PWTLSFBF2	0.039	2.279

Table 2.9: Bias and RMSE for linear slope estimated for age, gender and interaction: unweighted, fully weighted, standard weight pooling estimator, and linear spline weight pooling estimator (without and with fractional Bayes Factor priors). Units present in parenthesis.

ment or poststratification may reduce variance as well as bias, but this is often not the case in practice. This variance inflation could be countered by a prediction estimation process that accounts for the unequal probability of selection. First a mean structure for sampled observations y is proposed, such as a linear combination including another covariate $X\beta$. Then one could fit the proposed model, draw parameters from their estimated distributions, or posterior distributions in case under a Bayesian framework, and create the predicted y based on drawn parameters. Generating a posterior predictive distribution of the population under this model allows estimation of the posterior predictive distribution of the population quantity of interest, such as a mean or regression parameter. Generating from a model that accommodates interactions between the probability of selection and the subpopulation quantity of interest corresponding to the population quantity of interest among subjects with similar probabilities of selection mimics the design-based, fully-weighted estimator and helps to account for any discrepancies between the model population quantity of interest and the true model generating the population data or non-ignorable selection models, but at the same time can inflate noise when selection is ignorable and model specification is approximately correct. “Weight pooling” models rely on Bayesian model averaging to generate predictive estimators that collapse interaction terms to the degree when data suggest they are not needed in the model.

The linear spline weight pooling model we presented provides an advantage in three ways over a preliminary weight pooling model developed in Elliott (2008, 2009) when considered in a linear regression framework. First, it may eliminate the need for the fractional Bayes factor prior that is necessary to induce sufficient pooling over the fully-weighted estimator. Second, in settings where there is no association between the probability of selection and the regression slope, the previous weight pooling model tends to have credible intervals that are far too narrow, leading to severe undercoverage, whereas the linear spline weight pooling estimator has only slight

undercoverage. All forms of the weight pooling estimator either match the efficiency of the fully weighted estimator when the unweighted estimator is substantially biased due to model misspecification, or have substantially increased efficiencies over the fully weighted estimator when model misspecification is modest or non-existent. The proposed linear spline weight pooling estimator tends to be somewhat more efficient than the original weight pooling estimator in the latter settings. The application results are more mixed, but it must be kept in mind that estimating bias and RMSE requires the fully weighted estimator to be unbiased, which is typically not the case in small sample settings, especially for non-linear population statistics such as the regression coefficients considered.

A variety of extensions are possible to the work proposed here. It is expected that any estimator from pooling model functions as a shrinkage estimator between estimation from unweighted model and fully weighted model, representing a trade-off between bias and variance. While the above assumption remains correct, however, in application the selected pooling model may perform worse than both unweighted and fully-weighted ones. One example is that where use of weights is not required for bias correction, small weight strata may pooled together, but large weight strata remain separate, leading to unnecessary variance inflation comparing to fully weighted model, while the ideal result is a weight pooling model collapse all large weight strata together to perform a variance reduction similar to weight trimming. Thus instead of using non-informative prior, more specific prior encouraging the collapse of large weight strata could help achieve better results, and may be of value to explore further here. Second, in this chapter we present weight strata ordered by inclusion probabilities. In general, the order of weight strata could be either based on inclusion probabilities, or some other natural ordering. For example, if probabilities of inclusion vary by age, it is appropriate to define weight strata by individuals with close ages. Also, our analysis assumes that the weights are not available to the analyst for the non-sampled

elements of the population. Finite population Bayesian approaches that incorporate weights or other design variables directly into prediction if they are available for the non sampled population are considered in Zheng and Little (2003, 2004, 2005) and Chen et al. (2010, 2012).

Though the Bayesian approach has been in vogue for several decades, it has received less attention in practice in survey settings for at least two reasons. First, it is well known that unless the variables Z used in formulating the design are unrelated to the survey outcomes, model-based estimates which ignore the sample design features can be biased (*Little, 1983*). It was believed that conceptualizing a model that incorporates all design features such as unequal probabilities of selection, clustering and stratification was too difficult (*Hansen et al., 1983*) and even if one was to conceptualize such a model, its implementation in practice would be computationally impossible. But with the advent of powerful computers and modern Bayesian computational methods, it is now possible to develop and implement realistic models to address complex design features, nonresponse, coverage and measurement errors in developing both population level and subpopulation level estimates, leading to new approaches to improve over existing design-based approaches in many practical settings. Our results in this manuscript we believe contribute to this effort.

CHAPTER III

Weight Smoothing for Generalized Linear Models Using a Laplace Prior

3.1 Introduction

Studies based on data sampled with unequal inclusion probabilities typically apply case weights equal to the inverse of probabilities of inclusion to reduce or remove the bias in estimators of population quantities of descriptive interest, 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; Pfeiffermann, 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 scheme is informative, weighted estimators typically have reduced 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, and lead to an overall larger mean square

error (MSE) than would be the case if the weights are ignored or at least controlled in some fashion.

Weight trimming, or “winsorization,” is used to control the variation in weights, or more precisely, to cap the weights at some value w_0 , and redistribute the values above w_0 among the rest (*Alexander et al.*, 1997; *Kish*, 1992; *Potter*, 1990). Various approaches have been developed in creating different criteria to determine the cap value based on data. Some example includes NAEP method by Potter (1988), which set the cutoff point equal $\sqrt{c \sum_{i \in s} w_i^2 / n}$, where c was chosen in an ad-hoc manner. Cox and McGrath (1981) approached it by estimating the cutoff point value which optimized the empirical MSE estimated by $M\hat{S}E(\hat{\theta}_t) = (\hat{\theta}_t - \hat{\theta}_w)^2 - \hat{V}ar(\hat{\theta}_t) + 2\sqrt{\hat{V}ar(\hat{\theta}_t)\hat{V}ar(\hat{\theta}_w)}$, where $\hat{\theta}_w$ was the fully weighted estimator, and $\hat{\theta}_t$, $t = 1, \dots, T$, was the weight trimmed estimator, with t denoting various trimming levels, from 1 as the unweighted estimator to T as the fully-weighted estimator. Chowdbury et al. (2007) suggested treating weights as coming from a skewed cumulative distribution (e.g., an exponential distribution), and using the upper 1% of the fitted distribution as a cut point for weight trimming. Beaumont (2008) proposed a generalized design-based method, replacing the actual weights with values predicted on some form of response and design variables. Details of these design-based approaches are summarized in (Henry 2012).

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 can be obtained by building a model that contains indicators for weight strata together with interaction terms between 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 strata are smoothed if data provide little evidence of difference between them, 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 (2008, 2009) extended the application to linear and generalized linear models, and discussed different settings for the random effect priors, namely exchangeable, autoregressive, linear and nonparametric random slopes, and evaluated their performances.

In this chapter we consider extending the weight smoothing approach by the 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 non-ignorable sampling. In addition, considering the prevailing performance of Laplace prior in sparse model selection, we expect the hierarchical model to properly smooth the strata when data provide no evidence in difference among strata, even under simple mean and covariance matrix settings such as exchangeable random priors, while maintaining its bias-reduction feature when it is needed. We evaluate the performance of our proposed model in a simulation study, under both model misspecification and informative sampling, for both numerical and dichotomous outcomes, and compare it with competing methods. This chapter 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

a simulation study, 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 on Dioxin Dataset from NHANES and Partner of Child Passenger Safety Dataset. Section 5 provides a summary discussion.

3.2 Weight Smoothing Methodology

3.2.1 Finite Bayesian Population Inference

For finite Bayesian population inference, we model the the population data Y : $Y \sim f(Y|\theta, Z)$, where Z are 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 semi-parametric or non-parametric 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

$$p(Y_{nob} | Y_{obs}, I, Z) = \frac{\int \int p(Y_{nob} | Y_{obs}, Z, \theta, \phi) p(I | Y, Z, \theta, \phi) p(Y_{obs} | Z, \theta) p(\theta, \phi) d\theta d\phi}{\int \int \int p(Y_{nob} | Y_{obs}, Z, \theta, \phi) p(I | Y, Z, \theta, \phi) p(Y_{obs} | Z, \theta) p(\theta, \phi) d\theta d\phi dY_{nob}} \quad (3.1)$$

where Y_{nob} consists of the $N - n$ unobserved cases in the population, and ϕ models the inclusion indicator I . 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 (3.1) reduces to

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

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). Notice that if inference about quantities $Q(Y|X)$ involving covariates X is desired (e.g., regression slope), noninformative or ignorable sample designs can be relaxed to have distribution of I depend on X .

3.2.2 Weight Prediction

Beaumont (2008) proposed a model-assisted method, tamping down the extreme values in weights by replacing them with their predicted 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, i.e. $I_i = 1$ as i th unit sampled and $I_i = 0$ otherwise, $Y = (Y_1, \dots, Y_N)^T$ the vector of survey response variable, 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 $\tilde{w}_i = E_M(w_i|I_i = 1, z_i, y_i)$, or sometimes reduced to $\tilde{w}_i = E_M(w_i|I_i = 1, y_i)$. Beaumont (2008) discussed two estimators, 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 $v_i > 0$ were 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 sampled data, then the re-weighted estimator of the survey response variable of interest is obtained using the predicted weights.

3.2.3 Weight Smoothing

In general, weight smoothing stratifies data by inclusion probabilities, and applies a hierarchical model treating strata means as random effects, thus achieves trimming via shrinkage. Considering the population mean as the quantity of interest, an

example weight smoothing model is as following:

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

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

where $\mu = (\mu_1, \dots, \mu_H)$, $\phi = (\phi_1, \dots, \phi_H)$, and $h = 1, \dots, H$ indexes different “weight strata” defined, e.g., by same or similar inclusion probabilities. We assume ϕ , D , and σ^2 all have weak or non-informative priors. Notice that the weight strata are not necessarily ordered by inclusion probabilities, but could be in a more natural ordering, 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 $\hat{\mu}_h = E(\mu_h|y)$. 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; Lazzeroni 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 $\hat{\mu}_h = w_h \bar{y}_h + (1 - w_h) \tilde{y}$, for $w_h = \tau^2 n_h / (\tau^2 n_h + \sigma^2)$ and $\tilde{y} = (\sum_h n_h / (n_h \tau^2 + \sigma^2))^{-1} \sum_h n_h / (n_h \tau^2 + \sigma^2) \bar{y}_h$. 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 $\tilde{y} = \frac{\sum_h n_h \bar{y}_h / \sigma^2}{\sum_h n_h / \sigma^2} = \bar{y}$ 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 stratum differ, the weight smoothing model achieves a “data-driven” compromise between the weighted estimator, which is design consistent but may be highly inefficient, and unweighted estimator, which is fully efficient when the assumption of independent between inclusion probability and mean of Y holds, but is likely biased otherwise.

3.2.4 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

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

where $a_i(\phi)$ is a known function of (nuisance) scale parameter ϕ , 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 $g(\cdot)$ 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, $Var(y_i | \theta_i) = a_i(\phi) V(\mu_i)$, 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 simplified results that $V(\mu_i) = 1/g'(\mu_i)$. Some examples include Gaussian (linear) regression, where $a_i(\phi) = \sigma^2$ and the canonical link $g(\mu_i) = \mu_i$; logistic regression, where $a_i(\phi) = n_i^{-1}$ and the canonical link $g(\mu_i) = \log(\mu_i / (1 - \mu_i))$, and Poisson regression, where $a_i(\phi) = 1$ and the canonical link $g(\mu_i) = \log(\mu_i)$.

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

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

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 moderately 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 $g'(\mu_i) = 1$,

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

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

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

where $\text{expit}(\cdot) = \exp(\cdot)/(1 + \exp(\cdot))$. In practice, approximate posterior distributions of B can be obtained by replacing the observed y_{hi} with predicted values $g(x'_{hi}\hat{\beta}_h)$ for each draw of $\hat{\beta}_h$ and obtaining the pseudo-MLE for the chosen regression model.

3.2.5 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 normal distribution prior which restricts the variation between random effect term and prior mean in an $L2$ manner, Laplace prior measures by the $L1$ distance. According to Eltoft et al.(2006), the general form of Multivariate Laplace distribution is given by:

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

where y is a d -dimensional random variables $y = (y_1, \dots, y_d)$; $K_m(x)$ 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, such a format is inconvenient for application. The alternative approach is to represent Laplace distribution as a scale mixture of normals with an exponential mixing density. By creating a set of latent mixing

variables $D_\tau = \text{diag}(\tau_1^2, \dots, \tau_{H_p}^2)$, and applying exchangeable random slope (XRS) setting, we reach the two level hierarchical form of Laplace Prior for β :

$$\begin{aligned}
(\beta_1^T, \dots, \beta_H^T)^T | \beta_h^*, D_\tau, \sigma^2 &\sim \text{MVN}(\beta_h^*, \sigma^2 D_{\tau h}) \\
\beta_h^* | \sigma_0^2 &\sim \text{MVN}(0, \sigma_0^2 I_p) \\
D_{\tau h} &= \text{diag}(\tau_{h1}^2, \dots, \tau_{hp}^2) \\
\sigma^2, \tau_1^2, \dots, \tau_{H_p}^2 &\sim 1/\sigma^2 \prod_{j=1}^{H_p} \frac{\lambda^2}{2} e^{-\lambda^2 \tau_j^2 / 2} \\
\lambda^2 &\sim \text{Gamma}(r, \delta).
\end{aligned}$$

The first level of the model depends on the distribution assumption of the generalized linear model used. In this paper, 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:

$$\begin{aligned}
y_{hi} | x_{hi}, \beta_h, \sigma^2 &\sim N(x_{hi}^T \beta_h, \sigma^2) \\
(\beta_1^T, \dots, \beta_H^T)^T | \beta_h^*, D_\tau, \sigma^2 &\sim \text{MVN}(\beta_h^*, \sigma^2 D_{\tau h}) \\
\beta_h^* | \sigma_0^2 &\sim \text{MVN}(0, \sigma_0^2 I_p) \\
D_{\tau h} &= \text{diag}(\tau_{h1}^2, \dots, \tau_{hp}^2) \\
\sigma^2, \tau_1^2, \dots, \tau_{H_p}^2 &\sim 1/\sigma^2 \prod_{j=1}^{H_p} \frac{\lambda^2}{2} e^{-\lambda^2 \tau_j^2 / 2} \\
\lambda^2 &\sim \text{Gamma}(\gamma = 1, \delta = 1.78).
\end{aligned}$$

Following the deduction in Park & Casella (2008), the analytical forms of all fully

conditional distributions of β , σ^2 etc are achievable, and the posterior predictive distribution could be obtained through a Gibbs Sampler as below. A detailed derivation is attached in the Appendix B.1.

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

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:

$$\begin{aligned}
y_{hi} | x_{hi}, \beta_h &\sim \prod_{h=1}^H \prod_{i=1}^{n_h} \left(\frac{\exp(x_{hi} \beta_h)}{1 + \exp(x_{hi} \beta_h)} \right)^{y_{hi}} \left(\frac{1}{1 + \exp(x_{hi} \beta_h)} \right)^{1 - y_{hi}} \\
(\beta_1^T, \dots, \beta_H^T)^T | \beta_h^*, D_{\tau} &\sim MVN(\beta_h^*, D_{\tau h}) \\
\beta_h^* | \sigma_0^2 &\sim MVN(0, \sigma_0^2 I_p) \\
D_{\tau h} &= \text{diag}(\tau_{h1}^2, \dots, \tau_{hp}^2) \\
\tau_1^2, \dots, \tau_{Hp}^2 &\sim \prod_{j=1}^{Hp} \frac{\lambda^2}{2} e^{-\lambda^2 \tau_j^2 / 2} \\
\lambda^2 &\sim Gamma(r = 1, \delta = 1.78).
\end{aligned}$$

When the first level is not normally distributed, the full conditional distribution of

β does not belong to any known distributions, and thus direct sampling is impossible. Instead we apply 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 previous Gibbs Sampler algorithm, and are directly drawn from their fully conditional distributions as below: (full derivation in Appendix B.2)

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

3.3 Simulation Study

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

3.3.1 Hierarchical Weight Smoothing Model for Ordinary Linear Regression

We generate a population of $N = 20,000$ for ordinary linear regression. The predictor X is uniformly distributed on the interval from 0 to 10, and is equally divided into 20 strata with a range of 0.5 each. The response variable Y is then generated as a spline function of X with cutpoints between strata as knots. Three sets of coefficients are applied separately, so the pattern of $Y | X$ varies from straight slope to increasing curve and decreasing curve.

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

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

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

$$\beta_b = (0, 0, 0, 0, 0, 0, 0, .5, .5, .5, .5, 1, 1, 1, 1, 2, 2, 2, 2, 4, 4)$$

$$\beta_c = (0, 11, -4, -4, -2, -2, -2, -2, -1, -1, -1, -1, -0.5, -0.5, -0.5, -0.5, 0, 0, 0, 0, 0)$$

From the population, a sample of $n = 1000$ is selected without replacement, according to inclusion probabilities equal to $\pi_i = (1 + i/30) * i/2$ for the i th stratum. Thus the ratio between the maximum and minimum of weights is about 35, and the sample size of each stratum is always greater than three. Z is created as $Z = I \otimes X$, where $I = c(I_1, \dots, I_h)$ is an indicator vector stating if the current observation belongs to i th stratum. Z is centered within each column with respect to each stratum (for computation convenience), and used as predictor in the simulations.

Our inferential target is $B = (\sum_{i=1}^N \tilde{X}_i \tilde{X}_i')^{-1} \sum_{i=1}^N \tilde{X}_i Y_i$ 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 may be most efficient. Population variance σ^2 varies among 10, 10^3 and 10^5 , creating varying level of variance influence compared to possible bias;

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 1000 to approximate a non-informative prior; the prior for λ follows a gamma hyperprior with parameter $r = 1$ and $\delta = 1.78$, as suggested by Park and Casella (2008). All other parameters in simulation are initialized at zero, except for variance estimator σ^2 , which is initialized at one. A Gibbs Sampler method is applied, that is, for each iteration, all parameters are sequentially drawn from the full conditional distribution. Then to obtain the estimate from 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 10000 times, with a burn-in of 2000. Diagnostic plots are generated to assure the algorithm's convergence. Bias, RMSE and 95% coverage are recorded for comparison. 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 four variations of the model-assisted estimators from Beaumont (2008): predicted weights on y only (PREDY); predicted weights on degree 5 polynomial of y (PREDY5); 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 according to the fully weighted estimator. Results are provided in Table 3.1, 3.2 and 3.3.

