University of Massachusetts Amherst
ScholarWorks@UMass Amherst

Doctoral Dissertations | Dissertations and Theses

March 2022

Development of Bayesian temporal models for global health estimation in data-limited settings

Zhengfan Wang
University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/dissertations_2

Recommended Citation

Wang, Zhengfan, "Development of Bayesian temporal models for global health estimation in data-limited settings" (2022). *Doctoral Dissertations*. 2480.
https://doi.org/10.7275/26867436.0 https://scholarworks.umass.edu/dissertations_2/2480

This Open Access Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

DEVELOPMENT OF BAYESIAN TEMPORAL MODELS FOR GLOBAL HEALTH ESTIMATION IN DATA-LIMITED SETTINGS

A Dissertation Presented

by

ZHENGFAN WANG

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 2022

Biostatistics and Epidemiology

© Copyright by Zhengfan Wang 2022

All Rights Reserved

DEVELOPMENT OF BAYESIAN TEMPORAL MODELS FOR GLOBAL HEALTH ESTIMATION IN DATA-LIMITED SETTINGS

A Dissertation Presented

by

ZHENGFAN WANG

Approved as to style and content by:

Leontine Alkema, Chair

John Staudenmayer, Member

Jing Qian, Member

Lisa Chasan-Taber, Department Chair
Biostatistics and Epidemiology

ACKNOWLEDGMENTS

First of all, I would like to acknowledge my advisor, Leontine Alkema. My graduate research and dissertation were generously supported by her for 4 years. As an international student, there are many difficulties, challenges in study, daily life and career exploring. With the help of Leontine, I gradually developed a passion for Bayesian analysis and modeling. I am very proud to have her as my advisor. I would like to thank all my committee members Jing Qian, and John Staudenmayer for their comments and supports for my dissertation.

I would like to also thank all members of the Core Stillbirth Estimation Group and the members of the Technical Advisor Group of the UN Inter-agency Group for Child Mortality Estimation for providing input and feedback on estimating stillbirth. I would like to also thank Bruno Masquelier, Stephane Helleringer and Emily Peterson for their contribution on estimating survival probability in sibling survival history data.

I also want to thank members from our lab, Emily Peterson, Chuchu Wei, Herb Susmann, Guandong (Eliot) Yang, Greg Guranich and other participants who joined our working group meetings. Thanks for their opinions and discussion for my research projects. I also thank my friends, classmates, roommates (Yukun Li, Jingyao Hou, Yiding Zhang, Rui Zhang), and staff and professors who helped me a lot from the Department of Biostatistics and Epidemiology. I feel so lucky to have spent the years in such a friendly and collaborative lab and department. I also thank Xiangrong Kong who led and supported me as well.

Finally, I would like to thank my parents who always regard my health, education and happiness as the top priorities, and thank my significant other, Ruocheng Dong, for her company.

ABSTRACT

DEVELOPMENT OF BAYESIAN TEMPORAL MODELS FOR GLOBAL HEALTH ESTIMATION IN DATA-LIMITED SETTINGS

FEBRUARY 2022

ZHENGFAN WANG

B.Sc., BEIJING NORMAL UNIVERSITY & HONG KONG BAPTIST

UNIVERSITY UNITED INTERNATIONAL COLLEGE

M.Sc., GEORGETOWN UNIVERSITY

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Leontine Alkema

Estimation of health indicators globally is complicated because of data sparsity and data quality issues, especially in low and middle income countries without well-functioning registration systems. This dissertation introduces Bayesian methods for the estimation of stillbirth rates and adult mortality in data-limited settings.

Motivated by statistical challenges in the estimation of stillbirth rates globally, we develop a Bayesian hierarchical temporal sparse regression model (BHTSRM). Bayesian hierarchical temporal regression models combine a hierarchical regression model with a temporal smoothing process. This type of model has been used for estimating health indicators for multiple populations in data-sparse settings to track high-quality data while producing covariate-driven estimates for populations with

limited or no data. To extend its usage to settings where the number of candidate covariates is large relative to data availability, we propose the use of BHTSRMs that impose sparsity by using horseshoe priors on regression coefficients. We also develop a method to adjust observations with alternative stillbirth definitions and account for varying levels of uncertainty associated with different data sources in fitting the BHTSRM to stillbirth data. The proposed model has been used by the United Nations to estimate stillbirth rates globally.

To facilitate prediction based on BHTSRMs, we propose an associated variable selection method: horseshoe shrinkage parameter reference distribution variable selection (HSS-VS). We check the performance of the new method through simulation exercises and use it for variable selection in the estimation of stillbirth rates.

In low and middle income countries without well-functioning registration systems, sibling survival history (SSH) data can be used to estimate adult mortality but it may be subject to substantial reporting errors. We propose a new Bayesian survival model to estimate age-cohort specific survival probabilities from SSH data while accounting for bias and uncertainty introduced by SSH reporting errors. In the model, the cumulative hazard function is captured with a two-dimensional spline function. We apply it to estimate adult survival in Senegal.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS ..... **iv**

ABSTRACT ..... **vi**

LIST OF TABLES ..... **xi**

LIST OF FIGURES ..... **xiv**

CHAPTER

1. GLOBAL ESTIMATION AND PROJECTION OF STILLBIRTH RATE **1**

1.1 Introduction 2
1.2 Data 4

1.2.1 Database construction 4
1.2.2 Notation 5
1.2.3 Covariates 8

1.3 Exclusion based on the ratio of SBR to NMR 11

1.3.1 Predictive model for the SBR to NMR ratio 12
1.3.2 Exclusion procedure 14

1.4 Methods for SBR estimation 14

1.4.1 SBR estimation model summary 14
1.4.2 Estimation of data quality parameters 15

1.4.2.1 Variance of $\log(y_i)$ 15
1.4.2.2 Source type bias ψ_j and measurement error variance term σ_j^2 17

1.4.3 Definition adjustment 18

1.4.4 Bayesian hierarchical temporal sparse regression model 22
1.4.5 Computation 25
1.4.6 Model validation and comparison 25