Under β_a , where the model is correctly specified, all methods yield unbiased results, and the unweighted estimator maintains the best efficiency, with an approxi-

	$\sigma^2 = 10^{1.5}$			$\sigma^2 = 10^{3.5}$			$\sigma^2 = 10^{5.5}$		
	Bias	RMSE	cover	Bias	RMSE	cover	Bias	RMSE	cover
UNWT	-0.013	0.730	0.95	0.009	0.693	0.95	-0.766	0.713	0.96
FWT	-0.006	1	0.94	-0.012	1	0.93	-1.253	1	0.96
HWS	-0.006	1.047	0.96	-0.019	0.978	0.95	-1.107	0.768	0.99
XRS	-0.001	1.489	1	0.016	0.720	0.96	-0.765	0.716	0.96
PREDY	-0.014	0.805	0.96	0.025	0.698	0.95	-0.781	0.711	0.97
PREDY5	0.108	1.511	0.52	0.104	0.756	0.95	-0.785	0.717	0.97
PREDYX	-0.005	0.803	0.97	-0.016	0.783	0.94	-1.145	0.796	0.95
PREDYX5	-0.005	0.978	0.94	-0.019	0.953	0.94	-1.540	0.983	0.95

Table 3.1: Comparison of various estimators of slope B_1 under β_a linear spline setting. Bias and RMSE under populations with residual variance $10^{1.5}$, $10^{3.5}$ and $10^{5.5}$ from following model: unweighted, fully weighted, hierarchical weight smoothing, exchangeable random effect and weight prediction by y, degree 5 polynomial of y, linear combination of x and y, and degree 5 polynomial of x,y.

	$\sigma^2 = 10^{1.5}$			$\sigma^2 = 10^{3.5}$			$\sigma^2 = 10^{5.5}$		
	Bias	RMSE	cover	Bias	RMSE	cover	Bias	RMSE	cover
UNWT	1.980	10.201	0	1.993	2.441	0.02	1.204	0.726	0.95
FWT	-0.006	1	1	-0.005	1	0.92	-1.252	1	0.96
HWS	-0.006	0.453	0.97	-0.042	0.947	0.96	-1.354	0.774	0.99
XRS	-0.103	1.008	0.95	1.769	2.213	0.04	1.203	0.729	0.94
PREDY	0.963	4.977	0	1.794	2.228	0.03	1.174	0.722	0.94
PREDY5	1.059	5.466	0	1.791	2.212	0.03	1.184	0.730	0.95
PREDYX	0.368	2.050	0.32	0.385	0.919	0.90	-0.746	0.792	0.95
PREDYX5	0.018	1.000	1	0.012	0.955	0.96	-1.515	0.983	0.96

Table 3.2: Comparison of various estimators of slope B_1 under β_b linear spline setting. Bias and RMSE under populations with residual variance $10^{1.5}$, $10^{3.5}$ and $10^{5.5}$ from following model: unweighted, fully weighted, hierarchical weight smoothing, exchangeable random effect and weight prediction by y, degree 5 polynomial of y, linear combination of x and y, and degree 5 polynomial of x,y.

	$\sigma^2 = 10^{1.5}$			$\sigma^2 = 10^{3.5}$			$\sigma^2 = 10^{5.5}$		
	Bias	RMSE	cover	Bias	RMSE	cover	Bias	RMSE	cover
UNWT	-1.874	6.227	0	-1.836	2.177	0.01	-2.611	0.758	0.9
FWT	-0.006	1	1	-0.020	1	0.97	-1.257	1	0.95
HWS	-0.005	0.337	0.85	0.069	0.937	0.96	-0.772	0.772	0.99
XRS	-0.549	1.872	0.06	-1.721	2.052	0.01	-2.609	0.761	0.91
PREDY	-0.009	1.258	0.97	-1.738	2.075	0.01	-2.627	0.756	0.90
PREDY5	-0.167	1.131	0.99	-1.256	1.757	0.37	-2.593	0.761	0.88
PREDYX	-0.327	1.416	0.75	-0.729	1.091	0.75	-1.861	0.809	0.93
PREDYX5	-0.020	1.019	1	-0.055	0.965	0.96	-1.557	0.983	0.95

Table 3.3: Comparison of various estimators of slope B_1 under β_c linear spline setting. Bias and RMSE under populations with residual variance $10^{1.5}$, $10^{3.5}$ and $10^{5.5}$ from following model: unweighted, fully weighted, hierarchical weight smoothing, exchangeable random effect and weight prediction by y, degree 5 polynomial of y, linear combination of x and y, and degree 5 polynomial of x,y.

mate 30% decrease in RMSEs comparing to fully weighted estimator. The original weight smoothing method under XRS tends to provide unstable results, inflating the variance when population signal is strong, but achieving similar RMSEs as the unweighted estimator when the population signal is weak relative to the noise. Our model, under the same XRS assumption but with a Laplace prior, gives more stable results that resemble the fully weighted estimator when variance is low, but increases in efficiency as population variance increases. Both the XRS and HWS estimators have correct to somewhat conservative coverage when the linear model is correctly specified. Most model-assisted estimators have improved RMSEs comparing to the fully weighted estimator, with the exception of PREDY5, which has unstable results and poor nominal coverage when $\sigma^2 = 10^{1.5}$.

For scenarios under β_b and β_c , the unweighted estimator of B is biased, and the fully weighted estimator strongly prevails over unweighted estimator with respect to both RMSE and 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 sophisticated enough 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, though it suffers a moderate drop in coverage under β_c and $\sigma^2 = 10^{1.5}$. Most of the model-assisted estimators are insufficiently structured to reduce bias in the small-to-medium residual variance settings, except for PREDYX5, which mimics the fully weighted estimator and thus has little savings in relative RMSEs under any of the scenarios.

3.3.2 Hierarchical Weight Smoothing Model for Logistic Regression

Following Elliott (2007), we set up population in two approaches: 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 following:

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

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

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 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 equals 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 r range from 0.05 to 0.95 to represent different level 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 maximum weight and minimum weight, and the correlation between selection probability and Y varies from 0 to 30 % as the correlation between Z^* and Y increases from .05 to .95. 20 strata of equal size are created by pooling observations with similar selection probabilities together.

From this population, samples with $n = 1000$ 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 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 10000 times, with a burn-in of 2000.

We compare the properties of our Laplace model (HWS) with same major competitors as in the linear regression setting, with the exception of (PREDY5): since Y

	$c = 0$			$c = .45$			$c = .80$		
	Bias	RMSE	Cover	Bias	RMSE	Cover	Bias	RMSE	Cover
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
PREDY	0.014	0.765	0.95	0.082	1.614	0.27	0.136	3.115	0
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 3.4: Comparison under model misspecification. Bias and RMSE under populations with underlying model quadratic coefficient 0, .45 and .80 from following model: unweighted, fully weighted, hierarchical weight smoothing, exchangeable random effect and weight prediction by y, degree 5 polynomial of y, linear combination of x and y, and degree 5 polynomial of x,y.

is a binary variable, higher-order polynomials are not relevant. Bias and nominal 95% coverage are recorded directly, while RMSE is rescaled according to fully weighted estimator. Results are provided in Table 3.4 and 3.5.

While comparing different models under model misspecification setting, the unweighted model has increased bias since the population model is less correctly specified, resulting in a change from efficient estimate to a poor estimate (Relative RMSEs range from 69.7% to 281.9% of FWT's as C increases) and poor coverage as misspecification increases. The exchangeable random slope model estimator is not robust, with bias similar to unweighted model, and larger RMSEs 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, although coverage suffers to a moderate degree. The weight prediction models PREDY and PREDYX perform similar to unweighted estimate, gaining efficiency when models are correctly specified, and suffering when misspecification increases. PREDYX5, which predicts weights with a degree five polynomial of both x and y, essentially mimics the fully-weighted estimator.

	$r = .05$			$r = .50$			$r = .95$		
	Bias	RMSE	Cover	Bias	RMSE	Cover	Bias	RMSE	Cover
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
PREDY	0.055	1.034	0.70	0.071	1.233	0.54	0.062	1.079	0.64
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

Table 3.5: Comparison under informative sampling. Bias and RMSE under populations with correlation between Z and Y equal to .05, .50 and .95 from following model: unweighted, fully weighted, hierarchical weight smoothing, exchangeable random effect and weight prediction by y, degree 5 polynomial of y, linear combination of x and y, and degree 5 polynomial of x,y.

Under informative sampling, the unweighted estimator has only slightly larger RMSEs 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 12% compared to the FWT estimator, although coverages suffer to a moderate degree except when the sampling is highly informative. PREDY is modestly biased but has poor coverage (perhaps not surprising given that Y is binary), while PREDYX improves RMSE by up to 17% while having only slight undercoverage. PREDYX5 again mimics the fully weighted model.

3.4 Application

3.4.1 Application on Dioxin Data from NHANES

To demonstrate the performance of our method in 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, 1250 representative adult subjects were selected under 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 like age and gender are also available from the survey. The sampled data are stratified into 25 strata, within each consists of 2 Masked Variance Units (MVU's) for proper variance estimation procedure, with survey weights provided as well. Due to technical limit, 674 readings are below limit of detection, and are imputed through multiple imputation using the model described in Chen et al. (2010), resulting in 5 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 values 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 (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 the true estimated square bias of regression coefficient

Model	Bias($\times 10^{-3}$)	RMSE($\times 10^{-3}$)
UNWT	-1.262	3.265
WT	0	3.888
HWT	-0.086	1.214

Table 3.6: Regression of log TCDD on Age. Bias and RMSE for linear slope estimated for age: unweighted, fully weighted and hierarchical weight smoothing.

Model	Bias($\times 10^{-2}$)	RMSE($\times 10^{-1}$)
UNWT	-8.219	1.248
WT	0	0.637
HWT	0.589	0.607

Table 3.7: Regression of log TCDD on Gender. Bias and RMSE for linear slope estimated for gender: unweighted, fully weighted and hierarchical weight smoothing.

$\hat{\beta}$ given by $\max((\hat{\beta} - \hat{\beta}_w)^2) - \hat{V}_{01}$, where $\hat{V}_{01} = \hat{V}ar(\hat{\beta}) + \hat{V}ar(\hat{\beta}_w) - 2\hat{C}ov(\hat{\beta}, \hat{\beta}_w)$. To fully account for the design feature, all variance/covariance estimates are calculated via jackknife as $\hat{V}ar(\hat{\beta}_w) = \sum_h \frac{k_h-1}{k_h} \sum_{i=1}^{k_h} (\hat{\beta}_{w_{(hi)}} - \hat{\beta}_w)^2$, $\hat{\beta}_{w_{(hi)}} = (X'W_{(hi)}X)^{-1}XW_{(hi)}y$, 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 $j \notin h, j \notin i$, $\frac{k_h-1}{k_h}w_j$ for all elements $j \in h, j \notin i$, and 0 for elements $j \in h, j \in i$. $\hat{V}ar(\hat{\beta})$ and $\hat{C}ov(\hat{\beta}_w, \hat{\beta})$ are calculated accordingly, and estimates from five imputed replicate datasets are combined with Rubin's formula. The results are based on 10000 iterations with first 2000 draws discarded as burn-in. And the resulting Biases and RMSEs are summarized in Tables 3.6-3.9.

For the first two models of log TCDD on age and gender separately, the estima-

Model	Age		Gender	
	Bias($\times 10^{-4}$)	RMSE($\times 10^{-3}$)	Bias($\times 10^{-2}$)	RMSE($\times 10^{-2}$)
UNWT	-9.067	3.296	-0.159	9.017
WT	0	3.895	0	6.161
HWS	-0.841	1.227	1.058	5.659

Table 3.8: 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.

Model	Age		Gender		Interaction	
	Bias ($\times 10^{-4}$)	RMSE ($\times 10^{-3}$)	Bias ($\times 10^{-2}$)	RMSE ($\times 10^{-1}$)	Bias ($\times 10^{-3}$)	RMSE ($\times 10^{-3}$)
UNWT	-5.063	3.758	2.882	1.591	-0.880	3.285
WT	0	2.661	0	3.259	0	7.335
HWS	-9.142	2.048	-6.530	1.282	1.646	2.667

Table 3.9: 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.

tions of the single predictor from unweighted model appear to be biased comparing to fully weighted model, resulting in estimated biases about 40% and 70% of RMSEs. However, the weighted model also fails to provide a efficient estimate for effect on age, supported by a RMSE of 3.888, larger than 3.265 from the unweighted model. Meanwhile, the hierarchical weight smoothing model shows its ability to improve efficiency, both reducing the biases comparing to unweighted model, and maintaining RMSEs similar to or smaller than fully weighted model depend on the severity of variance inflation.

As more predictors enter the model, the estimated bias rapidly decreases in scale, leading to a scenario that both bias and inflation in variance could dominate the overall RMSEs, and neither unweighted model nor 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 RMSEs comparing to either unweighted estimator or fully weighted estimator.

3.4.2 Application on Partner for Child Passenger Safety Data

In this section, we use Partners for Child Passenger Safety dataset to demonstrate our method’s performance under logistic regression setting. Unit observations in the dataset are damaged vehicles disproportionately sampled from State Farm claim records between December 1998 and December 2005, when at least one child occupant less than 15 years of age gets involved in a model year 1990 or newer State Farm-

insured vehicle. The focus of the study is children's consequential injuries, defined by either facial lacerations or other injuries rated 2 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 on this rare outcome, the overall population is divided into three strata based on injury status – vehicles with at least one child occupant screened positive for injury, vehicles with all child occupants reported receiving medical treatment but screened negative for injury, and vehicles with no occupant receiving medical treatment – and is crossed with two strata defined by whether the vehicle is drive-able or not. Since the stratification is associated with risk of injury, and cannot be fully explained by other auxiliary variable, the sampling design is informative, with weights varying from 1 to 50, and 9% of weights lying outside 3 times their standard deviations.

As determined by Winston, Kallan, Elliott, Menon and Durbin (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(compact 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 previous section, then the Gibbs sampler is executed for 10,000 iterations with 2,000 burn-in, and odds ratios are compared with unweighted and fully weighted model.

The estimated odds ratios for compacted extended cab pickups indicator don't vary much from unadjusted model to fully adjusted model, while unweighted regression and fully weighted regression lead to quite different result, from a OR of 3.534

	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)
XRS	8.681 _(2.790,27.007)	6.725 _(2.162,20.922)

Table 3.10: 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.

to 11.317 for unadjusted model, and from 3.448 to 13.890 when all other control variables are included. Both hierarchical weight smoothing model and exchangeable random effect model provide estimates lying in between the unweighted and weighted estimates, although estimation from HWS model tends to match estimation from fully weighted model. It is also worth noting that with similar point estimates, HWS model provides a considerable reduction in estimated standard deviation, leads to a smaller 95% confidence interval comparing to fully weighted model, an characteristic also presented in simulation study before.

3.5 Discussion

Generally, most methods for weight trimming, both design based and model based, handle sampling weights by achieving a balance between bias and variance, resulting in estimates usually lying between those from the unweighted model and fully weighted model. However, the weight smoothing model with Laplace prior shows the potential to provide a more efficient estimate than either unweighted model or fully weighted model at same time. This occurs especially when the model is misspecified, and population variance is small so the weight smoothing model is able to model the underlying data structure precisely, and yield an estimate with greatly reduced RMSE. However, this aggressive estimation comes at the cost of robustness, that is, the overly reduced variance could lead to poor coverage rate. As presented in the sim-

ulation, the HWS model suffers a moderate drop in coverage rate when population variance is small. It is worth exploring in future the model’s mechanism in reducing the overall RMSE, and the limit of the scenarios under which it still maintains reasonable coverage.

Comparing the results of the Laplace prior weight smoothing models with the model-assisted 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 approximately correct nominal coverage of credible intervals. In contrast, the model-assisted estimators can in some settings “oversmooth” weights when bias correction is needed and yield unstable estimates when the weight prediction is weak. The predicted weights in the model-assisted approach incorporate information from design variables, thus yield 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 statistics 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. Ultimately we find attempts to model weights rather than data misguided, as it 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: i.e., calibrated Bayes estimators (*Little, 2011*).

CHAPTER IV

Weighted Dirichlet Process Mixture Model to Accommodate Complex Sample Designs for Linear and Quantile Regression

4.1 Introduction

When estimating certain population quantities of interest from observations sampled with disproportional selection probabilities, the standard design-based inference approach considers population values as fixed, sampling indicators as random, and focuses on developing an estimator that is at least approximately unbiased with respect to the repeated sampling distribution. One example is the Horvitz and Thompson (H-T) estimator (*Horvitz and Thompson, 1952*) that targets population totals and means. The design-based approach does not make distributional assumptions (at least explicitly – the H-T estimator can be derived from a particular projection estimator of the population mean or total assuming a no-intercept linear relationship between the mean of the outcome and sampling probability, with the standard deviation of the residual proportional to the sample probability (*Zheng and Little, 2003*)). However, the design-based approach could be very inefficient under certain scenarios, such as when the sample size is small, the weights are highly variable, and/or the relationship between the quantity of interest and the probability of selection is weak.

For example, if one is interested in a population-level linear or generalized linear regression parameter, given the model is correctly specified, incorporating sampling weights in estimation is unnecessary for bias correction and will likely inflate variance (although if the errors are heteroscedastic and proportional to the probability of selection, a correct model will typically lead to fully weighted estimators). However, misspecified models and/or designs with non-ignorable inclusion mechanisms can lead to settings where weights are required for bias correction. Since “all models are wrong, but some are useful” (e.g., a linear approximation may be reasonable at a population level, but an unweighted estimator may enhance a modest quadratic trend), and since non-response or other features of the sample design not fully captured by design variables available to the analyst can lead to non-ignorable inclusion mechanisms, sampling weights are often used in estimating model parameters. For example, pseudo-maximum likelihood estimators use case weights in score equations and associated robust or “sandwich” estimators of variance to obtain inference (*Binder*, 1983).