1.5 Results 26

1.5.1 Data quality and data adjustments 26
1.5.2 Illustrative findings 27
1.5.3 Covariates 28
1.5.4 Model validation, comparison and sensitivity analyses 29

1.6 Discussion 32

2. SELECTION OF PREDICTORS BASED ON HORSESHOE PRIOR 35

2.1 Introduction 37
2.2 Summary of the BHTRM used for SBR estimation 38
2.3 Existing approaches for Bayesian variable selection 40

2.3.1 Expected log posterior predictive: approximate leave-one-out CV (loo) 41
2.3.2 Variable selection based on 95% credible intervals (CI) 42
2.3.3 Variable selection based on Projection method (proj) 42

2.4 HSS-VS method 44

2.4.1 Shrinkage measure ω_k for regression coefficient k in a sparse regression model 44
2.4.2 HSS-VS method details 45

2.4.2.1 Summary of implementation of HSS-VS method 46

2.5 Assessing performance of variable selection methods 48

2.5.1 Measure of prediction accuracy 48

2.6 Simulation study 50

2.6.1 Multiple regression 50
2.6.2 BHTRM 51

2.7 Case study: SBR estimation 55
2.8 Discussion 59

3. ESTIMATING SURVIVAL PROBABILITIES FROM SSH DATA 61

3.1 Introduction 61
3.2 Data 63

3.2.1 Health and Demographic Surveillance Data 64
3.2.2 Sibling Survival Histories 64
3.2.3 Data Summary 64

3.3 Methods 66

3.3.1 Objective and notation 66
3.3.2 Model for reporting errors 71
3.3.3 The B-surv Model 72

3.3.3.1 Likelihood function 72

3.3.4 B-Surv process model: parameterization of cumulative hazard function 74
3.3.5 Computation 77
3.3.6 Model evaluation and validation 78

3.3.6.1 Evaluating predictions of the birth year 78

3.3.7 Evaluating estimated survival parameters 79

3.4 Results 80

3.4.1 Reporting errors 80
3.4.2 B Surv model fits 81
3.4.3 B-surv model validation 84

3.5 Discussion 87

APPENDICES

A. SUPPLEMENTARY TABLES AND FIGURES FOR CHAPTER 1 88
B. SUPPLEMENTARY INFORMATION FOR CHAPTER 2 99
C. SUPPLEMENTARY TABLE AND FIGURES FOR CHAPTER 3 109

BIBLIOGRAPHY 113

LIST OF TABLES

Table **Page**

1.1 Data set $\mathcal{B}^-$ used for fitting the SBR estimation model by source and definition for countries in 2000-2019. For example, there are 75 countries with administrative data. "28 weeks" represents the standard definition. "22 weeks" and "24 weeks" represent 22/24 weeks or more of gestation; "500 grams" and "1000" grams represent birthweight $\geq$ 500/1000 grams. 8

1.2 Data availability in definition adjustment data sets $\mathcal{D}^d$. "22 weeks" and "24 weeks" represent 22/24 weeks or more of gestation; "500 grams" and "1000" grams represent birthweight $\geq$ 500/1000 grams. 18

1.3 Adjustments and standard deviation of alternative definition versus the 28 week definition. 26

1.4 Source type bias and source type standard deviation. 27

1.5 Estimates for regression coefficients under BHTSRM fit. Point estimates of regression coefficients, 95% credible interval given by the 2.5th and 97.5th percentiles of the posterior, and the standard deviation of the covariate prior to standardization. Details on covariates are given in Table A.3. 29

1.6 Validation results for SBR estimates. "N.test" represent the number of observations in the test set. Validation exercises "Recent", "Random", and "In-sample" represent leaving out recent observations, randomly leaving out 20% of all observations, and approximate leave-one-out validation, respectively. The outcome measures are as follows: mean of error, mean absolute error, and % of left-out observations below and above their respective 90% and 80% prediction intervals. Desirable* refers to outcomes for models that are unbiased and well calibrated. 30

1.7 Model comparison based on expected log pointwise predictive density and Pareto K diagnostic values. BHTSRM is our proposed model, and BHTRM is model with vague prior on covariates. The HS $\tau_0 = 0.001$ model stands for BHTSRM with $\tau_0 = 0.001$. Smooth1 and Smooth2 are two models with different settings of smoothers described in the text. When comparing models, larger ELPD value suggests better model performance. The percentage of high influential points (Pareto K values ¿ 1) for all models are presented in the "Pareto K diag." column, lower outcomes are preferred. 32

2.1 Linear regression simulation results. "ME" and "MAE" represent the mean error and mean absolute error. The values in "PIT(5%,10%,90%,95%)" are the desirable PIT values. "ACC", "FER" and "FIR" represent the accuracy rate, false exclusion rate, and false inclusion rate, respectively. 51

2.2 Summary of BHTRM simulation settings. "N", "σ_d^2", and ρ are number of observations, the three-group error variance, and correlation between covariates. "(K, K_1)" represents the number of candidate covariates K and relevant covariates K_1. $\beta^{(\neq 0)}$ represents the true coefficients. 52

2.3 Scenario 1: strong signals and median correlation 53

2.4 Scenario 2: weak signals and median correlation 54

2.5 Scenario 3: weak signals and high correlation 54

2.6 Scenario 4: strong signals and high correlation 54

2.7 Model comparison based on expected log pointwise predictive density and Pareto K diagnostic values. When comparing models, larger ELPD value suggests better model performance. If 95% CI for difference of ELPD contains 0, it suggests that the difference between two model is not significant. The percentage of high influential points (Pareto K values ¿ 1) for all models are presented in the "Pareto K diag." column, lower outcomes are preferred. 59

3.1 *Demographic characteristics of maternal siblings matched between SSH and HDSS data* $N = 1,294$. 65

3.2 *Summary statistics of reporting errors of matched siblings* 66

3.3 Posterior estimates of regression parameters: median estimate (Est.) and 95% credible intervals (95% CI). 81

3.4 The summary of residual for estimate birth year and survival time. PIT(5%, 95%) represents the PIT with the expected values 5% and 95%. $\text{cov}_{95\%}$ represents the 95% coverage rate. 85

A.1 Overview of estimates for regression coefficients under reference BHTSRM and BHTRM. "estimate" and "sd" represent the median and standard deviation of posterior samples. 91

A.2 Validation result aggregated across all countries (top) and broken down by high and low income countries (bottom). "N.test" represent the number of observations in the test set. Validation exercises "Recent" and "In-sample" represent leaving out recent observations, and approximate leave-one-out validation. The outcome measures are as follows: mean of error, mean absolute error, and % of left-out observations below (be) and above (ab) their respective 90% and 80% prediction intervals. "BHTRM" validation are results for models with vague prior on β. Desirable* refers to outcomes for models that are unbiased and well calibrated. 92

A.3 Candidate covariates with its sources and methodology 93

C.1 Descriptions of mathematical notations 109

LIST OF FIGURES

Figure **Page**

1.1 Data sets and exclusion steps. This chart summarizes the data sets used for estimating the SBR. Data sets are indicated in rectangle boxes and the processing steps are summarized by the thick arrows. The global SBR data set $\mathcal{B}$ consists of administrative data ("Admin.") , HMIS, survey and population studies ("Pop. study") including population-based prospectively data ("PopPros") and additional data ("PopLR"). 6

1.2 SBR data and estimates for 2000-2019 for selected countries. Posterior median point estimates from BHTSRM (red line) with 90% credible intervals (red area), and covariate-based estimates (dashed green line) with 90% credible intervals (green area) are shown. Observed but unadjusted observations are displayed as hollow symbols. Adjusted data (based on definition adjustments and accounting for survey biases where applicable) and data that do not require adjustments (non-survey data with standard definition) are shown for all source types. Colors indicate the definition of the observation. Error bars displayed with adjusted observations indicate 95% confidence interval of the SBR based on the observation, accounting for its estimated bias and error variance. Note that the y-axis varies across countries, and that data excluded based on the data quality assessment are not shown. 11

1.3 SBR to NMR ratio exclusion process. This chart summarizes the 2-step exclusion process based on SBR:NMR ratios. The thin arrows indicate the flow of data and parameters. 12

1.4 SBR estimation model overview. This chart summarizes the inputs and set up of the SBR estimation model. The input data includes SBR data set $\mathcal{B}^{-}$, covariate data set $\mathcal{C}$, and point estimates from the definition adjustment model (see Section 1.4.3). 16

2.1 Illustration of reference distribution Ω. The black line is the density and the dashed line is the threshold Ω_α. The estimated shrinkage parameters for two example candidate covariates are given by the circle with shrinkage parameter ω_1 and cross ω_2. Because $\omega_1 < \Omega_\alpha$ and $\omega_2 > \Omega_\alpha$, the two covariates are selected as signal and noise respectively. 47

2.2 Estimates of regression coefficients for full HS model, and for subsetted models obtained by different variable selection methods. 56

2.3 Standardized differences between SBR estimates for all countries obtained from the BHTSRM and subsetted models (y-axis), plotted against the SBR estimates from the BHTSRM. The variable selection method is given in the title of each plot. The regions are represented by colors. 57

2.4 SBR estimates (per 1,000 births) of different models for Afghanistan, Ireland, Malawi and Ukraine. The 95% credible interval of full HS model is shown by grey. 58

3.1 Lexis diagram illustration information collected from SSH. x-axis represents the calendar year. For alive sibling i where $s_i^{(ssh)}(t^*) = 1$, the current age is reported. For deceased sibling j where $s_j^{(ssh)}(t^*) = 0$, the age at death and time since death $d_j^{(ssh)}$ are collected in the survey. 67

3.2 Information on reporting errors, alive sibling i (Left) and deceased sibling j (Right). x-axis represents the calendar year. For alive sibling i, $s_i^{(ssh)}(t^*) = s_i(t^*) = 1$, and birth year error $e_i^{(birth)} = b_i^{(ssh)} - b_i$. For deceased sibling j, $s_j^{(ssh)}(t^*) = s_j(t^*) = 0$, the time since death error is $e_j^{(death)} = d_j^{(ssh)} - d_j$. 68

3.3 Graphical representation of the Bayesian hierarchical two-stage SSH error and survival models. Blue rectangle denotes SSH error model, green rectangle denotes outputs obtained from Stage I, used as inputs in Stage II, and red rectangle denotes SSH survival model). Solid arrows denote stochastic dependency. Hollow boxes contain observed data. 70

3.4 Spline coefficients α, capturing changes across cohorts and time. 76

3.5 Model comparison between fit to HDSS data and fit to SSH data while accounting for errors. 82

3.6 Model comparison between fit to SSH data and fit to SSH data without accounting for errors. 83

3.7 Model comparison between fit to HDSS data and fit to SSH data without accounting for errors. 84

3.8 Residual check for estimated birth year B_i and estimated survival time T_i since t_0 for deceased sibling. 85

3.9 An in-sample probability integral transform (PIT) of birth year and survival time in model fit to SSH data while accounting for errors. 86