An alternative approach uses Bayesian finite population inference, a model-based method that assumes a model for the observed data. The unobserved elements of the population are treated as missing data, and posterior predictive distributions of the population are generated by repeatedly imputing the unobserved elements of the population using draws from the posterior distribution of the parameters governing the data model. In the model-based setting, accounting for design elements in the model adds robustness to the inference. For example, if a linear relationship is assumed between a continuous covariate X and the (possibly transformed) mean of an outcome Y , unequal probabilities of selection associated with X can diminish or enhance non-linearities in the relationship between X and $E(Y | X)$ in the sampled data, leading to bias in estimating the population level linear approximation between X and $E(Y | X)$. One method to accommodate weights in a regression model setting

is to create dummy variables stratified by equal or approximately equal case weights and include indicators for weights strata and interactions between the covariates of interest and the weight strata in the regression model. Inference is made based on the posterior predictive distribution of the population level regression model of interest, which no longer needs to explicitly include the weight interactions, since they have been used to construct the posterior predictive distribution of the population to which the regression model of interest is fitted. This approach also suggest the possibility of developing models that retain the level of structure needed to incorporate the design features when necessary, but to default back to simpler models that ignore these features and are more efficient when the data suggest that they are not needed. Thus a carefully proposed model could provide an efficient estimator while maintains reasonable bias reduction.

For this purpose we investigate mixture models, which are able to model interaction between regression parameters and probabilities of inclusion by a combination of fixed number of component distributions. Mixtures of Gaussian distributions are often able to capture a large variety of skewed and overdispersed distributional forms (*McLachlan and Peel, 2000*), either directly for continuous data, or at a latent variable scale for binary, multinomial, or ordinal data. Dirichlet Process Mixture models (*Blackwell and MacQueen, 1973; MacEachern, 1994*) loosen the assumption of predetermined number of mixture components, and adapt a convenient mechanism to add or remove components from the model. Dunson, Pillai, and Park (2007) proposed a weighted Dirichlet Process Mixture Model, expanding the field to regression models, and adding extra flexibility by assigning weights to locations in population domain, which, in our application, create a natural link to weights from the complex sample design.

Quantile regression, although dating back to 1760 by Roger Boscovich (*Stigler, 1984*), has recently come into more widespread use in statistics (*Koenker, 2004*). In-

stead of modeling the conditional mean of an outcome as a linear function of given covariate values, quantile regression extends the concept to estimate the conditional median or various quantiles of outcome as linear functions of covariates, allowing for more detailed and robust inference, especially for situations in which common assumptions for linear regression are violated. Yu K and Moyeed (2001) successfully embedded quantile regression in Bayesian setting by an analogue of asymmetric Laplace distribution with non-informative prior, making it possible to tackle quantile regression problem with MCMC methods. While quantile regression is commonly used with population survey data, methodological exploration in the complex sample setting has been somewhat limited, and little if any work has explored the effect of weight trimming in the quantile regression setting.

In this study we develop a Weighted Dirichlet Process Mixture Model (WDPM model) to estimate quantity of interest from complex survey design data, in order to build data-driven inference that captures a wide variety of normal and non-normal distributions in a fashion that is sensitive to unequal probability of selection aspects of the sample design, but also offers increased efficiency when data permit. In addition to the linear regression setting, we also consider applications to quantile regression. Because the WDPM models are highly flexible and can generate predictive distributions that are accurate in tails of the distribution, they are a natural choice to consider for model-based methods to obtain population quantile regression estimates. The chapter is organized as follows. In Section 2 we review the theory of Bayesian finite population inference, quantile regression, finite mixture models, and Dirichlet Process mixture models. Section 3 reviews the Weighted Dirichlet Process Mixture model, and extends the WDPM model to incorporate survey weights. Section 4 provides a simulation study, and compares bias, coverage and RMSE of the proposed method with standard methods, under both linear model and quantile regression settings. Section 5 demonstrates the method's performance by applications on Dioxin

Dataset from NHANES. Section 6 provides a summary discussion.

4.2 Background Methodology

4.2.1 Bayesian Finite Population Inference

To introduce Bayesian finite population inference, we denote the sample design variables (selection probabilities, cluster indicator, stratum variables) by Z , and the population data Y is modeled as $Y \sim f(Y|\theta, Z)$. The distribution f could be either highly parametric, with a low dimension θ , or semi-parametric or “non-parametric” model with a high-dimension θ . An example of the former would be a normal model with common mean and variance, while an example of the latter would be a spline or Dirichlet Process Model. Let N be the number of elements in the population, Y_{obs} consist of the n observed data elements, and Y_{nob} consist of the $N - n$ unobserved cases in the population. Considering Y_{nob} as missing data, their posterior predictive distribution is given by:

$$p(Y_{nob} | Y_{obs}, I, Z) = \frac{\int \int p(Y_{nob} | Y_{obs}, Z, \theta, \phi) p(I | Y, Z, \theta, \phi) p(Y_{obs} | Z, \theta) p(\theta, \phi) d\theta d\phi}{\int \int \int p(Y_{nob} | Y_{obs}, Z, \theta, \phi) p(I | Y, Z, \theta, \phi) p(Y_{obs} | Z, \theta) p(\theta, \phi) d\theta d\phi dY_{nob}}$$
(4.1)

where ϕ models the inclusion indicator I . If ϕ and θ have independent priors, and the sampling design is ignorable, that is, I only depends on $Y_{obs}|Z$, the formula of predictive posterior distribution reduces to

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

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*). This approach can be extended to build inferences on a function

of the population data, say $Q(Y)$.

4.2.2 Quantile Regression

Quantile regression is a general class of linear models that focuses on estimating either the median or other quantiles of the response variable conditional on covariates. Consider a real valued random variable Y with cumulative distribution function $F_Y(y) = P(Y \leq y)$. Then for any $\tau \in [0, 1]$, the τ -th quantile of Y is defined by:

$$Q_Y(\tau) = F_Y^{-1}(\tau) = \inf\{y : F_Y(y) \geq \tau\}.$$

Thus $Q_Y(1/2)$ refers to median, $Q_Y(1/4)$ refers to first quartile (25th percentile), and so forth. The quantile function provides a complete characterization of Y with various values of τ . To solve for the τ -th quantile numerically, we define the piecewise linear loss function

$$\rho_\tau(y) = y(\tau - I(y < 0))$$

where I equals one if $y < 0$ is satisfied, and zero otherwise. The τ -th quantile of Y , namely u , is calculated by minimizing the expected loss of $Y - u$

$$\min_u E(\rho_\tau(Y - u)) = \min_u (\tau - 1) \int_{-\infty}^u (y - u) dF_Y(y) + \tau \int_u^{\infty} (y - u) dF_Y(y).$$

Thus

$$\hat{u} = \underset{u \in R}{\operatorname{argmin}} E(\rho_\tau(Y - u))$$

Assuming a random sample of Y , y_i , $i = 1, \dots, n$, the sample analogue of τ th-quantile is attained by solving the following minimization problem:

$$\hat{u} = \underset{u \in R}{\operatorname{argmin}} \sum_{i=1}^n (\rho_\tau(y_i - u))$$

Now we extend to regression setting. Let x_i , $i = 1, \dots, n$ be a $p \times 1$ vector of regressors.

The τ -th conditional quantile function is then given by $Q_{Y_i|X_i}(\tau) = X_i^T \beta_\tau$, and one

can obtain β_τ by solving:

$$\hat{\beta}_\tau = \underset{\beta \in R^p}{\operatorname{argmin}} E(\rho_\tau(Y_i - X_i\beta)).$$

The sample analogue:

$$\hat{\beta}_\tau = \underset{\beta \in R^p}{\operatorname{argmin}} \sum_{i=1}^n \rho_\tau(y_i - x_i\beta) \quad (4.2)$$

is usually solved by the simplex method (*Murty*, 1983).

Yu and Moyeed (2001) suggest a likelihood form based on the asymmetric Laplace distribution:

$$f_\tau(u) = \tau(1 - \tau) \exp\{-\rho_\tau(u)\}$$

where $\rho_\tau(u)$ has the same form of the loss function stated above. Thus the likelihood function could be written as:

$$L(y|\beta) = \tau^n(1 - \tau)^n \exp\left\{-\sum_i \rho_\tau(y_i - x_i^T \beta)\right\}.$$

Differentiating $L(y|\beta)$ yields the objective function given in (4.2).

Finally, to apply quantile regression in unequal probability of selection survey designs, we introduce a weighted likelihood. That is, given survey weight w_i , $i = 1, \dots, n$, β_τ is solved from the following pseudo-likelihood:

$$L_w(y|\beta) = \tau^{\tilde{N}}(1 - \tau)^{\tilde{N}} \exp\left\{-\sum_i w_i \rho_\tau(y_i - x_i^T \beta)\right\}$$

where $\tilde{N} = \sum_i w_i$.

4.2.3 Finite Gaussian Mixture Model

The general form of Finite Gaussian Mixture Model is:

$$\begin{aligned}
 Y_i|C_i = c, \mu_c, \sigma_c^2 &\sim N(\mu_c, \sigma_c^2) \quad c = 1, \dots, K \\
 C_i = c|p_{i1}, \dots, p_{ik} &\sim MULTI(1; p_{i1}, \dots, p_{ik}) \\
 \log\left(\frac{p_{ic}}{p_{i1}}\right) &= g(\alpha_c, w_i, z_i, x_i)
 \end{aligned}$$

where C_i is the class membership identifying the mixture component for each observation. The probability of class membership is assumed to be a function of the sampling weight w_i , and possibly other observed covariates x_i as well.

This approach can easily be adapted in regression models, applying sampling weights w_i in a cumulative logit model form:

$$\begin{aligned}
 Y_i|C_i = c, \mu_c, \sigma_c^2 &\sim N(\mu_c, \sigma_c^2) \quad c = 1, \dots, K \\
 C_i = c|p_{i1}, \dots, p_{ik} &\sim MULTI(1; p_{i1}, \dots, p_{ik}) \\
 \eta_{ij} = \Phi(\gamma_j - f(w_i, \alpha)), & \text{ for } \eta_{ij} = \sum_{k=1}^j p_{ik}, \quad j = 1, \dots, K - 1.
 \end{aligned}$$

Given a sufficiently large K , a finite Gaussian mixture model can maintain robustness in the presence of regression model misspecification, as well as skewness and overdispersion in the residual error term. Yet when the data permit, fitting a simpler model with a small value of K could lead to increased efficiency. In addition, $f(w_i, \alpha)$ could take the form of a simple parametric model (e.g., linear in w_i), or semi-parametric (e.g., linear P-splines) for further flexibility.

4.2.4 Dirichlet Process Mixture Model

The Finite Gaussian Mixture Model can be written in a more general form of

$$f(y_i|x_i) = \int N(y_i|\phi_i)G_{x_i}(\phi_i)$$

where G_{x_i} is multinomial. An alternative approach defines G_x as an element in an uncountable collection of probability measures $G_x \sim DP(\alpha G_0)$, where DP denotes a Dirichlet Process (*Ferguson, 1973*) centered at base measure G_0 with precision α . This leads to standard DP mixture models (*MacEachern, 1994*) and avoids explicitly specifying number of components K in advance.

Expressing the Dirichlet process in "stick-breaking" form leads to:

$$G = \sum_{h=1}^{\infty} p_h \delta_{\theta_h}, \quad \frac{p_h}{\prod_{l=1}^{h-1} p_l} \sim \text{BETA}(1, \alpha)$$

where δ_{θ_h} degenerate at θ , and $\{\theta_h\}$ are atoms generated from G_0 . Use of a Polya urn scheme (*Blackwell and MacQueen, 1973*) integrates out the infinite dimensional G and provides an easier form for simulation:

$$\phi_i | \phi^{(i)}, \alpha \sim \frac{\alpha}{\alpha + n - 1} G_0 + \frac{1}{\alpha + n - 1} \sum_{j \neq i} \delta_{\phi_j}.$$

That is, a new draw of observation i could be from the same component as a existing observation with probability $1/(\alpha + n - 1)$, or initiate a new draw from base measure G_0 with probability $\alpha/(\alpha + n - 1)$.

The drawback of standard DP mixture emerges in regression setting where the posterior predictive distribution of $y_i | x_i$ at a given draw of $\phi_h = (\beta_h, \sigma_h^2)$ is generated from $\frac{\alpha}{\alpha+n} N(y_i | x_i, \beta_0, \sigma_0^2) + \sum_{h=1}^K \frac{n_h}{\alpha+n} N(y_i | x_i, \beta_h, \sigma_h^2)$, K and S denoting the configuration of ϕ into K distinct values, n_h as the number of observations assigned to component h , and β_0 and σ_0^2 as further independent draws from G_0 . The conditional posterior predictive distribution of y takes a linear form of x :

$$E(y_i^{rep} | x_i, \beta_0, \dots, \beta_K, \sigma_0^2, \dots, \sigma_K^2, K, S) = \sum_{h=1}^K \frac{n_k}{\alpha + n} x_i \beta_h = x_i \bar{\beta}$$

where $\bar{\beta} = \sum_{h=1}^K \frac{n_k}{\alpha+n} \beta_h$. This restricts the model's ability to capture non-linear patterns in data.

4.3 Weighted Dirichlet Process Mixture Model

The weighted Dirichlet Process Mixture Model (Dunson et al. 2007) is a more flexible extension of standard DP mixture model, that allows the DP prior itself to be a mixture of DP random basis measures.

Assuming $G_{X_j}^* \sim DP(\alpha G_0)$, $j = 1, \dots, n$ random basis measures at each distinct covariate value, the actual DP prior is built as a mixture model of $G_{X_j}^*$:

$$G_{\mathbf{x}} = \sum_{j=1}^n b_j(\mathbf{x}) G_{\mathbf{x}_j}^*$$

$$b_j(\mathbf{x}) = \frac{\gamma_j \exp(-\psi \|\mathbf{x} - \mathbf{x}_j\|)}{\sum_l^n \gamma_l \exp(-\psi \|\mathbf{x} - \mathbf{x}_l\|)}$$

The form of b_x grants high weights in G_x to subjects with x_j closer to x , encourage clustering of subjects that are near to each other in the covariate space. Weight γ is designed to add extra “weight” at specific locations where data tend to make a larger impact. The smoothing parameter ψ is included to control the degree to which G_x loads across multiple draws from $DP(\alpha G_0)$. Note that the standard DP mixture model is a special case of the weighted DP mixture model, where $b_j(\mathbf{x}) = 1/n$ for all j .

To obtain values of γ_j and ψ in a data-driven manner, ψ is assigned a truncated log-normal hyperprior, $\psi \sim \log -N(\mu_\psi, \sigma_\psi^2)$, $\psi \in (0, 5)$. The choice for weight γ is more subtle to avoid either single dominating weight or uniformly distributed weights equivalent to a standard DP mixture model. Here we consider the hyperprior $\gamma_j \sim \text{gamma}(\alpha_\gamma, \beta_\gamma)$ which favors a few dominant locations.

To complete the Bayesian specification of the linear regression model for $N(y_i | \phi_i)$

with $\phi_i = (\beta_i, \sigma_i^2)$ and $\beta_i = (\beta_{i1} \dots \beta_{ip})$, we assume:

$$\beta_i \sim N(\beta, \sigma_i^2 \Sigma_\beta)$$

$$\tau_i = \sigma_i^{-2} \sim \text{Gamma}(a_\tau, b_\tau)$$

$$\beta \sim N(\beta_0, V_{\beta_0})$$

$$\Sigma_\beta^{-1} \sim \text{Wishart}((v_0 \Sigma_0)^{-1}, v_0)$$

$$b_\tau \sim \text{Gamma}(a_0, b_0)$$

To make inference on quantities of interest, e.g., the population regression parameters, we apply the data augmentation method. First, using the analytical form of all conditional probabilities outlined in Dunson et al. (2007), we obtain draws from the posterior distribution of parameters from the WDPM model using Gibbs Sampler. We then obtain a draw of y_i^{rep} , the posterior predictive distribution of y_i at x_i conditional on a draw of all other parameters as

$$w_{i0} N(y_i^{rep}; x_i^T \beta_0, \sigma_0^2 + x_i^T \Sigma_\beta x_i) + \sum_{h=1}^K w_{ih} N(y_i^{rep}; x_i^T \beta_h, \sigma_h^2)$$

where $w_{i0}(x_i) = \frac{\alpha b_j(x_i)}{\alpha + \sum_{l \neq i} I(C_{s_i} = j)}$, $w_{ih}(x_i) = \frac{b_{C_h}(x_i) \sum_{m \neq i} I(S_m = h)}{\alpha + \sum_{l \neq i} I(C_{s_i} = C_h)}$ for $S = (S_1, \dots, S_n)$ mapping n subjects into K distinct clusters and $C(C_1, \dots, C_K)$ denoting the K cluster themselves.

A draw from the posterior distribution of the population regression slope

is then obtained from a multivariate normal with mean $(X^T W^* X)^{-1} X^T W^* \tilde{y}$ with $\tilde{y} = \sum_{h=0}^K w_{ih} x_i^T \beta_h$, and variance $(X^T W^* X)^{-1} \left\{ \sum_l w_i^* x_i \left(\sum_{h=0}^K w_{ih} \sigma_h^2 + w_{i0}^2 x_i^T \Sigma_\beta x_i \right) x_i^T \right\} (X^T W^* X)^{-1}$,

where w_i^* are the sample weights and W^* is a diagonal matrix of the w_i^* . If the sam-

pling fraction is non-trivial, the predictive regression slope can be obtained by joining

the predicted and observed outcomes in the finite mixture setting. The detailed Gibbs

Sampler steps are in Appendix C. Efficiency is gained when K is small, leading to a

linear prediction unaffected by survey weights. Large values of K can accommodate

non-linearities that lead to bias if the survey weights are ignored.