3.10 Probability integral transform check of survival probability for deceased siblings given HDSS reported birth year b_i and HDSS reported survival time t_i. 86

A.1 Histogram of estimates of θ_i. There are 73 observations. 89

A.2 Estimates of β and 95% CI for reference BHTSRM and BHTRM. 90

C.1 Exploratory data analysis. Distribution of birth year error and time since death error. 110

C.2 Exploratory data analysis. Relationship between errors and covariates. 111

C.3 Residual plot from SSH error model. 112

CHAPTER 1

GLOBAL ESTIMATION AND PROJECTION OF STILLBIRTH RATE

Estimation of stillbirth rates globally is complicated because of the paucity of reliable data from countries where most stillbirths occur. We compiled data and developed a Bayesian hierarchical temporal sparse regression model for estimating stillbirth rates for 195 countries from 2000 to 2019. The model combines covariates with a temporal smoothing process so that estimates are data-driven in country-periods with high-quality data and determined by covariates for country-periods with limited or no data. Horseshoe priors are used to encourage sparseness. The model adjusts observations with alternative stillbirth definitions and accounts for various sources of uncertainty. In-sample goodness of fit and out-of-sample validation results suggest that the model is reasonably well calibrated. The model is used by the UN Inter-agency Group for Child Mortality Estimation to monitor the stillbirth rate for 195 countries.

1.1 Introduction

The United Nations Inter-agency Group for Child Mortality Estimation (UN IGME) defines a stillbirth as a baby born with no signs of life at 28 weeks or more of gestation (UN Inter-agency Group for Child Mortality Estimation [2020]), consistent with the International Classification of Diseases (ICD-11, World Health Organization [2019]) definition of a "late gestation fetal death". Prior estimates highlighted the large global burden of stillbirths with an estimated 2.6 million stillbirths for the year 2015 (Blencowe et al. [2016]). Ending preventable stillbirths is one of the core goals of the UN's Global Strategy for Women's, Children's and Adolescents' Health from 2016 until 2030 (Kuruvilla et al. [2016]) and the Every Newborn Action Plan (ENAP, World Health Organization [2014]). These global initiatives aim to reduce the stillbirth rate (SBR, the number of stillbirths per 1,000 total births) to 12 or fewer stillbirths per 1,000 births in every country by 2030.

Monitoring of SBRs is challenging because of data paucity in countries where most stillbirths occur. Estimates of SBRs for a country can be derived from administrative data from registration systems (e.g., civil registration and vital statistics (CRVS) and medical birth and death registries). The reliability of SBR estimates from such data sources depends on the accuracy and completeness of reporting and recording of stillbirths and live births. Not all countries maintain an accurate, timely, and complete registration system for stillbirths. Moreover, in many low- and middle-income countries (LMICs), stillbirths are not reported in registration systems at all. For such countries, stillbirth data can be obtained from health management information systems (HMIS), with limitations similar to the registration systems: stillbirth data from registries and HMIS may be reported in different stillbirth definitions, and may be biased due to underreporting, misclassification, and other data quality issues. Lastly, SBR data can be obtained from household surveys and population-based studies but – in addition to limitations similar to the other data sources regarding definitions –

these data are typically not available for all years of interest and may be subject to potentially large biases and/or non-sampling errors.

Blencowe et al. [2016] produced estimates of the SBR for all countries, from 2000 to 2015. Yearly estimates for developed countries with high-quality data were obtained from the data directly, using a Loess smoother. Estimates for all other countries were obtained from a regression model with country-specific intercepts and global regression coefficients. The main limitation of this work is the use of the regression model for countries with limited data: resulting trend estimates are covariate-driven, even if available data suggest deviations away from covariate-predicted trends. In addition, a stepwise approach was taken to carry out variable selection, which underestimates uncertainty since the model selection process is not accounted for.

In this paper, we propose a new approach to estimating the SBR for all countries, using a Bayesian hierarchical temporal sparse regression model (BHTSRM). The model is used by the UN IGME to monitor the SBR globally (UN Inter-agency Group for Child Mortality Estimation [2020], Hug et al. [2021]). Our approach updates and extends the work of Blencowe et al. [2016]. As its name implies, BHTSRM combines a hierarchical regression model with a temporal smoothing process. This type of model produces estimates that track high-quality data while producing covariate-driven trend estimates for countries with limited or no SBR data. While this kind of model has been used for estimating global health indicators in other settings, e.g., in Alkema et al. [2017], prior work does not address sparsity. Here we extend upon previous work by introducing sparsity-inducing priors for estimating regression coefficients. In particular, we use horseshoe priors (Piironen and Vehtari [2017b]) to shrink the less important coefficients toward zero, which makes BHTSRM an approach that can deal with a large number of covariates.

As compared to Blencowe et al. [2016], our proposed model also introduces new statistical approaches to address various data quality issues. Firstly, we propose a

statistical procedure for data exclusion based on comparing observed ratios of SBR to the neonatal mortality rate (NMR). Secondly, we add to the model an estimation approach to incorporate observations with alternative definitions of a stillbirth (e.g., based on 22 weeks gestational age or 1000 grams birthweight) while accounting for the additional uncertainty associated with such observations.

This paper is organized as follows: in Section 3.2, we provide an overview of data sources and definitions that are available for measuring SBR. We introduce the exclusion of data based on the ratio of SBR to NMR in Section 1.3. We describe the SBR estimation model in Section 1.4, including the BHTSRM. In Section 3.4, we present estimates of SBR, data quality parameters and validation results. Last, we conclude with a discussion of limitations and future research directions in Section 3.5.