4.4 Simulation Study

In this section we evaluate the application of Weighted Dirichlet Process Mixture Model in complex survey design in two scenarios: ordinary linear regression and quantile regression settings. For each setting, the target of interest is the population slope. The competing methods are the unweighted estimator, the fully-weighted estimator and weight trimming estimator with cutpoint set at three times standard deviation. Biasness, relative root of mean square error (RMSE) and coverage of 95% confidence or credible intervals are calculated to assess the performance.

4.4.1 Weighted Dirichlet Mixture Model for Ordinary Linear Regression

For ordinal linear regression setting, a population of $N = 20,000$ is generated. The predictor X is evenly distributed into 10 strata on the interval from 0 to 10, each with a range of 1. And within each stratum X is uniformly distributed. Then the response variable Y is created from a spline function of X , with cutpoints between strata as knots. Three sets of coefficients are considered to represent different models setting of $Y | X$, including convex curve, concave curve and straight slope:

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

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

$$\beta_a = c(0, 0, 0, 0, .5, .5, 1, 1, 2, 2, 4)$$

$$\beta_b = c(0, 11, -4, -2, -2, -1, -1, -0.5, -0.5, 0, 0)$$

$$\beta_c = c(0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0).$$

Under β_a and β_b , linear model assumption is misspecified, and weights are necessary to correct the corresponding biasness. Under β_c , the model is correctly specified, and it would be most efficient to ignore the sample weight.

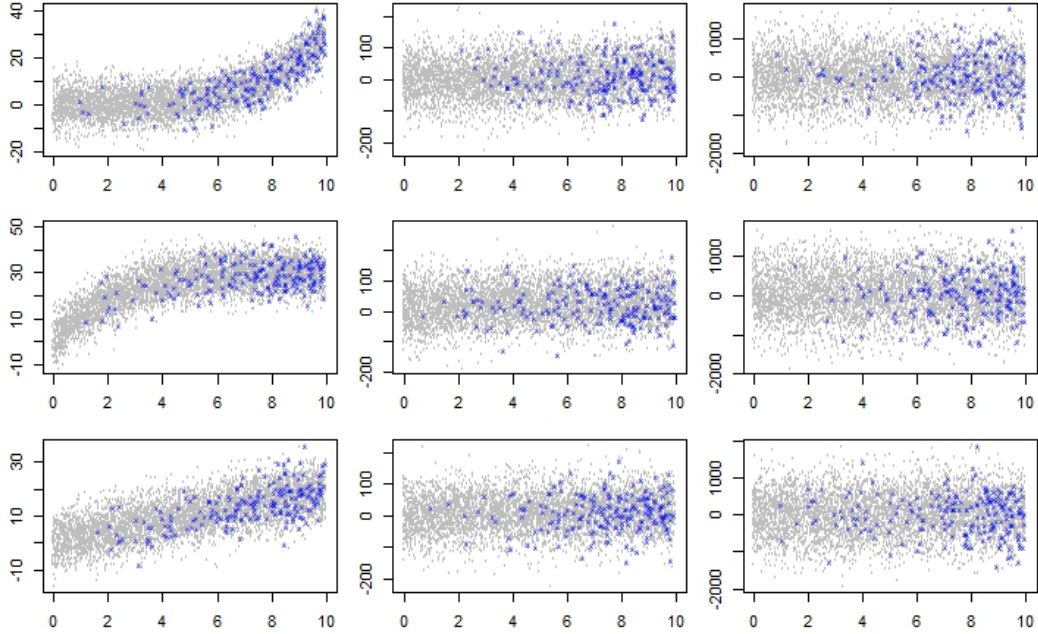


Figure 4.1: Scatter plot of generated population vs covariate. Population are marked grey in background and sample in blue cross. From left to right, $\sigma^2 = 10^{1.5}, 10^{3.5}, 10^{5.5}$; from top to bottom, setting $\beta_a, \beta_b, \beta_c$.

Population variance σ^2 varies among $10^{1.5}$, $10^{3.5}$ and $10^{5.5}$, creating varying level of variance influence from revealing a moderate curving pattern, to completely overwhelming the local difference in the population slope. Also note that slope in setting β_a changes more dramatically where data are most densely sampled, while the reverse happens in β_b , suggesting that a complex model is needed to correctly capture the two different scenarios.

Altogether 200 samples are repeatedly drawn from the population. Each sample consists of 200 observations that are sampled with selection probabilities proportional to $\pi_i = (1+i)*i$ for i th stratum. This guarantees that the maximum weight is about 55 times than the minimum weight. Different settings of populations and samples are presented in Figure 4.1.

The target quantity of interest is $B = (\sum_{i=1}^N \tilde{X}_i \tilde{X}_i')^{-1} \sum_{i=1}^N \tilde{X}_i Y_i$ for $\tilde{X}_i = (1 \ X_i)'$, the least-squares linear approximation of the population slope.

The hyperprior parameters are pre-specified as follows. For the prior on weight functions, we let $a_\gamma = 0.01$, $b_\gamma = 2$, and $\xi = 0.01$. For hyper-priors on basis distribution parameters, we have $\beta_0 = 0$, $V_{\beta_0} = 1000 \times I_p$, $v_0 = 1$, $\Sigma_0 = I_p$, $a_\tau = 0.1$, $a_0 = 0.1$, $b_0 = 0.1$, $\mu_\psi = \log(30)$ and $\sigma_\psi = 0.5$. In application, we also restrict the value of ψ in a range from zero to five.

A Gibbs Sampler as previously described is applied, that is, the new distribution of S , DP weights, number of atom distributions and parameters within each atom are drawn sequentially from the full conditional distributions. The unobserved Y are then generated by re-creating the missing population from mean of all components, and the estimate target population slope is obtained by fully weighted regression on observed and predicted Y . All slope and variance parameters for different components in the Gibbs sample are initialized using estimated global beta coefficients and variance to achieve faster convergence. The first 5000 iterations are dropped as burn-in, and the following 10000 iterations are kept to form the distribution of estimated parameters. Diagnostic plots are generated to assure the algorithm's convergence. The process repeats for all 200 samples to provide the empirical distribution for the repeated measures properties.

We compare the properties of our Weighted Dirichlet Process Mixture model (WDPM) with major competitors, including the unweighted model (UNWT), fully weighted model (FWT) and a standard ad-hoc weight trimming method with threshold at three times standard deviation of the weights(WT3). Biases and nominal 95% coverages are recorded directly, while RMSEs are rescaled according to fully weighted estimator. Results are provided in Table 4.1, 4.2, and 4.3.

For the first two scenarios, the model is misspecified as linear, and the unweighted method tends to be biased, leads to an overall larger RMSE and lower coverage rate comparing to fully weighted model. However, as residual variance increases, the gain in efficiency gradually overcomes the loss in accuracy, and at large variance level,

	$\sigma^2 = 10^{1.5}$			$\sigma^2 = 10^{3.5}$			$\sigma^2 = 10^{5.5}$		
	Bias	RMSE	cover	Bias	RMSE	cover	Bias	RMSE	cover
UNWT	1.184	2.499	0	1.120	0.692	0.87	0.475	0.599	0.93
FWT	-0.355	1	0.72	-0.297	1	0.92	0.289	1	0.91
WT3	-0.024	0.531	0.98	-0.039	0.748	0.94	-0.185	0.751	0.94
WDPM	-0.297	0.702	0.98	-0.258	0.490	0.99	-0.723	0.431	0.92

Table 4.1: Comparison of various estimators of slope B_1 under β_a linear spline setting. Bias and relative RMSE under populations with residual variance $10^{1.5}$, $10^{3.5}$ and $10^{5.5}$ from following model: Unweighted, Fully Weighted, Weight Trimming and weighted Dirichlet Process Mixture Model.

	$\sigma^2 = 10^{1.5}$			$\sigma^2 = 10^{3.5}$			$\sigma^2 = 10^{5.5}$		
	Bias	RMSE	cover	Bias	RMSE	cover	Bias	RMSE	cover
UNWT	-0.973	1.464	0	-1.047	0.665	0.87	-1.692	0.600	0.93
FWT	0.559	1	0.67	0.629	1	0.90	1.214	1	0.91
WT3	0.085	0.434	0.99	0.070	0.736	0.92	-0.076	0.750	0.94
WDPM	0.203	0.532	1	0.205	0.326	0.98	-0.601	0.418	0.90

Table 4.2: Comparison of various estimators of slope B_1 under β_b linear spline setting. Bias and relative RMSE under populations with residual variance $10^{1.5}$, $10^{3.5}$ and $10^{5.5}$ from following model: Unweighted, Fully Weighted, Weight Trimming and weighted Dirichlet Process Mixture Model.

	$\sigma^2 = 10^{1.5}$			$\sigma^2 = 10^{3.5}$			$\sigma^2 = 10^{5.5}$		
	Bias	RMSE	cover	Bias	RMSE	cover	Bias	RMSE	cover
UNWT	-0.007	0.599	0.93	-0.072	0.599	0.93	-0.717	0.599	0.93
FWT	0.007	1	0.91	0.065	1	0.91	0.650	1	0.91
WT3	-0.002	0.751	0.94	-0.016	0.751	0.94	-0.162	0.751	0.94
WDPM	-0.035	0.503	1	-0.128	0.456	1	-1.733	0.393	0.90

Table 4.3: Comparison of various estimators of slope B_1 under β_c linear spline setting. Bias and relative RMSE under populations with residual variance $10^{1.5}$, $10^{3.5}$ and $10^{5.5}$ from following model: Unweighted, Fully Weighted, Weight Trimming and weighted Dirichlet Process Mixture Model.

unweighted method results have better RMSEs comparing to fully weighted method, suggesting that the model misspecification could be ignored. The weight trimming method has an overall better performance comparing to fully weighted method, maintaining the necessary bias-correction while improving the efficiency and 95% coverage. However, the weighted Dirichlet Process Mixture estimator demonstrates a dominating performance across all settings, maintaining more than 50% reduction in RMSE comparing to fully weighted model. And when the residual variance is large, it leads to more efficient estimates than unweighted estimates.

Under β_c , where the model is correctly specified, all methods yield unbiased results. The biases for $\sigma^2 = 10^{5.5}$ are due to simulation error, enhancing the instability in the estimating. Here the WDPM method yields the maximum reduction in RMSE, reducing RMSE by 50-60% over the fully weighted method, while the unweighted estimator consistently reduces RMSE by about 40% over the fully weighted method, and weight trimming method reduces by 20%. The UNWT, FWT, and WT3 coverage rates are a little low due to the instability caused by small sample size. Meanwhile, the WDPM estimator provides conservative coverage across all residual variance settings.

4.4.2 Weighted Dirichlet Mixture Model for Quantile Regression

To assess the performance of the WDPM method in quantile regression, we consider heavy tailed, skewed, and bimodal distributions. As in the linear regression setting, a population of 20000 is generated, and samples of 200 are drawn from the population. All covariates X are uniformly distributed on interval of $(0, 10)$. Our inferential target is the linear population slope of X on first quartile (25th percentile), median and third quartile (75th percentile) of $Y : \underset{\beta_{0\tau}, \beta_{1\tau}}{\operatorname{argmin}} \sum_{i=1}^N p_{\tau}(y_i - \beta_{0\tau} - \beta_{1\tau}x_i)$, for $\tau = .25, .5, .75$. We estimate bias, RMSE, and coverage from 200 independent simulations.

For the long tail setting, we consider a non-central t distribution with five degree

of freedom, and selection probabilities π_i related to covariate X :

$$\begin{aligned} X_i &\sim \text{Uniform}(0, 10) \\ Y_i|X_i &\sim T(\mu = x_i, df = 5) \\ \pi_i &\propto (1 + \lceil x_i \rceil) \times \lceil x_i \rceil \\ i &= 1, \dots, N = 20,000. \end{aligned}$$

For skewed distributed setting, we consider a Gamma distribution, and selection probabilities related to covariate X :

$$\begin{aligned} X_i &\sim \text{Uniform}(0, 10) \\ Y_i|X_i &\sim \text{Gamma}(k = x_i^{1.5}/5 + 1, \theta = 2) \\ \pi_i &\propto (1 + \lceil x_i \rceil) \times \lceil x_i \rceil \\ i &= 1, \dots, N = 20,000. \end{aligned}$$

For bimodal distribution, we consider the following mixture with weight α_i related to x_i :

$$\begin{aligned} X_i &\sim \text{Uniform}(0, 10) \\ Y_i|X_i, \alpha_i &\sim \alpha_i N(x_i, 16) + (1 - \alpha_i) N(-5, 16) \\ \alpha_i &\sim \text{Bernoulli}(x_i/10) \\ \pi_i &\propto (1 + \lceil x_i \rceil) \times \lceil x_i \rceil \\ i &= 1, \dots, N = 20,000. \end{aligned}$$

Under first scenario, the linear model is correctly specified, with an over-dispersed residual. Thus we expect all estimates to be unbiased, with unweighted method

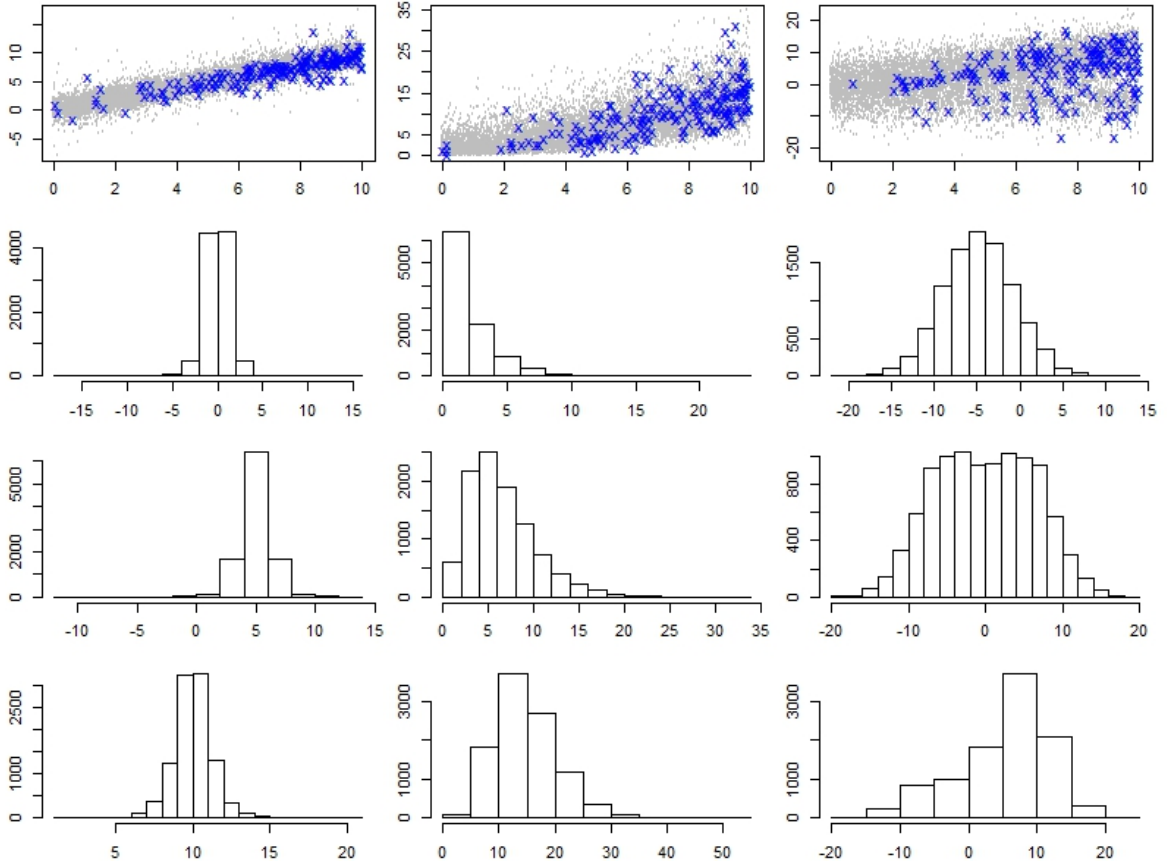


Figure 4.2: Plot of quantile regression simulation settings. Population as grey in background and sample in blue cross. From left to right: t distribution, exponential distribution and bimodal distribution; From top to bottom: Scatter plot, Histogram of Y at X=0, 5, and 10.

gaining efficiency, and WDPM model correcting the coverage rate. For the other two scenarios, biases from unweighted estimate are introduced due to non-linearity in x_i combined with sampling probabilities that are a function of x_i , and we expect WDPM estimates obtain similar performance as fully weighted estimates. Detailed population settings and samples are also demonstrated in Figure 4.2.

The hyperprior parameters are pre-specified as in the linear regression setting: $a_\gamma = 0.01$, $b_\gamma = 2$, and $\xi = 0.01$. For hyper-priors on basis distribution parameters, we have $\beta_0 = 0$, $V_{\beta_0} = 1000 \times I_p$, $v_0 = 1$, $\Sigma_0 = I_p$, $a_\tau = 0.1$, $a_0 = 0.1$, $b_0 = 0.1$, $\mu_\psi = \log(30)$ and $\sigma_\psi = 0.5$.

	First Quartile			Median			Third Quartile		
	Bias	RMSE	cover	Bias	RMSE	cover	Bias	RMSE	cover
UNWT	-0.002	0.504	0.98	0.005	0.502	0.96	-0.009	0.469	0.98
FWT	-0.022	1	0.91	-0.007	1	0.93	0.001	1	0.91
WT3	-0.012	0.718	0.99	-0.003	0.692	1	-0.007	0.656	1
WDPM	0.016	0.510	0.98	0.004	0.495	0.98	-0.008	0.455	0.98

Table 4.4: Comparison across various estimators of slope B_1 under non central T distribution. Bias and relative RMSE of estimates for 1st quartile, median and 3rd quartile of the outcome from following model: Unweighted, Fully Weighted, Weight Trimming and weighted Dirichlet Process Mixture Model.