1.2 Data

1.2.1 Database construction

SBR data were compiled by the UN IGME from various sources for the year 2000 and onwards. The majority of data collected on stillbirths were obtained from administrative data systems and health management information systems (HMIS). UN IGME conducts an annual country consultation to solicit up-to-date administrative data on stillbirths from ministries of health or national statistics offices. Population-based study data were obtained from a review of the academic literature and a WHO data call to maternal-newborn health experts. Nationally representative household surveys (e.g., Demographic and Health Surveys, Multiple Indicator Cluster Surveys, Reproductive Health Surveys) are another source of stillbirth data.

After data were compiled, general exclusion rules were applied. The evaluation and assessment for data quality were applied to all data sources based on pre-defined exclusion criteria. Data were excluded if they lacked information on definition or data collection systems, if the proportion of reported stillbirths with unknown gesta-

tional age or birthweight was above 50 per cent, if data were internally inconsistent, or if coverage of live births in administrative data systems was estimated as below 80 per cent. Vital registration data with incomplete coverage of child deaths were also excluded, where incompleteness was taken from the WHO CRVS completeness assessment (WHO Department of Information, Evidence, and Research [2018]).

1.2.2 Notation

We use lowercase Greek letters for unknown parameters and uppercase Greek letters for variables which are functions of unknown parameters (modeled estimates). Roman letters indicate variables that are known or fixed, including data (in lowercase) and estimates provided by other sources or the literature (in uppercase).

Data compilation and general exclusion resulted in a global database of observed SBR values. Observations are available across countries over time and are indexed by i; For each i, $c[i]$ refers to the country for which the i-th observation was recorded, and $t[i]$ to the calendar year of observation i. Index $j[i]$ is used to refer to the source category of observation i. We define an observed value y_i as the SBR calculated from the number of reported stillbirths z_i and number of live births q_i from a given source for a country-period with $y_i = z_i/(z_i + q_i)$. Periods referring to calendar years when available, or longer if the source does not provide information on annual SBR. In the database, data source types are categorized as (1) administrative data; (2) HMIS data; (3) household survey data; and (4) population-based study data. Among population-based studies, we distinguish between population-based prospectively-collected data with recruitment prior to 28 weeks of gestation, and follow-up to at least 28 days for live births, referred to here as PopPros data (Bose et al. [2015], Ahmed et al. [2018]) and additional data (PopLR).

We denote the set of all available observations resulting after the general exclusion step as $\mathcal{B}$. The data set $\mathcal{B}$ forms the basis of all analyses, as outlined in Figure 1.

Firstly, an exclusion procedure is introduced for observations in the global data set $\mathcal{B}$ based on the ratio of SBR y_i to NMR o_i. The NMR o_i is calculated from the number of neonatal deaths m_i and number of live births q_i with $o_i = m_i/q_i$. The ratio of SBR to NMR is analyzed using the PopPros data set $\mathcal{P}$. The details of the exclusion are described in Section 1.3. Other subsets of data set $\mathcal{B}$ are used for fitting the definition adjustment model and the SBR estimation model.

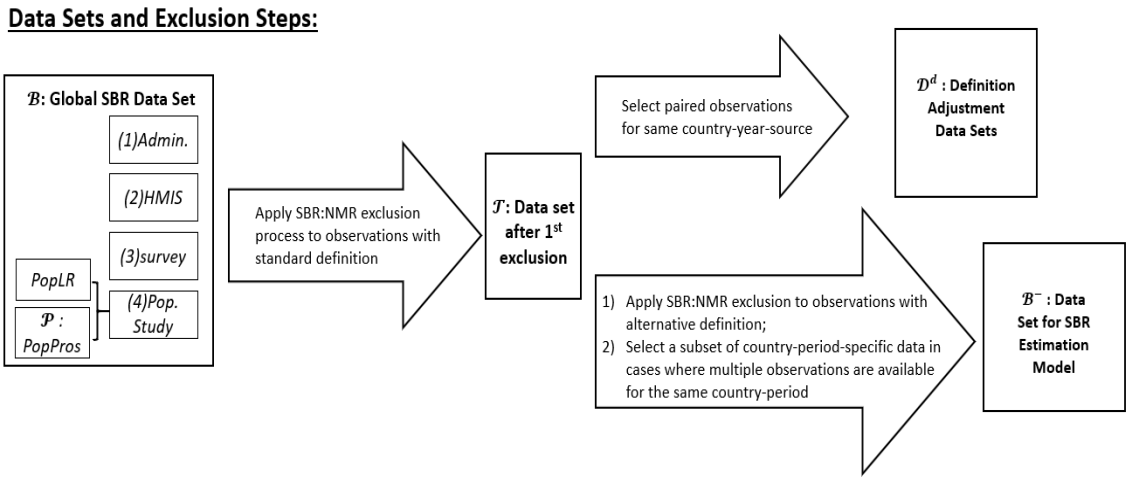


Figure 1.1. Data sets and exclusion steps. This chart summarizes the data sets used for estimating the SBR. Data sets are indicated in rectangle boxes and the processing steps are summarized by the thick arrows. The global SBR data set $\mathcal{B}$ consists of administrative data ("Admin.") , HMIS, survey and population studies ("Pop. study") including population-based prospectively data ("PopPros") and additional data ("PopLR").

To allow for international comparison, we focus on estimating SBRs reported using the standard definition (gestational age $\geq$ 28 weeks). In fitting the SBR model, we used data based on the standard definition when available. However, for a subset of country-periods in $\mathcal{B}$, stillbirths were reported using an alternative definition only, based on birthweight or a different gestational age cut-off. Four kinds of alternative definitions are incorporated in the analysis: definitions referring to a baby born with no signs of life at (1) 24 weeks or more of gestation; (2) 22 weeks or more; (3) birthweight $\geq$ 1000 grams; and (4) birthweight $\geq$ 500 grams. To use these observations

for estimating the SBR, we estimated adjustments and uncertainties associated with alternative definition d using the definition adjustment data set $\mathcal{D}^d$. The data set and definition adjustment model are given in Section 1.4.3.

We denote the subset of observations used for SBR estimation by $\mathcal{B}^-$. This database is obtained after (i) excluding observations that are identified as outlying based on the SBR to NMR ratio exclusion approach, and (ii) selecting a subset of country-period-specific data in cases where multiple observations are available for the same country-period, see Figure 1.1. The approach in (ii) is as follows: (iia) we keep observations from non-administrative data only if administrative data are not available and (iib) if observations are recorded in multiple definitions, we select only one definition based on the following order of preference: (1) standard definition, (2) birthweight $\geq$ 1000 grams; (3) 22 weeks or more of gestation; (4) 24 weeks or more of gestation; (5) birthweight $\geq$ 500 grams. There are 1531 observations from 133 countries in this SBR model data set $\mathcal{B}^-$. Table 1.1 summarizes the breakdown of observations based on definition and source.

Data availability is illustrated for selected countries in Figure 1.2. Data availability ranges in the selected countries from no included data in Afghanistan to an annual time series of national administrative data based on the standard 28 weeks definition for Ireland. Botswana, Malawi, Uganda and Ukraine are examples of countries with SBR data from multiple sources, available for selected periods only. In Ukraine, SBR data are available from 2007 to 2017 from administrative systems but recorded using 22 weeks definition. In Uganda, the only available data comes from surveys and population-based studies. In Malawi, available data sources are HMIS, population-based studies, and household surveys.

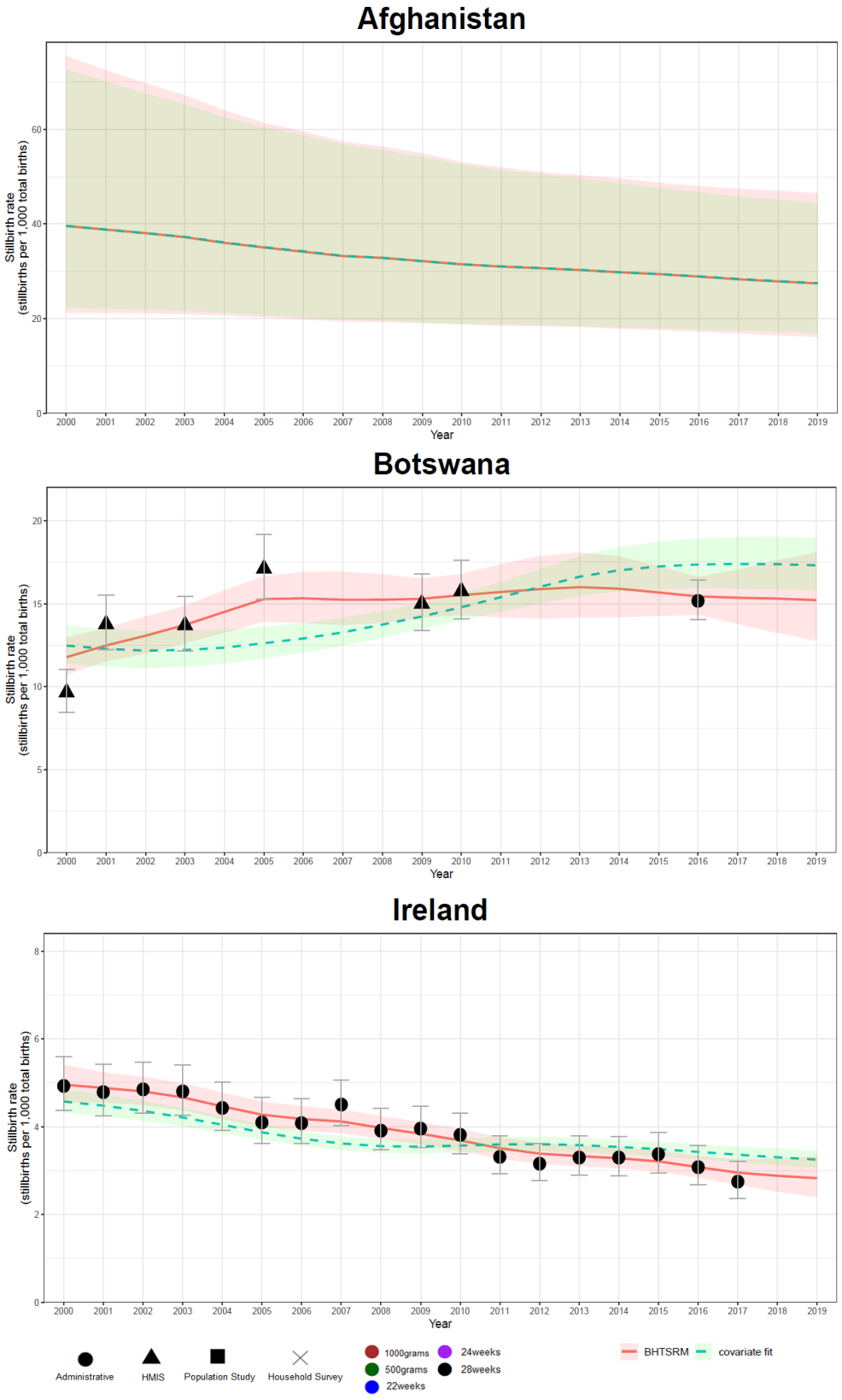
1.2.3 Covariates

Blencowe et al. [2016] identified a large number of candidate covariates for estimating SBR based on a conceptual framework. The framework includes distal determinants such as socio-economic factors, demographic and biomedical factors, associated perinatal outcome markers, and access to health care. Covariate database $\mathcal{C}$ contains information on the 16 covariates for all country-years. Further details are given in Supplementary Table A.3.