	First Quartile			Median			Third Quartile		
	Bias	RMSE	cover	Bias	RMSE	cover	Bias	RMSE	cover
UNWT	0.235	1.971	0.53	0.202	1.315	0.71	0.212	1.052	0.81
FWT	0.010	1	0.94	0.006	1	0.97	0.050	1	0.93
WT3	0.092	1.100	0.78	0.082	0.887	0.92	0.092	0.855	0.96
WDPM	0.154	1.299	0.82	0.078	0.665	0.93	0.061	0.555	0.93

Table 4.5: Comparison across various estimators of slope B_1 under Gamma distribution. Bias and relative RMSE of estimates for 1st quartile, median and 3rd quartile of the outcome from following model: Unweighted, Fully Weighted, Weight Trimming and weighted Dirichlet Process Mixture Model.

Similar to the simulation with ordinary linear regression, the Gibbs sampler is used to generate the posterior predictive distribution of the population. After generating the population from predictive posterior distribution, quantile regression is used to obtain the population slope of x_i on the median, first and third quartile of the outcome. Within each simulation there are 15000 iterations, with the first 5000 dropped as burn-in. Diagnostic plots are generated to assure the algorithm's convergence. Results are provided in Table 4.4, 4.5 and 4.6.

For the population created from a long-tail T distribution, the unweighted method has the best performances across all quantile estimates, prevailing in both efficiency and coverage, since bias-reduction from weighting could be ignored in this scenario. The fully weighted method has somewhat reduced coverage due to the instability caused by the small sample size. And weight trimming method shows major improve-

	First Quartile			Median			Third Quartile		
	Bias	RMSE	cover	Bias	RMSE	cover	Bias	RMSE	cover
UNWT	0.638	1.545	0.41	0.248	0.875	0.87	0.009	0.538	0.96
FWT	-0.038	1	0.94	0.030	1	0.96	0.051	1	0.92
WT3	0.203	0.943	0.73	0.128	0.764	0.97	0.065	0.724	0.99
WDPM	-0.098	0.652	0.97	-0.131	0.626	0.97	-0.148	0.540	0.94

Table 4.6: Comparison across various estimators of slope B_1 under Bimodal distribution. Bias and relative RMSE of estimates for 1st quartile, median and 3rd quartile of the outcome from following model: Unweighted, Fully Weighted, Weight Trimming and weighted Dirichlet Process Mixture Model.

ment RMSE and obtains conservative coverage rate. Meanwhile, the WDPM method maintains stable results across median, first and third quartiles, provides about 45% reductions in RMSEs comparing to fully weighted method or 20% comparin to weight trimming method, with conservative coverage rates.

In second scenario, where skewed population distribution and model misspecification both occur, weighting becomes necessary, and the fully weighted method has better performances with respect to bias at lower quartiles, as anticipated. The unweighted method fails to correct the biasness, and has poor coverage and larger RMSE for all except 3rd quartile, where the impact of biases is offset by the reduction in variance. The weight trimming method has improved performance in median and third quartile comparing to fully weighted method, but suffers minor drop in coverage rate in first quartile. The WDPM model has smaller RMSE than any other model for the 50th and 75th percentile, with approximately correct coverage. For the 25th percentile, the WDPM model suffers from some reduction in coverages plus increase in biases and RMSEs due to the model’s inability to completely capture the behavior of the low percentile of the outcome for small values of x due to extremely small sample size, see Figure 4.2.

In the bimodal setting, bias reduction is required for estimation in first quartile and median, but not in third quartile, since it closely approximates a pure linear

model. The fully weighted method successfully reduces biases and performs better with respect to RMSE than unweighted one in first quartile and median, but loses in efficiency in third quartile. Weight trimming method acts as a upgraded version of fully weighted method, showing better result in all but the coverage of first quartile. The WDPM model provides a large improvement comparing to fully weighted model in first quartile and median, reducing RMSE by about 30% to 40% RMSE. It also maintains the improvement even comparing to weight trimming method. When the unweighted model has better RMSE in third quartile, WDPM model closely follows its performance. Both fully weighted method and WDPM model have satisfactory coverage rates.

4.4.3 Weighted Dirichlet Mixture Model for Quantile Regression with Binary Covariate

In this subsection, we conduct a simulation study expanding the application of the weighted Dirichlet Mixture Model to quantile regression with a binary covariate. To be more specific, we focus on the bimodal population setting, assessing performance differences between unweighted quantile regression, weighted quantile regression, weight trimming estimate and WDPM model, to help in understanding the result from application on Dioxin data in the next section.

	First Quartile			Median			Third Quartile		
	Bias	RMSE	cover	Bias	RMSE	cover	Bias	RMSE	cover
UNWT	0.046	0.978	0.98	-0.012	0.968	0.98	0.004	1.023	0.94
FWT	0.049	1	0.98	0.012	1	0.98	0.007	1	0.96
WT3	0.044	0.995	1	0.001	0.999	1	0.013	0.987	0.98
WDPM	-0.052	0.608	0.98	0.038	0.117	1	0.073	0.703	0.96

Table 4.7: Comparison across various estimators of slope B_1 under Bimodal distribution with Binary covariates. Bias and RMSE of estimates for 1st quartile, median and 3rd quartile of the outcome from following model: Unweighted, Fully Weighted, Weight trimming by 3 SD and weighted Dirichlet Process Mixture Model.

The bimodal distributed population is created as follows:

$$X_i \sim \text{Bernoulli}(0.5)$$

$$\alpha_i \sim \text{Bernoulli}(0.5)$$

$$Y_i|X_i, \alpha_i \sim N(0.5 * X_i + 5 * \alpha_i, sd = 1)$$

$$\pi_i \propto 15 * X_i + N(0, 1) + 7$$

$$i = 1, \dots, N = 20,000.$$

Here π usually ranges from 2.5 to 26, resulting in a ratio of approximately 9 between the maximum and minimum selection probabilities. Fifty samples with size of 200 are drawn from the population with various probabilities of selection defined above. Simulations on each sample consist of 10000 iterations, with first 5000 dropped as burn-in. Bias, RMSE and coverage are assessed with 50 independent simulations, and results are displayed in Table 4.7.

The population setting introduces no potential biasness through varies probabilities of selection. And the result suggests that all models provide consistent results with good coverage in first quartile and third quartile, while WDPM reduces the overall RMSEs by 30%. However, when estimating population slope for the median, the true RMSEs from unweighted, weighted and weight trimming models are greatly

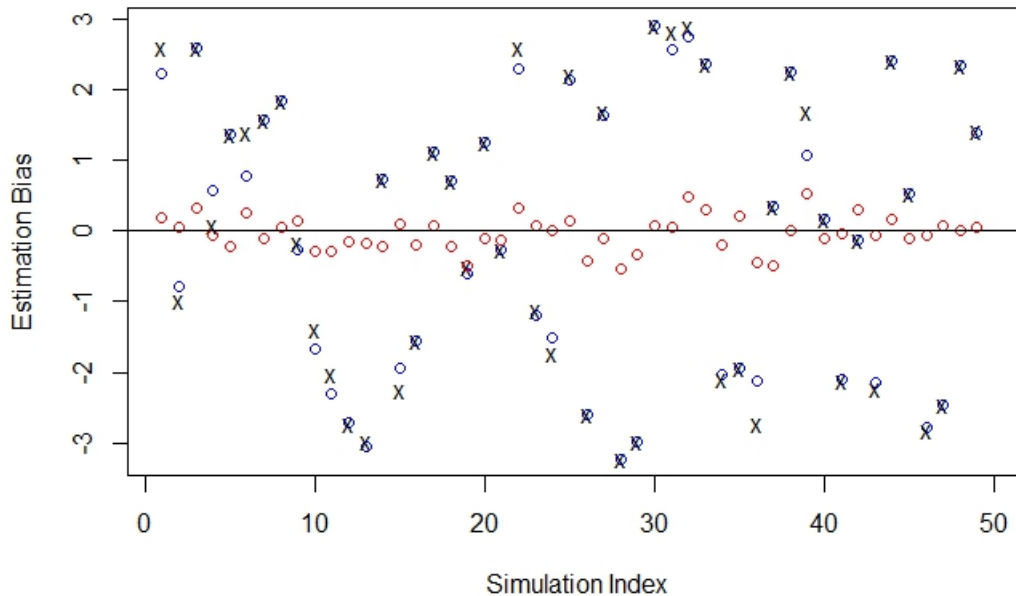


Figure 4.3: Figure of biases of estimated population slope for median from all three methods. Biasness of point estimates from each simulation are plotted sequentially. Blue circles mark the unweighted quantile regression, black crosses mark the fully weighted quantile regression, and red circles mark WDPM estimates.

increased, indicating unstable estimation.

To explore these results further, Figure 4.3 plots bias for all three approaches for fifty simulations. This suggests that under certain situation, both unweighted method and weighted method often provide similar estimates far away from true value. WDPM is more robust for those situations, providing stable estimates.

4.5 Application on Dioxin data from NHANES

We apply the weighted Dirichlet Process prior to an analysis relating age and gender to the blood level of dioxin in a representative sample of US adult. We use data from the 2003-2004 National Health and Nutrition Examination Survey (NHANES) dataset. The 2013-2014 NHANES is a multi-stage, unequal probability-of-selection

survey, consisting of 25 strata and 2 masked PSU per stratum. Survey weights adjust for oversample of race/ethnic minority older subjects, as well as non-response adjustments at both the first stage of recruitment and the second stage of biomarker collection. Both blood bio-markers and demographic information are collected and included in the model. The bio-marker we are interested in is 2,3,7,8-tetracholorodibenzo-p-dioxin (TCDD), usually a compound resulting from incomplete combustion in incineration, paper and plastics manufacturing and smoking. More than half TCDD readings are below limit of detection, and are imputed five times through multiple imputation described in *Chen et al. (2010b)*. A jackknife method is used to compute variances to fully account for sample design features, and Rubin’s formula (*Rubin, 1987*) is used to combine inferences from each of the multiply-imputed datasets.

4.5.1 Linear Regression Setting

We fit four linear regression models to assess the impact of age, gender, and their interactions on log transformed blood TCDD. Hyper-priors are set the same as in the simulation study, and unweighted, fully weighted and weighted Dirichlet Process Mixture estimates are compared in bias and RMSE, where the fully weighted version is treated as the "true" value in corresponding calculation. Note that there exists correlations between the weighted estimator and other estimators, with the true estimated square bias of regression coefficient $\hat{\beta}$ given by $max((\hat{\beta} - \hat{\beta}_w)^2) - \hat{V}_{01}$, where $\hat{V}_{01} = \hat{V}ar(\hat{\beta}) + \hat{V}ar(\hat{\beta}_w) - 2\hat{C}ov(\hat{\beta}, \hat{\beta}_w)$. And to fully account for all design feature, all variance/covariance estimates are calculated via jackknife as $\hat{V}ar(\hat{\beta}_w) = \sum_h \frac{k_h-1}{k_h} \sum_{i=1}^{k_h} (\hat{\beta}_{(hi)} - \hat{\beta})^2$, where $\hat{\beta}_{(hi)}$ denotes the β estimator by excluding the i_{th} MVU in h_{th} stratum, and the case weights utilized in the fully-weighted and WDPM analysis

Model	Est(10^{-2})	Bias(10^{-3})	RMSE(10^{-3})	95%CI(10^{-2})
UNWT	3.309	-1.262	3.265	(2.827,3.790)
FWT	3.435	0	3.888	(2.669,4.201)
WT3	3.435	-0.002	5.495	(2.669,4.2)
WDPM	3.351	-0.835	0.658	(3.222,3.481)

Table 4.8: Regression of log TCDD on Age. Bias, RMSE and 95% CI for linear slope estimated for age in unweighted, fully weighted, weight trimming by 3 SD and weighted Dirichlet Process Mixture models.

Model	Est(10^{-1})	Bias(10^{-1})	RMSE(10^{-1})	95%CI(10^{-1})
UNWT	1.539	-0.822	1.248	(0.028,3.051)
FWT	2.361	0	0.637	(1.107,3.616)
WT3	2.367	0.006	0.900	(1.113,3.621)
WDPM	0.674	1.248	1.565	(0.578,0.771)

Table 4.9: Regression of log TCDD on Gender. Bias, RMSE and 95% CI for linear slope estimated for gender in unweighted, fully weighted, weight trimming by 3 SD and weighted Dirichlet Process Mixture models.

are given by:

$$w_j^* = \begin{pmatrix} w_j & \text{if } j \notin h, j \in i \\ \frac{k_h-1}{k_h} & \text{if } j \in h, j \notin i \\ 0 & \text{if } j \in h, j \in i \end{pmatrix}$$

$\hat{Var}(\hat{\beta})$ and $\hat{Cov}(\hat{\beta}_w, \hat{\beta})$ are calculated accordingly, and estimates from five imputed replicate datasets are combined with Rubin's formula. The result in the WDPM estimate was based on 10000 iterations after discarding 2000 draws as burn-in. The resulting estimates, biases, RMSEs and 95% confidence intervals are summarized in the Tables 4.9-4.11, where the fully weighted estimator is treated as unbiased. Dioxin levels are positively associated with age plus being male, with the age association being stronger among males, although the interaction is significant only for the unweighted model.

In general, the survey weights have less impact on estimating the effect of age, but play a crucial role in estimating the effect of gender, thus usually unweighted

Age	Est(10^{-2})	Bias(10^{-3})	RMSE(10^{-3})	95%CI(10^{-2})
UNWT	3.356	-0.907	3.296	(2.860,3.851)
FWT	3.446	0	3.895	(2.679,4.214)
WT3	3.446	-0.001	5.505	(2.679,4.213)
WDPM	3.295	-1.520	0.993	(3.099,3.490)
Gender	Est(10^{-1})	Bias(10^{-1})	RMSE(10^{-1})	95%CI(10^{-1})
UNWT	2.556	0.016	0.902	(1.084,4.028)
FWT	2.540	0	0.616	(1.326,3.754)
WT3	2.546	0.006	0.87	(1.333,3.758)
WDPM	1.064	-1.476	1.373	(0.066,1.469)

Table 4.10: Regression of log TCDD on Age and Gender. Bias, RMSE and 95% CI for linear slope estimated for age and gender in unweighted, fully weighted, weight trimming by 3 SD and weighted Dirichlet Process Mixture models.

Age	Est(10^{-2})	Bias(10^{-4})	RMSE(10^{-3})	95%CI(10^{-2})
UNWT	2.842	-5.063	3.758	(2.261,3.423)
WT	2.893	0	2.661	(2.368,3.417)
WT3	2.893	-0.001	3.764	(2.368,3.417)
WDPM	2.847	-4.577	2.444	(2.365,3.328)
Gender	Est(10^{-1})	Bias(10^{-2})	RMSE(10^{-1})	95%CI(10^{-1})
UNWT	-1.946	2.882	1.591	(-5.081,1.189)
WT	-2.234	0	3.259	(-8.655,4.186)
WT3	-2.229	0.006	4.606	(-8.645,4.188)
WDPM	-1.482	7.526	2.758	(-6.915,3.952)
Interaction	Est(10^{-2})	Bias(10^{-3})	RMSE(10^{-3})	95%CI(10^{-2})
UNWT	0.963	-0.880	3.285	(0.316,1.610)
WT	1.051	0	7.335	(-0.394,2.496)
WT3	1.051	0	10.367	(-0.393,2.495)
WDPM	0.874	-1.769	7.093	(-0.523,2.271)

Table 4.11: Regression of log TCDD on Age, Gender and Interaction. Bias, RMSE and 95% CI for linear slope estimated for age, gender and interaction in unweighted, fully weighted, weight trimming by 3 SD and weighted Dirichlet Process Mixture models.

estimates have smaller RMSEs in estimated coefficient of age, and fully weighted and weight trimming estimates have smaller RMSEs in estimated coefficient of gender. Consequently the WDPM model has much better performance than the weighted and weight trimming, and even than the unweighted model, when estimating the effect of age on dioxin blood levels. The effect of gender appears to be biased toward the null, lead to larger RMSE increase than other models except in the joint age \times gender model.

4.5.2 Quantile Regression Setting

In this section we evaluate the performance of WDPM estimator in the quantile regression setting based on the same dioxin dataset from the NHANES study. We again focus on the impact of age on the first quartile, median and third quartile of log blood TCDD. While estimating bias and RMSE, results from weighted quantile regression are considered as unbiased, and jackknife and Rubin's formula are applied for complex survey scheme and multiple imputation. The result for the WDPM estimator is based on 10000 iterations after discarding 2000 draws as burn-in. The resulting biasness and RMSE are summarized in Tables 4.12-4.15. The scatter plot of the sample and linear slope for first quartile, median and third quartile of logTCDD by age are presented in Figure 4.4. The impact of age (older ages have higher TCDD level) and gender (males having higher TCDD levels) is greater at median levels than at the first and third quartiles, whereas interactions (stronger age association among males than females) are greater in the first and third quartile than in the median.