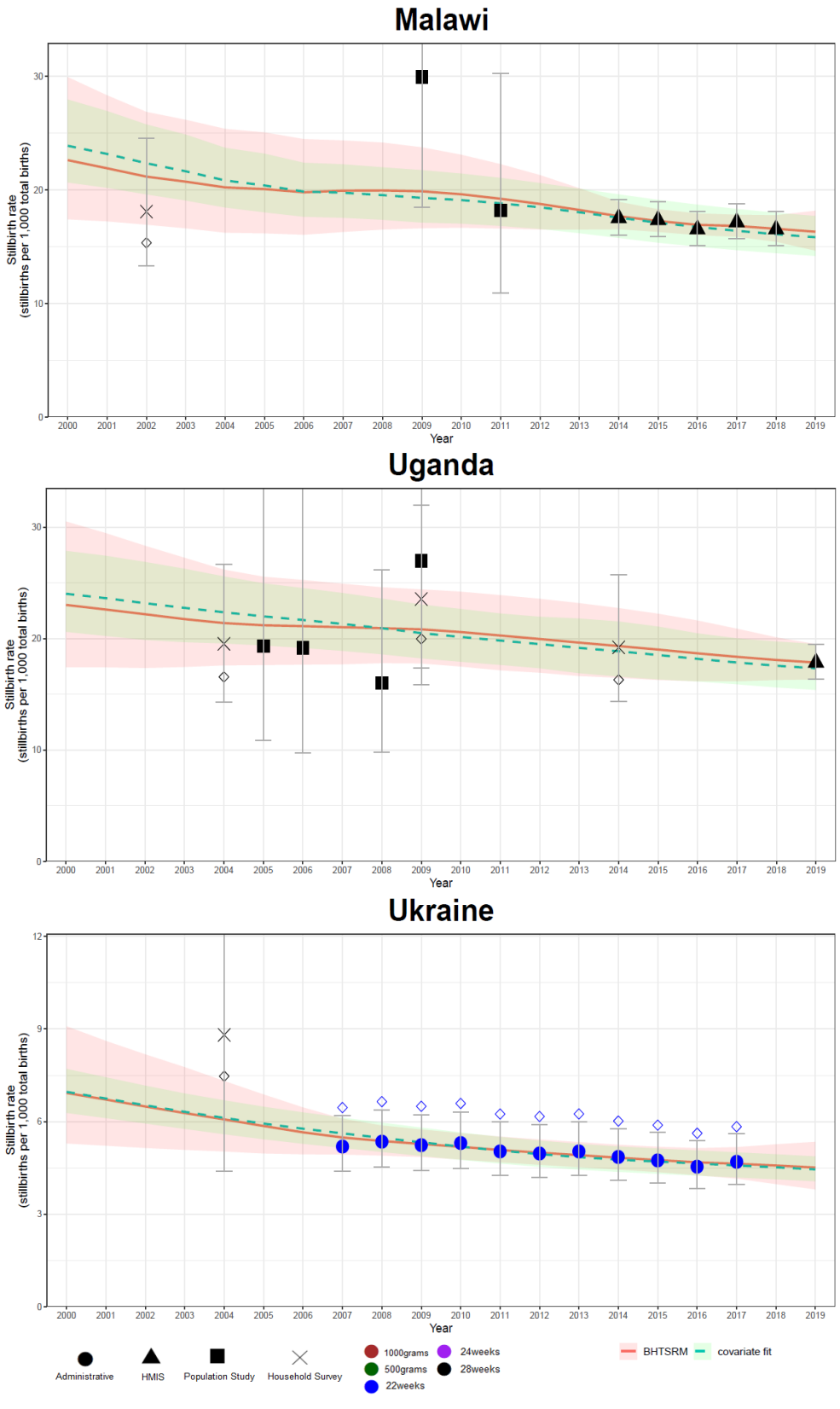
Data Source	Number of Countries	Number of Obs
Administrative	75	1157
HMIS	26	162
Household Survey	44	95
Population Study	23	117

Definition	Number of Countries	Number of Obs
28 weeks	124	1220
24 weeks	3	44
22 weeks	15	85
1000 grams	20	146
500 grams	5	36

Table 1.1. Data set $\mathcal{B}^-$ used for fitting the SBR estimation model by source and definition for countries in 2000-2019. For example, there are 75 countries with administrative data. "28 weeks" represents the standard definition. "22 weeks" and "24 weeks" represent 22/24 weeks or more of gestation; "500 grams" and "1000" grams represent birthweight $\geq$ 500/1000 grams.



– Figure continued on next page –



– Figure caption on next page –

Figure 1.2. SBR data and estimates for 2000-2019 for selected countries. Posterior median point estimates from BHTSRM (red line) with 90% credible intervals (red area), and covariate-based estimates (dashed green line) with 90% credible intervals (green area) are shown. Observed but unadjusted observations are displayed as hollow symbols. Adjusted data (based on definition adjustments and accounting for survey biases where applicable) and data that do not require adjustments (non-survey data with standard definition) are shown for all source types. Colors indicate the definition of the observation. Error bars displayed with adjusted observations indicate 95% confidence interval of the SBR based on the observation, accounting for its estimated bias and error variance. Note that the y-axis varies across countries, and that data excluded based on the data quality assessment are not shown.