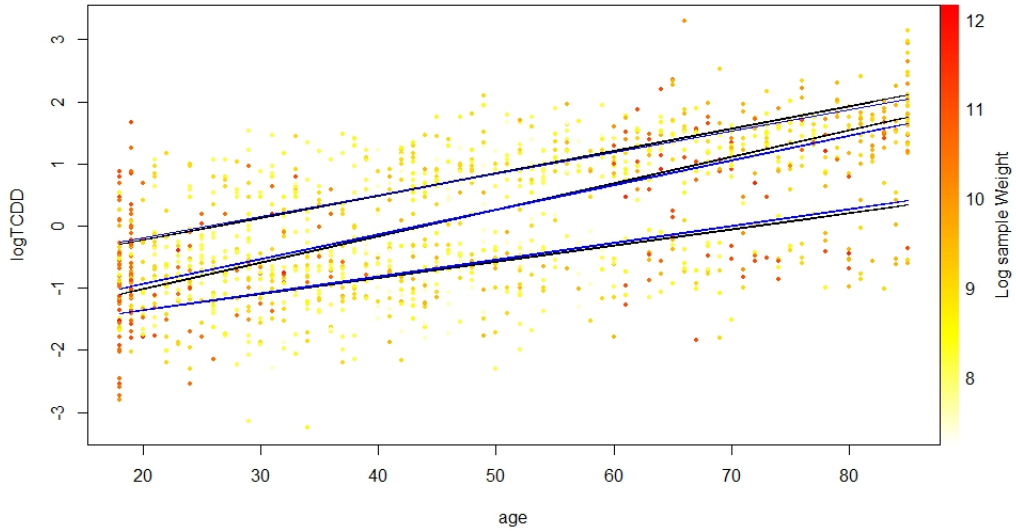


Figure 4.4: Quantile regression of log TCDD on age. Sample presented as background dots, color represents weighting in log form, and lines from top to bottom are linear slope for first quartile, median and third quartile, black for WDPM, blue for fully-weighted .

	First Quartile			
	Est(10^{-2})	Bias(10^{-3})	RMSE(10^{-3})	95% CI(10^{-2})
UNWT	2.780	7.123	3.507	(2.089,3.471)
FWT	2.708	0	6.762	(1.376,4.041)
WT3	2.709	0.080	9.557	(1.378,4.041)
WDPM	2.772	6.321	0.866	(2.601,2.942)
	Median			
	Est(10^{-2})	Bias(10^{-3})	RMSE(10^{-3})	95% CI(10^{-2})
UNWT	3.983	-2.758	5.489	(3.247,4.719)
FWT	4.259	0	4.702	(3.332,5.185)
WT3	4.258	-0.012	6.648	(3.331,5.184)
WDPM	4.157	-1.018	0.495	(4.059,4.255)
	Third Quartile			
	Est(10^{-2})	Bias(10^{-3})	RMSE(10^{-3})	95% CI(10^{-2})
UNWT	3.400	-1.276	3.428	(2.872,3.927)
FWT	3.527	0	3.967	(2.746,4.309)
WT3	3.526	-0.011	5.586	(2.747,4.305)
WDPM	3.546	0.188	0.537	(3.440,3.652)

Table 4.12: Quantile regression of log TCDD on age. Bias, RMSE and 95%CI for linear slope estimated for age in unweighted, fully weighted, weight trimming by 3 SD and weighted Dirichlet Process Mixture models.

	First Quartile			
	Est(10^{-1})	Bias(10^{-1})	RMSE(10^{-1})	95% CI(10^{-1})
UNWT	0.550	-0.599	2.384	(-3.131,4.231)
FWT	1.149	0	1.497	(-1.801,4.099)
WT3	1.154	0.005	2.121	(-1.803,4.11)
WDPM	0.147	-1.002	0.061	(0.026,0.267)
	Median			
	Est(10^{-1})	Bias(10^{-1})	RMSE(10^{-1})	95% CI(10^{-1})
UNWT	2.949	-0.470	2.372	(-1.094,6.991)
FWT	3.418	0	2.029	(-0.578,7.414)
WT3	3.424	0.006	2.870	(-0.578,7.426)
WDPM	0.884	-2.534	1.537	(0.728,1.040)
	Third Quartile			
	Est(10^{-1})	Bias(10^{-1})	RMSE(10^{-1})	95% CI(10^{-1})
UNWT	1.792	-0.865	1.082	(0.216,3.367)
FWT	2.657	0	0.798	(1.085,4.229)
WT3	2.657	0	1.092	(1.132,4.182)
WDPM	1.100	-1.557	1.339	(0.992,1.208)

Table 4.13: Quantile regression of log TCDD on gender. Bias, RMSE and 95% CI for linear slope estimated for gender in: unweighted, fully weighted, weight trimming by 3 SD and weighted Dirichlet Process Mixture model.

First Quartile				
Age	Est(10^{-2})	Bias(10^{-3})	RMSE(10^{-3})	95% CI(10^{-2})
UNWT	3.142	1.131	6.396	(1.882,4.402)
FWT	3.029	0	12.424	(0.582,5.477)
WT3	3.029	0.004	17.627	(0.574,5.485)
WDPM	2.713	-3.161	1.293	(2.458,2.968)
Gender	Est(10^{-1})	Bias(10^{-1})	RMSE(10^{-1})	95% CI(10^{-1})
UNWT	3.574	0.662	2.291	(-0.099,7.248)
FWT	2.913	0	1.116	(0.713,5.112)
WT3	2.918	0.005	1.563	(0.735,5.1)
WDPM	1.089	-1.824	1.459	(0.706,1.471)
Median				
Age	Est(10^{-2})	Bias(10^{-3})	RMSE(10^{-3})	95% CI(10^{-2})
UNWT	3.964	-1.657	3.140	(3.450,4.478)
FWT	4.130	0	3.678	(3.405,4.855)
WT3	4.129	-0.01	5.182	(3.407,4.851)
WDPM	4.083	-0.469	0.862	(3.913,4.253)
Gender	Est(10^{-1})	Bias(10^{-1})	RMSE(10^{-1})	95% CI(10^{-1})
UNWT	3.109	-0.280	0.807	(1.518,4.699)
FWT	3.388	0	2.225	(-0.995,7.772)
WT3	3.412	0.024	3.157	(-0.99,7.815)
WDPM	1.135	-2.253	0.525	(0.615,1.655)
Third Quartile				
Age	Est(10^{-2})	Bias(10^{-3})	RMSE(10^{-3})	95% CI(10^{-2})
UNWT	3.438	-1.291	5.483	(2.641,4.235)
FWT	3.567	0	4.588	(2.663,4.471)
WT3	3.566	-0.008	6.47	(2.664,4.468)
WDPM	3.465	-1.015	0.898	(3.289,3.642)
Gender	Est(10^{-1})	Bias(10^{-1})	RMSE(10^{-1})	95% CI(10^{-1})
UNWT	1.652	-0.279	0.649	(0.539,2.764)
FWT	1.931	0	0.648	(0.655,3.207)
WT3	1.926	-0.005	0.906	(0.659,3.192)
WDPM	0.802	-1.129	0.936	(0.500,1.105)

Table 4.14: Quantile regression of log TCDD on age and gender. Bias, RMSE and 95% CI for linear slope estimated for age and gender in unweighted, fully weighted, weight trimming by 3 SD and weighted Dirichlet Process Mixture models.

	First Quartile			
Age	Est(10^{-2})	Bias(10^{-3})	RMSE(10^{-3})	95% CI(10^{-2})
UNWT	1.938	0.328	7.505	(0.459,3.417)
FWT	1.905	0	11.843	(-0.428,4.238)
WT3	1.905	-0.005	16.689	(-0.421,4.23)
WDPM	2.083	1.774	6.574	(0.787,3.378)
Gender	Est(10^{-1})	Bias(10^{-1})	RMSE(10^{-1})	95% CI(10^{-1})
UNWT	-6.287	1.292	5.91	(-17.93,5.356)
FWT	-7.579	0	8.869	(-25.051,9.893)
WT3	-7.577	0.003	12.51	(-25.008,9.855)
WDPM	-5.893	1.686	4.814	(-15.377,3.59)
Interaction	Est(10^{-2})	Bias(10^{-3})	RMSE(10^{-2})	95% CI(10^{-2})
UNWT	2.394	-1.609	1.003	(0.418,4.369)
FWT	2.555	0	1.923	(-1.233,6.343)
WT3	2.556	0.012	2.707	(-1.217,6.329)
WDPM	1.951	-6.035	1.108	(-0.231,4.134)
	Median			
Age	Est(10^{-2})	Bias(10^{-3})	RMSE(10^{-3})	95% CI(10^{-2})
UNWT	3.499	-0.67	2.05	(3.095,3.902)
FWT	3.566	0	3.157	(2.944,4.187)
WT3	3.564	-0.011	4.458	(2.942,4.187)
WDPM	3.397	-1.686	3.067	(2.793,4.001)
Gender	Est(10^{-1})	Bias(10^{-1})	RMSE(10^{-1})	95% CI(10^{-1})
UNWT	-0.227	0.539	1.912	(-3.993,3.539)
FWT	-0.765	0	6.058	(-12.7,11.169)
WT3	-0.765	0.001	8.555	(-12.685,11.156)
WDPM	-1.486	-0.721	1.948	(-5.325,2.352)
Interaction	Est(10^{-2})	Bias(10^{-3})	RMSE(10^{-3})	95% CI(10^{-2})
UNWT	0.723	-1.703	2.797	(0.171,1.274)
FWT	0.893	0	9.05	(-0.89,2.676)
WT3	0.894	0.011	12.763	(-0.885,2.673)
WDPM	0.937	0.445	4.682	(0.015,1.86)
	Third Quartile			
Age	Est(10^{-2})	Bias(10^{-3})	RMSE(10^{-3})	95% CI(10^{-2})
UNWT	3.313	-2.318	5.311	(2.514,4.112)
FWT	3.545	0	5.468	(2.467,4.622)
WT3	3.543	-0.015	7.718	(2.467,4.62)
WDPM	3.461	-0.840	2.822	(2.905,4.017)
Gender	Est(10^{-1})	Bias(10^{-1})	RMSE(10^{-1})	95% CI(10^{-1})
UNWT	0.297	-1.383	1.818	(-3.284,3.877)
FWT	1.679	0	2.449	(-3.146,6.505)
WT3	1.666	-0.013	3.453	(-3.157,6.489)
WDPM	1.292	-0.387	6.363	(-11.244,13.828)

Continued on next page

Table 4.15 – continued from previous page

Interaction	Est(10^{-1})	Bias(10^{-1})	RMSE(10^{-1})	95% CI(10^{-1})
UNWT	2.474	1.943	2.961	(-0.336,0.831)
FWT	0.531	0	3.989	(-0.733,0.839)
WT3	0.546	0.015	5.595	(-0.727,0.836)
WDPM	3.034	2.503	8.392	(-1.35,1.957)

Table 4.15: Quantile regression of log TCDD on age, gender and interaction. Bias, RMSE and 95% CI for linear slope estimated for age, gender and interaction in unweighted, fully weighted, weight trimming by 3 SD and weighted Dirichlet Process Mixture model.

The patterns for quantile regression applied on NHANES study are not consistent across different quartiles. In general, when estimating the population slope of age in the first quartile of outcomes, the unweighted method is clearly more efficient than the fully-weighted one, reducing the RMSE by almost 50%. The performance of the unweighted and fully weighted estimators of the median and third quartile of outcomes are much closer between the two methods, usually within less than 15% differences in RMSEs, and with no one method besting the other across all locations and settings. When dealing with gender, the fully weighted estimate is usually favored in the main-effect only model with respect to RMSE. However, in the two-way interaction model, weights have less impact, and efficiency again becomes the dominating factor in the overall performance, leading the unweighted models to be favored. Since very few weights actually fall out the range of mean plus or minus three standard deviation, the weight trimming method makes little modification, closely resembles the fully weighted method results and obtains larger RMSEs due to the way they are calculated.

The WDPM method always provides estimates with smaller variance. For age, the differences between the WDPM estimates and the fully-weighted estimates are small, this reduction in variability leads to large reduction in RMSE across all quartiles.

For gender, the WDPM results are quite different from the other methods' results, which are more similar to each other. This is consistent with our simulation finding that both the unweighted and full weighted estimator of linear trends can be highly biased in this setting, while the WDPM approach yields nearly unbiased estimates. Hence we do not fully trust the bias estimates for gender, although we cannot know the truth in this setting.

4.6 Discussion

Generally, fully weighted estimators are applied when sampling weights are involved and bias correction is a primary concern. However, when the sample scheme and weights are unrelated to quantity of interest, these estimators can introduce extra variance and lead to substantial losses in efficiency. Other design based methods like ad-hoc weight trimming usually target a tradeoff between accuracy and efficiency. Meanwhile, by proposing a robust model on for the population distribution, finite population Bayesian inference has the potential to develop estimators that have reduced RMSE relative to both unweighted and weighted estimates. Here we adapt a weighted Dirichlet Process Mixture model to capture underlying data structures, and prevent issues like “over-smoothing” that simpler models can suffer from. For certain simulations considered, WDPM estimates provided a 70% reduction in RMSE comparing to fully weighted estimates, with approximate nominal 95% interval coverage rates. Similar reductions are also observed in application, particularly with respect to gender-adjusted effects of age on dioxin in NHANES study.

In addition, the weighted Dirichlet Process Mixture model can be used to model more than just means and/or slopes within subgroups of population. By including parameters for differing variances among components, our model provides a flexible method capable of precisely replicating non-normal population distributions, enabling inference on population quantiles or coefficients from quantile regression. Our simu-

lations also show promising results in quantile regression settings, suggesting overall reduction in RMSEs with sufficient nominal coverage rates.

Our approach could be extended in a number of ways. We have focused on continuous outcomes, but extensions to binary or multinomial models are straightforward by modeling latent variables under a weighted DPM frame. Also, highly skewed distributions could be modeled using alternative skewed shape measures such as gamma distribution to further reduce number of components required and simplify the model, although with heavily skewed populations and/or very small number of sampled observations in certain areas of the prediction space, the WDPM model may provide “over-smoothed” fits leading to poor coverage. It is crucial to test and understand those boundaries before we consider WDPM as a routine method in handling complex survey data.

A final aspect of the method is that the improvement in overall RMSE from complex Bayesian model is based on intensive computation. Based on current available computation facilities, results for small samples with hundreds of observations could be obtained in a reasonable amount of time, but when sample size escalates beyond several thousand cases, or the target quantity of interest has complicated form, the computing time quickly raises to an intolerable scale. (This is also a factor in our limiting simulations to 200 per scenario, even with access to a large number of computing cores.) We expect that with the fast development of hardware as well as parallel processing, we could soon put WDPM into application, analyzing data with large sample size.

CHAPTER V

Conclusion

Comparing to traditional design-based approaches, finite population Bayesian inference provides an alternative perspective in analyzing data from complex survey design. By assuming that the data follows certain distribution models conditioned on some unknown parameters, we can make inference based on posterior predictive distributions conditioned on samples and weights. By using models flexible enough to capture various patterns from data, we expect the estimates to achieve a balance between bias-correction and efficiency in a data-driven manner.

In this manuscript, we study three types of candidate Bayesian models for complex sample survey analysis: linear spline models combined with variable selection and Fractional Bayes Factor that mimic “weight-trimming”; hierarchical “weight-smoothing” models using Laplace prior that treat interactions between weight strata and quantity of interest as random effects and form shrinkage type estimators; and Weighted Dirichlet Process Mixture Model using advanced mixture model known to approximate skewed or multimodal distributions with small sample sizes. Both simulations and applications are conducted to assess their performances against competing methods in estimating population slope under various population settings. Bias, RMSE and 95% confidence interval coverage rate are used to make the comparison.

In general, we observe considerable improvement in overall RMSE comparing to

our proposed methods comparing with traditional competing methods like weight trimming, etc. The Bayesian methods approximate the unweighted method when data suggest no evidence of difference among strata, but correct the bias when it is necessary, successfully provide estimates that are self-tuned in a data-driven manner as we expected. More interestingly, finite population Bayesian inference often provides estimates with better RMSEs that outperform both unweighted and fully weighted estimates, given that clear non-linear trends between outcome and covariate are observed. The detailed mechanism that causes this feature is to be investigated in future work.

We also learn the limitations of Bayesian Inference from the simulation studies, mainly appearing as overfitting and unstable modeling. Due to the "model-based" nature of these methods, fitting the Bayesian model correctly and precisely on the given data is crucial to obtaining excellent estimates. With sophisticated models like we proposed, it could often correct the issue of over-smoothing (under-fitting and closely resembling unweighted estimates), which is common among simpler models. However, it is potentially risky that the Bayesian models could by mistake identify noise in the sample as pattern, overfit to the sample and leads to biased result and low coverage rate. This happens in the simulation study when strong non-linear curves are observed in the sampled dataset. Also, unstable estimates could appear when sampled observations are severely unevenly distributed. With few or no observations within certain area, the Bayesian model cannot correctly estimate underlying parameters responsible for that area, and create heavily biased estimate when target quantity of interest is heavily rely on those poorly estimated parameters. In simulation studies, population settings with very low probabilities of selection in certain range of the covariate often result in increased biases in estimates. Further studies for a deeper understanding on these limits are also an important next step.

Another obstacle that limits the application of Bayesian inference methods is the

computational burden. The model flexibility in adapting various data patterns is somewhat equivalent to model complexity, and their solutions rely on MCMC-type method like Gibbs Sampler and Metropolis Steps which are computation intensive algorithms. According to our simulations, with sample size under 5000 for "weight-pooling" or "weight-smoothing" model, or under 500 for WDPM model, the algorithms take from merely minutes to acceptably few hours. However, as sample size increases, the computing time quickly becomes unbearable, limits their application on large datasets or target quantities of interest as more complicated functions of population. So currently, we focus on models' applications in obtaining highly efficient estimates from small samples, and hope that in the near future, we can extend their use to larger samples with the rapid development of computer hardware and parallel processing methods.

APPENDICES

APPENDIX A

Metropolis Algorithm For Advanced Weight Pooling Method

A.1 Metropolis Algorithm for Advanced Weight Pooling Method

To identify different pooling pattern, we sort the data in ascending order by weights, and create H potential strata of equal size by identifying $H - 1$ cut points, namely $\tau_1, \dots, \tau_{H-1}$. A vector V of $H - 1$ dichotomous elements v_1, v_2, \dots, v_{H-1} , is also created to mark each of the cut points. That is, if the adjacent strata are pooled together, thus the cut point in between is ignored, the corresponding element in the vector should be coded as 0, otherwise 1. So a complete pooling model would result in V consisting of all 0s, and no pooling model as a V of all 1s. Using binary coding, the unique ID of each pooling pattern is obtained by letting $l = \sum_{i=1}^{H-1} (2^{i-1} v_i)$. The Metropolis algorithm starts with V contains all 0s.

Step 1:

Propose a new l' by first randomly selecting one element in V , namely the j th element, and let $V' = (v_1, v_2, \dots, v_{j-1}, 1 - v_j, v_{j+1}, \dots)$. Then l' is calculated as $l' = \sum_{i=1}^{H-1} (2^{i-1} v'_i)$.

Notice that the process is equivalent for all elements, thus $p(L = l) = 1/2^{H-1}$ for all l , meaning that all models have equal probability of been proposed. The corresponding design matrix $Z_{l'}$ is also calculated so that the i th row $Z_{l'i} = x_i(1, (h - \tau_1)_+, \dots, (h - \tau_{H-1})_+)$ '

Step 2:

To determine the acceptance of the proposed model, the ratio of $p(L = l'|y, X)/p(L = l|y, X) = p(y|L = l', X)/p(y|L = l, X)$, is calculated where $p(y|L = l, X) \propto |\Psi_l|^{1/2}[\Delta_l - \theta_l^T \Psi_l \theta_l]^{-(n+a)/2}$,

for

$$\psi_l = ((Z_l^T Z_l) + \Sigma_0)^{-1},$$

$$\theta_l = (Z_l^T Z_l)b + \Sigma_0 \beta_0,$$

$$\Delta_l = b^T (Z_l^T Z_l)b + \beta_0^T \Sigma_0^{-1} \beta_0 + Q_l^2 + as^2,$$

$$b = (Z_l^T Z_l)^{-1} Z_l^T y,$$

$$Q_l^2 = y^T (I_{p_{H^*}} - H_l)y,$$

$$H_l = Z_l (Z_l^T Z_l)^{-1} Z_l^T.$$

$p(y|L = l', X)$ is calculated similarly by replacing Z_l with $Z_{l'}$.

Step 3:

Accept the new pooling model l' with probability $\min(1, p(L = l'|y, X)/p(L = l|y, X))$, otherwise, stay at $L = l$.

Step 4:

Under current model selection l' , full conditional distribution of parameter β s and σ^2 are available for direct draws:

$$\text{Draw } \sigma^2 \text{ from } \sigma^2 | L = l, y, X \sim \text{Inv}_\chi^2(n + a, \Delta + l - \theta_l^T \psi_l \theta_l)$$

$$\text{Draw } \beta_l \text{ from } \beta_l | \sigma^2, y, X \sim N(\Gamma_l A_l, \sigma^2 \Gamma_l), A_l = Z_l^T y + \Sigma_0^{-1} \beta_0, \Gamma_l = [\Sigma_0^{-1} + (Z_l^T Z_l)]^{-1}$$

Step 5:

To assess the posterior predictive distribution of population slope, $P(B|Y)$, the point estimate can be obtained by first creating the predicted y s on sampled parameters as $Z_l\beta_l$, then estimating B from fully weighted regression model on predicted Y s.

Repeated step 1 to step 5 until the estimated B is stable. The pooled estimated B is a sample of its posterior predictive distribution, on which the inference on population slope B is based.

APPENDIX B

Gibbs Sampler For Weight Smoothing Model

B.1 Full Conditional Distribution for Linear Model

To derive the fully conditional distribution of the linear model for Gibbs sampler, first we start with the hierarchical model:

$$Y_h \sim MVN(X_h\beta_h, \sigma^2 I_{n_h})$$

$$\beta_h = (\beta_{h1}, \dots, \beta_{hp})^T, h = 1, \dots, H$$

$$\beta_h \sim MVN(\beta_h^*, \sigma^2 D_{\tau h})$$

$$\beta_h^* \sim MVN(0_p, \sigma_0^2 I_p)$$

$$D_{\tau h} = \text{diag}(\tau_{h1}^2, \dots, \tau_{hp}^2)$$

$$\sigma^2 \sim 1/\sigma^2$$

$$\tau_{hi}^2 \sim \frac{\lambda^2}{2} e^{-\lambda^2 \tau_{hi}^2 / 2}$$

$$\lambda^2 \sim \text{Gamma}(\gamma, \delta)$$

Ignoring all constants, we reduce the formula to the kernel of likelihood of y , and all other conditional probabilities:

$$\begin{aligned}
p(Y|\beta, \sigma^2) &\propto (\sigma^2)^{-n/2} \prod_{h=1}^H \exp\left\{-\frac{1}{2}(Y_h - X_h\beta_h)^T (\sigma^2 I_{n_h})^{-1} (Y_h - X_h\beta_h)\right\} \\
p(\beta|\beta^*, \sigma^2, D_\tau) &\propto (\sigma^2)^{-Hp/2} \prod_{h=1}^H |D_{\tau h}|^{-1/2} \exp\left\{-\frac{1}{2}(\beta_h - \beta_h^*)^T (\sigma^2 D_{\tau h})^{-1} (\beta_h - \beta_h^*)\right\} \\
f(\beta^*) &\propto \prod_{h=1}^H \exp\left\{-\frac{1}{2}\beta_h^{*T} (\sigma_0^2 I_p)^{-1} \beta_h^*\right\} \\
f(\sigma^2) &\propto 1/\sigma^2 \\
f(\tau^2|\lambda^2) &\propto (\lambda^2)^{Hp} \prod_{h=1}^H \prod_{i=1}^p \exp(-\lambda^2 \tau_{hi}^2/2) \\
f(\lambda^2) &\propto (\lambda^2)^{\gamma-1} \exp(-\delta\lambda^2)
\end{aligned}$$

Since β_h s from different strata are independent, we write separately the kernel of posterior distribution of β_h , which is proportional to the product of likelihood of y and β_h prior.

$$\begin{aligned}
p(\beta_h|rest) &\propto \exp\left\{-\frac{1}{2\sigma^2}[(Y_h - X_h\beta_h)^T (Y_h - X_h\beta_h) + (\beta_h - \beta_h^*)^T D_{\tau h}^{-1} (\beta_h - \beta_h^*)]\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}[\beta_h^T X_h^T X_h \beta_h - 2Y_h^T X_h \beta_h + \beta_h^T D_{\tau h}^{-1} \beta_h - 2\beta_h^{*T} D_{\tau h}^{-1} \beta_h]\right\} \\
&= \exp\left\{-\frac{1}{2\sigma^2}[\beta_h^T (X_h^T X_h + D_{\tau h}^{-1})\beta_h - 2(Y_h^T X_h + \beta_h^{*T} D_{\tau h}^{-1})\beta_h]\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}[(\beta_h - (X_h^T X_h + D_{\tau h}^{-1})^{-1}(Y_h^T X_h + D_{\tau h}^{-1} \beta_h^*))^T (X_h^T X_h + D_{\tau h})^{-1} \right. \\
&\quad \left. (\beta_h - (X_h^T X_h + D_{\tau h}^{-1})^{-1}(Y_h^T X_h + D_{\tau h}^{-1} \beta_h^*))]\right\}
\end{aligned}$$

Which suggests that $\beta_h|rest \sim MVN(A^{-1}(X_h^T Y_h + D_{\tau h}^{-1} \beta_h^*), \sigma^2 A^{-1})$, $A = X_h^T X_h + D_{\tau h}^{-1}$.

Similarly, we derive the kernel of fully conditional distribution of other parameters as

follows:

$$\begin{aligned}
p(\beta_h^* | rest) &\propto \exp\left\{-\frac{1}{2}[(\beta_h^* - \beta_h)^T (\sigma^2 D_{\tau h})^{-1} (\beta_h^* - \beta_h) + \beta_h^{*T} (\sigma_0^2 I)^{-1} \beta_h^*]\right\} \\
&\propto \exp\left\{-\frac{1}{2}[\beta_h^{*T} ((\sigma^2 D_{\tau h})^{-1} + (\sigma_0^2 I)^{-1}) \beta_h^* - 2\beta_h^T (\sigma^2 D_{\tau h})^{-1} \beta_h^*]\right\} \\
&\propto \exp\left\{-\frac{1}{2}(\beta_h^* - ((\sigma^2 D_{\tau h})^{-1} + (\sigma_0^2 I)^{-1})^{-1} (\sigma^2 D_{\tau h})^{-1} \beta_h)^T ((\sigma^2 D_{\tau h})^{-1} + (\sigma_0^2 I)^{-1}) \right. \\
&\quad \left. (\beta_h^* - ((\sigma^2 D_{\tau h})^{-1} + (\sigma_0^2 I)^{-1})^{-1} (\sigma^2 D_{\tau h})^{-1} \beta_h)\right\}
\end{aligned}$$

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

$$\begin{aligned}
p(\sigma^2 | rest) &\propto (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2}(\sigma^2)^{-1} \sum_{h=1}^H (Y_h - X_h \beta_h)^T (Y_h - X_h \beta_h)\right\} * \\
&\quad (\sigma^2)^{-Hp/2} \exp\left\{-\frac{1}{2}(\sigma^2)^{-1} \sum_{h=1}^H (\beta_h - \beta_h^*)^T (D_{\tau h})^{-1} (\beta_h - \beta_h^*)\right\} * (\sigma^2)^{-1} \\
&= (\sigma^2)^{-(n/2+Hp/2)-1} \exp\left\{-\frac{1}{2}(\sigma^2)^{-1} \left[\sum_{h=1}^H (Y_h - X_h \beta_h)^T (Y_h - X_h \beta_h) + \right. \right. \\
&\quad \left. \left. \sum_{h=1}^H (\beta_h - \beta_h^*)^T (D_{\tau h})^{-1} (\beta_h - \beta_h^*)\right]\right\}
\end{aligned}$$

$$\begin{aligned}
\sigma^2 | rest &\sim InvGamma((n + Hp)/2, \frac{1}{2} \left[\sum_{h=1}^H (Y_h - X_h \beta_h)^T (Y_h - X_h \beta_h) + \right. \\
&\quad \left. \sum_{h=1}^H (\beta_h - \beta_h^*)^T (D_{\tau h})^{-1} (\beta_h - \beta_h^*)\right])
\end{aligned}$$

$$\begin{aligned}
p(1/\tau_{hi}^2 | rest) &\propto (\tau_{hi}^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \frac{(\beta_{hi} - \beta_{hi}^*)^2}{\sigma^2 \tau_{hi}^2}\right) * \exp\left(-\frac{\lambda^2 \tau_{hi}^2}{2}\right) * d(\tau_{hi}^2) \\
&\propto (1/\tau_{hi}^2)^{\frac{1}{2}} \exp\left[-\frac{1}{2} \left(\frac{(\beta_{hi} - \beta_{hi}^*)^2 (1/\tau_{hi}^2)}{\sigma^2} + \frac{\lambda^2}{1/\tau_{hi}^2}\right)\right] * (1/\tau_{hi}^2)^{-2} \\
&= (1/\tau_{hi}^2)^{-\frac{3}{2}} \exp\left[-\frac{1}{2} \left(\frac{(\beta_{hi} - \beta_{hi}^*)^2 (1/\tau_{hi}^2)^2 + \lambda^2 \sigma^2}{\sigma^2 (1/\tau_{hi}^2)}\right)\right] \\
&\propto (1/\tau_{hi}^2)^{-\frac{3}{2}} \exp\left[-\frac{1}{2} \frac{((1/\tau_{hi}^2) - \sqrt{\lambda^2 \sigma^2 / (\beta_{hi} - \beta_{hi}^*)^2})^2}{(\beta_{hi} - \beta_{hi}^*)^{-2} \sigma^2 (1/\tau_{hi}^2)}\right]
\end{aligned}$$

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

$$\begin{aligned}
p(\lambda^2|rest) &\propto (\lambda^2)^{Hp} \exp\left(-\frac{1}{2}\lambda^2 \sum_{h=1}^H \sum_{i=1}^p \tau_{hi}^2\right) * (\lambda^2)^{\gamma-1} \exp(-\delta\lambda^2) \\
&= (\lambda^2)^{Hp+\gamma-1} \exp\left[-\lambda^2\left(\frac{1}{2} \sum_{h=1}^H \sum_{i=1}^p \tau_{hi}^2 + \delta\right)\right] \\
\lambda^2 &\sim \text{Gamma}(Hp + \gamma, \frac{1}{2} \sum_{h=1}^H \sum_{i=1}^p \tau_{hi}^2 + \delta)
\end{aligned}$$

B.2 Full Conditional Distribution for Logistic Model

$$y_{hi}|X_{hi}, \beta_h, \sim \text{Binomial}(p = \text{logit}(x_{hi}\beta_h))$$

$$\beta_h = (\beta_{h1}, \dots, \beta_{hp})^T, h = 1, \dots, H$$

$$\beta_h \sim \text{MVN}(\beta_h^*, \sigma^2 D_{\tau h})$$

$$\beta_h^* \sim \text{MVN}(0_p, \sigma_0^2 I_p)$$

$$D_{\tau h} = \text{diag}(\tau_{h1}^2, \dots, \tau_{hp}^2)$$

$$\tau_{hi}^2 \sim \frac{\lambda^2}{2} e^{-\lambda^2 \tau_{hi}^2 / 2}$$

$$\lambda^2 \sim \text{Gamma}(\gamma, \delta)$$

Similarly, we start with the hierarchical model, derive the kernel of the posterior distribution of all parameters, and reveal that they belongs to some known distribution families. The full conditional distribution of β_h doesn't belong to any known

distribution family, and the rest of parameters are presented below:

$$\begin{aligned}
p(Y|\beta) &= \prod_{h=1}^H \prod_{i=1}^{n_h} \left(\frac{\exp(x_{hi}\beta_h)}{1 + \exp(x_{hi}\beta_h)} \right)^{y_{hi}} \left(\frac{1}{1 + \exp(x_{hi}\beta_h)} \right)^{1-y_{hi}} \\
p(\beta|\beta^*, D_\tau) &\propto \prod_{h=1}^H |D_{\tau h}|^{-1/2} \exp\left\{-\frac{1}{2}(\beta_h - \beta_h^*)^T (D_{\tau h})^{-1} (\beta_h - \beta_h^*)\right\} \\
f(\beta^*) &\propto \prod_{h=1}^H \exp\left\{-\frac{1}{2}\beta_h^{*T} (\sigma_0^2 I_p)^{-1} \beta_h^*\right\} \\
f(\tau^2|\lambda^2) &\propto (\lambda^2)^{Hp} \prod_{h=1}^H \prod_{i=1}^p \exp(-\lambda^2 \tau_{hi}^2/2) \\
f(\lambda^2) &\propto (\lambda^2)^{\gamma-1} \exp(-\delta\lambda^2)
\end{aligned}$$

$$\begin{aligned}
p(\beta_h^*|rest) &\propto \exp\left\{-\frac{1}{2}[(\beta_h^* - \beta_h)^T D_{\tau h}^{-1} (\beta_h^* - \beta_h) + \beta_h^{*T} (\sigma_0^2 I)^{-1} \beta_h^*]\right\} \\
&\propto \exp\left\{-\frac{1}{2}[\beta_h^{*T} (D_{\tau h}^{-1} + (\sigma_0^2 I)^{-1}) \beta_h^* - 2\beta_h^{*T} D_{\tau h}^{-1} \beta_h]\right\} \\
&\propto \exp\left\{-\frac{1}{2}(\beta_h^* - (D_{\tau h}^{-1} + (\sigma_0^2 I)^{-1})^{-1} D_{\tau h}^{-1} \beta_h)^T (D_{\tau h}^{-1} + (\sigma_0^2 I)^{-1}) \right. \\
&\quad \left. (\beta_h^* - (D_{\tau h}^{-1} + (\sigma_0^2 I)^{-1})^{-1} D_{\tau h}^{-1} \beta_h)\right\}
\end{aligned}$$

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

$$\begin{aligned}
p(1/\tau_{hi}^2|rest) &\propto (\tau_{hi}^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \frac{(\beta_{hi} - \beta_{hi}^*)^2}{\tau_{hi}^2}\right) * \exp\left(-\frac{\lambda^2 \tau_{hi}^2}{2}\right) * d(\tau_{hi}^2) \\
&\propto (1/\tau_{hi}^2)^{\frac{1}{2}} \exp\left[-\frac{1}{2}((\beta_{hi} - \beta_{hi}^*)^2 (1/\tau_{hi}^2) + \frac{\lambda^2}{1/\tau_{hi}^2})\right] * (1/\tau_{hi}^2)^{-2} \\
&= (1/\tau_{hi}^2)^{-\frac{3}{2}} \exp\left[-\frac{1}{2} \left(\frac{(\beta_{hi} - \beta_{hi}^*)^2 (1/\tau_{hi}^2)^2 + \lambda^2}{(1/\tau_{hi}^2)} \right)\right] \\
&\propto (1/\tau_{hi}^2)^{-\frac{3}{2}} \exp\left[-\frac{1}{2} \frac{((1/\tau_{hi}^2) - \sqrt{\lambda^2 / (\beta_{hi} - \beta_{hi}^*)^2})^2}{(\beta_{hi} - \beta_{hi}^*)^{-2} (1/\tau_{hi}^2)}\right]
\end{aligned}$$

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

$$\begin{aligned}
p(\lambda^2|rest) &\propto (\lambda^2)^{Hp} \exp\left(-\frac{1}{2}\lambda^2 \sum_{h=1}^H \sum_{i=1}^p \tau_{hi}^2\right) * (\lambda^2)^{\gamma-1} \exp(-\delta\lambda^2) \\
&= (\lambda^2)^{Hp+\gamma-1} \exp\left[-\lambda^2\left(\frac{1}{2} \sum_{h=1}^H \sum_{i=1}^p \tau_{hi}^2 + \delta\right)\right] \\
\lambda^2 &\sim \text{Gamma}\left(Hp + \gamma, \frac{1}{2} \sum_{h=1}^H \sum_{i=1}^p \tau_{hi}^2 + \delta\right)
\end{aligned}$$

APPENDIX C

Gibbs Sampler for Weighted Dirichlet Process Mixture Model with Survey Weights

C.1 Gibbs Sampler for Linear Regression

Let $S = (S_1, \dots, S_n)$, $S_i = h$, $h \in (1, k)$ denote the configuration of i th observation to h th component.

Let θ_h denote the distinct parameters for h th component.

And let $C = (C_1, \dots, C_k)$, $C_h = j$ denote that θ_h is an atom from j th basis distribution.

Step 1: Update S_i

$$\begin{aligned}
S_i | \text{rest} &\sim \text{Multinomial}(0, 1, 2 \dots k) \\
Pr(S_i = h) &\propto \frac{b_{i, C_h^{(i)}} \sum_{m \neq i} \mathbf{1}(S_m^{(i)} = h)}{\alpha + \sum_{l \neq i} \mathbf{1}(C_{S_l^{(i)}}^{(i)} = C_h)} f(y_i | x_i, \theta_h) \\
Pr(S_i = 0) &\propto \sum_{j=1}^n \frac{\alpha b_{ij}}{\alpha + \sum_{l \neq i} \mathbf{1}(C_{S_l^{(i)}}^{(i)} = j)} \int f(y_i | x_i, \phi) dG_0(\phi) \\
y_i | x_i, \theta_h &\sim N(x_i \beta_h, \sigma_h^2) \\
\int f(y_i | x_i, \phi) dG_0(\phi) &= \frac{N_p(0; \beta_0, \Sigma_\beta) N(0; y_i, \alpha_\tau / \beta_\tau)}{N_p(0; \hat{\beta}_i, \hat{V}_{\beta_i})} \\
\hat{\beta}_h &= \hat{V}_{\beta_h} (\Sigma_\beta^{-1} \beta + \sigma_i^{-2} x_i y_i) \\
\hat{V}_{\beta_h} &= (\Sigma_\beta^{-1} + \sigma_i^{-2} x_i x_i')^{-1}
\end{aligned}$$

θ_h for new component is drawn similar to step 2 below.

Step 2: Update $\theta_h = (\beta_h, \sigma_h^2)$

$$\begin{aligned}
\sigma_h^{-2} | \text{rest} &\sim \text{Gamma}\left\{a_\tau + \frac{n_h}{2}, b_\tau + \frac{1}{2}(y^t y + \beta^T \Sigma_\beta^{-1} \beta - \beta_h^* \Sigma_\beta^{*-1} \beta_h^*)\right\} \\
\beta_h^* &= (X_h^T X_h + \Sigma_\beta^{-1})^{-1} (\Sigma_\beta^{-1} \beta + X_h^T y) \\
\Sigma_h^* &= (X_h^T X_h + \Sigma_\beta^{-1})^{-1} \\
\beta_h | \text{rest} &\sim N_p(\beta_h^*, \sigma_h^2 \Sigma_h^*) \\
b_\tau &\sim \text{Gamma}(a_0 + k a_\tau, b_0 + \sum_{j=1}^k \tau_j)
\end{aligned}$$

Step 3: Update C_h

$$\begin{aligned}
C_h | \text{rest} &\sim \text{Multinomial}(1, 2 \dots n) \\
Pr(C_h = j | \text{rest}) &= \frac{\prod_{i: S_i = h} b_{ij} \Psi(y_i - x_i \beta_h)}{\sum_{l=1}^n \prod_{i: S_i = h} b_{il}}
\end{aligned}$$

Step 4: Update γ_j

Let $K_{ij}^* = K_{ij} / \sum_{l \neq j} \gamma_l K_{il}$, the conditional likelihood for γ_j is

$$L(\gamma_j) = \prod_{i=1}^n \left(\frac{\gamma_j K_{ij}^*}{1 + \gamma_j K_{ij}^*} \right)^{\mathbf{1}(C_{S_i} = j)} \left(\frac{1}{1 + \gamma_j K_{ij}^*} \right)^{\mathbf{1}(C_{S_i} \neq j)}$$

This can be obtained by using $\mathbf{1}(C_{S_i} = j) = \mathbf{1}(Z_{ij}^* > 0)$, and following the Gibbs

Steps:

(a) $Z_{ij}^* \sim \text{Poisson}(\gamma_j \xi_{ij} K_{ij}^*) \mathbf{1}(Z_{ij}^* > 0)$ if $C_{S_i} = j$, 0 otherwise.

(b) $\xi_{ij} \sim \text{Gamma}(1 + Z_{ij}^*, 1 + \gamma_j K_{ij}^*)$

(c) $\gamma_j \sim \text{Gamma}(a_\gamma + \sum_{i=1}^n Z_{ij}^*, b_\gamma + \xi_{ij} K_{ij}^*)$

Step 5: Update ψ

$$\psi \sim \log -N(\mu_\psi, \sigma_\psi^2)$$

$$Pr(y|\psi) = \sum_{i=1}^n \prod_{j=1}^n b_{ij} \Phi((y_i - x_i \beta_{S_j}) / \sigma_{S_j}^2)$$

The posterior distribution is not in any known analytical form, and can be obtained from Metropolis Algorithm.

Step 6: Obtain draws of estimated population slope B

Obtain draws of estimated population slope B from posterior predictive distribution:

$$P(B|y, x) = \int p(B|y, x, \phi_k) dG_0(\phi_k)$$

This is achieved by first recreate the population through obtaining the predicted Y s from

$$y_i^{rep} \sim w_{i0}(x_i) N(x_i^T \beta_0, \alpha_\tau / \beta_\tau + x_i^T \Sigma_\beta x_i) + \sum_{h=1}^k w_{ih}(x_i) N(x_i^T \beta_h, \sigma_h^2)$$

then performing linear regression on predicted population to estimate population slope.

Step 7: Repeated Step 1 to 6.

The inference is based on the pooled draws from step 6 after certain burn-in period.

C.2 Gibbs Sampler for Quantile Regression

Let $S = (S_1, \dots, S_n)$, $S_i = h$, $h \in (1, k)$ denote the configuration of i th observation to h th component.

Let θ_h denote the distinct parameters for h th component.

And let $C = (C_1, \dots, C_k)$, $C_h = j$ denote that θ_h is an atom from j th basis distribution.

Step 1: Update S_i

$$\begin{aligned}
 S_i | \text{rest} &\sim \text{Multinomial}(0, 1, 2 \dots k) \\
 Pr(S_i = h) &\propto \frac{b_{i, C_h^{(i)}} \sum_{m \neq i} \mathbf{1}(S_m^{(i)} = h)}{\alpha + \sum_{l \neq i} \mathbf{1}(C_{S_l^{(i)}}^{(i)} = C_h)} f(y_i | x_i, \theta_h) \\
 Pr(S_i = 0) &\propto \sum_{j=1}^n \frac{\alpha b_{ij}}{\alpha + \sum_{l \neq i} \mathbf{1}(C_{S_l^{(i)}}^{(i)} = j)} \int f(y_i | x_i, \phi) dG_0(\phi) \\
 y_i | x_i, \theta_h &\sim N(x_i \beta_h, \sigma_h^2) \\
 \int f(y_i | x_i, \phi) dG_0(\phi) &= \frac{N_p(0; \beta_0, \Sigma_\beta) N(0; y_i, \alpha_\tau / \beta_\tau)}{N_p(0; \hat{\beta}_i, \hat{V}_{\beta_i})} \\
 \hat{\beta}_h &= \hat{V}_{\beta_h} (\Sigma_\beta^{-1} \beta + \sigma_i^{-2} x_i y_i) \\
 \hat{V}_{\beta_h} &= (\Sigma_\beta^{-1} + \sigma_i^{-2} x_i x_i')^{-1}
 \end{aligned}$$

θ_h for new component is drawn similar to step 2 below.

Step 2: Update $\theta_h = (\beta_h, \sigma_h^2)$

$$\sigma_h^{-2}|rest \sim Gamma\{a_\tau + \frac{n_h}{2}, b_\tau + \frac{1}{2}(y^t y + \beta^T \Sigma_\beta^{-1} \beta - \beta_h^* \Sigma_\beta^{*-1} \beta_h^*)\}$$

$$\beta_h^* = (X_h^T X_h + \Sigma_\beta^{-1})^{-1} (\Sigma_\beta^{-1} \beta + X_h^T y)$$

$$\Sigma_h^* = (X_h^T X_h + \Sigma_\beta^{-1})^{-1}$$

$$\beta_h|rest \sim N_p(\beta_h^*, \sigma_h^2 \Sigma_h^*)$$

$$b_\tau \sim Gamma(a_0 + k a_\tau, b_0 + \sum_{j=1}^k \tau_j)$$

Step 3: Update C_h

$$C_h|rest \sim Multinomial(1, 2 \dots n)$$

$$Pr(C_h = j|rest) = \frac{\prod_{i:S_i=h} b_{ij} \Psi(y_i - x_i \beta_h)}{\sum_{l=1}^n \prod_{i:S_i=l} b_{il}}$$

Step 4: Update γ_j

Let $K_{ij}^* = K_{ij} / \sum_{l \neq j} \gamma_l K_{il}$, the conditional likelihood for γ_j is

$$L(\gamma_j) = \prod_{i=1}^n \left(\frac{\gamma_j K_{ij}^*}{1 + \gamma_j K_{ij}^*} \right)^{\mathbf{1}(C_{S_i}=j)} \left(\frac{1}{1 + \gamma_j K_{ij}^*} \right)^{\mathbf{1}(C_{S_i} \neq j)}$$

This can be obtained by using $\mathbf{1}(C_{S_i} = j) = \mathbf{1}(Z_{ij}^* > 0)$, and following the Gibbs

Steps:

(a) $Z_{ij}^* \sim Poisson(\gamma_j \xi_{ij} K_{ij}^*) \mathbf{1}(Z_{ij}^* > 0)$ if $C_{S_i} = j$, 0 otherwise.

(b) $\xi_{ij} \sim Gamma(1 + Z_{ij}^*, 1 + \gamma_j K_{ij}^*)$

(c) $\gamma_j \sim Gamma(a_\gamma + \sum_{i=1}^n Z_{ij}^*, b_\gamma + \xi_{ij} K_{ij}^*)$

Step 5: Update ψ

$$\psi \sim \log -N(\mu_\psi, \sigma_\psi^2)$$

$$Pr(y|\psi) = \sum_{i=1}^n \prod_{j=1}^n b_{ij} \Phi((y_i - x_i \beta_{S_j}) / \sigma_{S_j}^2)$$

The posterior distribution is not in any known analytical form, and can be obtained from Metropolis Algorithm.

Step 6: Obtain draws of estimated population slope B

Obtain draws of estimated population slope B from posterior predictive distribution:

$$P(B|y, x) = \int p(B|y, x, \phi_k) dG_0(\phi_k)$$

This is achieved by first recreate the population through obtaining the predicted Y s from

$$y_i^{rep} \sim w_{i0}(x_i) N(x_i^T \beta_0, \alpha_\tau / \beta_\tau + x_i^T \Sigma_\beta x_i) + \sum_{h=1}^k w_{ih}(x_i) N(x_i^T \beta_h, \sigma_h^2)$$

then applying quantile regression on predicted population to estimate population slope for specific quartile/percentile of outcome.

Step 7: Repeated Step 1 to 6.

The inference is based on the pooled draws from step 6 after certain burn-in period.

BIBLIOGRAPHY

BIBLIOGRAPHY

- Alexander, C., S. Dahl, and L. Weidman (1997), Making estimates from the American Community Survey, *Paper presented at the 1997 Joint Statistical Meetings, Anaheim, CA*.
- Beaumont, J. (2008), A new approach to weighting and inference in sample surveys, *Biometrika*, *95*, 539–553.
- Binder, D. (1983), On the variances of asymptotically normal estimators from complex surveys, *International Statistical Review*, *51*, 279–292.
- Blackwell, D., and J. MacQueen (1973), Ferguson distributions via Polya urn schemes, *The Annals of Statistics*, *1*, 353–355.
- Chen, Q., M. Elliott, and R. Little (2010a), Bayesian Penalized Spline Model-Based Inference for Finite Population Proportions in Unequal Probability Sampling, *Survey Methodology*, *36*, 22–34.
- Chen, Q., M. Elliott, and R. Little (2010b), Bayesian Inference for Finite Population Quantiles from Unequal Probability Samples, *Survey Methodology*, *38*, 203–215.
- Chen, Q., et al. (2010c), Estimation of Background Serum 2,3,7,8-TCDD Concentrations Using Quantile Regression in the UMDES and NHANES Populations, *Epidemiology*, *21*, S51–S57.
- Chowdhury, S., M. Khare, and K. Wolter (2007), Weight Trimming in the National Immunization Survey, *Proceedings of the Joint Statistical Meetings, Section on Survey Research Methods, American Statistical Association*, pp. 2651–8.
- Cochran, W. (1977), *Sampling Techniques*, Wiley: New York.
- Cox, B., and D. 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*.
- Dunson, D., N. Pillai, and J.-H. Park (2007), Bayesian density regression, *Journal of the Royal Statistical Society, B69*, 163–183.
- Elliott, M. (2007), Bayesian weight trimming for generalized linear regression models, *Survey Methodology*, *33*, 23–24.

- Elliott, M. (2008), Model averaging methods for weight trimming, *Journal of Official Statistics*, 24, 517–540.
- Elliott, M. (2009), Model averaging methods for weight trimming in generalized linear regression models, *Journal of Official Statistics*, 25, 1–20.
- Elliott, M., and R. 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(5), 300–303.
- Ericson, W. (1969), Subjective Bayesian modeling in sampling finite populations, *Journal of the Royal Statistical Society*, 31, 195–234.
- Ferguson, T. (1973), Bayesian analysis of some nonparametric problems, *Annals of Statistics*, 1(2), 209–230.
- Ghosh, M., and G. Meeden (1986), Empirical Bayes estimation of means from stratified samples, *Journal of the American Statistical Association*, 81, 1058–1062.
- Hansen, M. H., W. G. Madow, and B. J. Tepping (1983), An evaluation of mode-dependent and probability sampling inferences in sample surveys, *Journal of American Statistical Association*, 78, 776–793.
- Henry, K., and R. V. Valliant (2012), Methods for Adjusting Survey Weights when Estimating a Total, *Proceedings of the 2012 Federal Committee on Statistical Methodology's Research Conference*.
- Holt, D., and T. Smith (1979), Poststratification, *Journal of the Royal Statistical Society*, 142, 33–46.
- Horvitz, D., and D. Thompson (1952), A generalization of sampling without replacement from a finite universe, *Journal of the American Statistical Association*, 47, 663–685.
- Kass, R., and A. Raftery (1995), Bayes Factors, *Journal of the American Statistical Association*, 90, 773–795.
- Kish, L. (1992), Weighting for unequal P_i , *Journal of Official Statistics*, 8, 183–200.
- Koenker, R. (2004), *Quantile Regression*, Cambridge University Press: Cambridge, UK.
- Lazzeroni, L., and R. Little (1998), Random-effects models for smoothing post-stratification weights, *Journal of Official Statistics*, 14, 61–78.
- Little, R. (1983), Estimating a finite population mean from unequal probability samples, *Journal of the American Statistical Association*, 78, 596–604.

- Little, R. (1991), Inference with survey weights, *Journal of Official Statistics*, 7, 405–424.
- Little, R. (1993), Poststratification: A modeler’s perspective, *Journal of the American Statistical Association*, 88, 1001–1012.
- Little, R. (2011), Calibrated Bayes, for Statistics in General, and Missing Data in Particular, *Statistical Science*, 26, 162–174.
- MacEachern, S. N. (1994), Estimating normal means with a conjugate style Dirichlet process prior, *Communications in Statistics - Simulation and Computation*, 23, 727–741.
- McCullagh, P., and J. Nelder (1989), *Generalized Linear Models*, 2nd Edition, CRC Press: Boca Raton, Florida.
- McLachlan, G., and D. Peel (2000), *Finite Mixture Models*, New York: Wiley.
- Murty, K. (1983), *Linear programming*, New York: Wiley.
- O’Hagan, A. (1995), Fractional Bayes Factors for model comparison, *Journal of the Royal Statistical Society, B57*, 99–138.
- Park, T., and G. Casella (2008), The Bayesian Lasso, *Journal of the American Statistical Association*, 103(482), 681–686.
- Pfeffermann, D. (1993), The role of sampling weights when modeling survey data, *International Statistical Review*, 61, 317–337.
- Potter, F. (1988), Survey of Procedures to Control Extreme Sampling Weights, *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 453–458.
- Potter, F. (1990), A study of procedures to identify and trim extreme sample weights, *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 225–230.
- Rizzo, L. (1992), Conditionally consistent estimators using only probabilities of selection in complex sample surveys, *Journal of the American Statistical Association*, 87, 1166–1173.
- Rubin, D. (1987), *Multiple Imputation for Non-Response in Surveys*, Wiley: New York.
- Sarndal, C. (1980), On π -inverse weighting verses best linear unbiased weighting in probability sampling, *Biometrika*, 67, 639–650.
- Skinner, C., D. Holt, and T. Smith (1989), *Analysis of Complex Surveys*, Wiley: New York.

- Skinner, C. J., and J. Rao (1996), Estimation in dual frame surveys with complex designs, *Journal of American Statistical Association*, 91.
- Stigler, S. (1984), Studies in the history of probability and statistics XL: Boscovich, Simpson and a 1760 manuscript note on fitting a linear relation, *Biometrika*, 71, 615–620.
- Winston, F., M. Kallan, M. Elliott, R. Menon, and D. Durbin (2002), Risk of injury to child passengers in compact extended pickup trucks, *Journal of the American Medical Association*, 287, 1147–1152.
- Yu, K., and R. Moyeed (2001), Bayesian quantile regression, *Statistics and Probability Letters*, 54, 437–447.
- Zheng, H., and R. Little (2003), Penalized spline model-based estimation of the finite populations total from probability-proportional-to-size samples, *Journal of Official Statistics*, 19, 99–117.
- Zheng, H., and R. Little (2004), Penalized spline nonparametric mixed models for inference about a finite population mean from two-stage samples, *Survey Methodology*, 30, 209–218.
- Zheng, H., and R. Little (2005), Inference for the population total from probability-proportional-to-size samples based on predictions from a penalized spline nonparametric model, *Journal of Official Statistics*, 21, 1–20.