1.3 Exclusion based on the ratio of SBR to NMR

Stillbirths are typically more poorly recorded than deaths of liveborn neonates, which are themselves under-recorded in many settings (Stanton et al. [2006] and Woods [2008]). We exclude data points whose stillbirths are likely to be underreported based on the ratio of observed SBR to NMR, making use of the fact that in settings where stillbirth case ascertainment is poor, the ratio of SBR to NMR is expected to be low.

We describe the approach in detail in the remainder of this section. In summary, we assume that each observed log-ratio is the sum of a setting-specific expected log-ratio and random error. We use the PopPros database $\mathcal{P}$ to build a model for the expected log-ratio. We then calculate observed log-ratios for all observations in the global data set $\mathcal{B}$ and exclude observations that – based on a comparison between the observation and its predictive distribution using the model for the expected log-ratio – are deemed subject to underreporting. The exclusion process is summarized in Figure 1.3.

The proposed approach improves upon the approach used previously for SBR estimation in Blencowe et al. [2016]. In the previously used approach, observations were excluded based on a percentile of the observed distribution of SBR to NMR

ratios. This approach did not account for varying uncertainty associated with the observed ratios and – contrary to our approach – the previous approach did not make explicit the probability of a false exclusion.

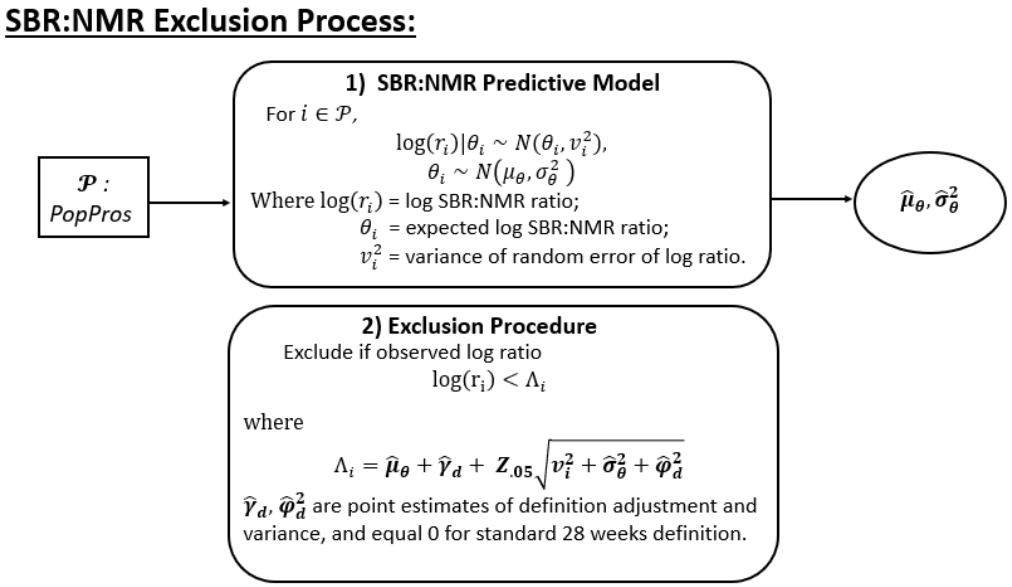


Figure 1.3. SBR to NMR ratio exclusion process. This chart summarizes the 2-step exclusion process based on SBR:NMR ratios. The thin arrows indicate the flow of data and parameters.

1.3.1 Predictive model for the SBR to NMR ratio

In the predictive model for the SBR to NMR ratio, we assume that each observed log-ratio is the sum of a setting-specific expected log-ratio and random error. This model is specified as follows. Let $r_i = y_i/o_i$ denote the observed ratio of SBR y_i to NMR o_i. We assume that

$$\log(r_i)|\theta_i \quad \sim \quad N(\theta_i, v_i^2), \tag{1.1}$$

where $\theta_i = E(\log(r_i))$ refers to the expected log-ratio of SBR to NMR and v_i^2 refers to the error variance.

The error variance v_i^2 is calculated using a Monte Carlo approximation. Specifically, denote z_i as the number of observed stillbirths and m_i as neonatal deaths. Then we have:

$$\begin{aligned} z_i|y_i &\sim Bin(g_i, y_i), \\ m_i|o_i &\sim Bin(q_i, o_i), \end{aligned}$$

where g_i refers to total births and q_i refers to the number of live births. Assuming independence between stillbirths and neonatal deaths, we obtain samples $(z_i^{(s)}, m_i^{(s)})$ and calculate the associated ratio $r_i^{(s)}$:

$$r_i^{(s)} = \frac{z_i^{(s)}/g_i}{m_i^{(s)}/q_i}.$$

The variance v_i^2 is given by the empirical variance of the samples $\log(r_i^{(s)})$.

We specify the distribution of the expected log-ratios θ_i as follows: assuming conditionally independence and a normal distribution, we set

$$\theta_i|\mu_\theta, \sigma_\theta^2 \sim N(\mu_\theta, \sigma_\theta^2), \tag{1.2}$$

with μ_θ referring to the mean log-ratio across different SBR and NMR settings and σ_θ^2 referring to variability across settings. We assign vague priors to μ_θ and σ_θ^2.

The model is fitted to data from PopPros data set $\mathcal{P}$. Based on the data collection procedure used by the studies in this data set, data are assumed to be based on complete reporting of stillbirths. The data set contains 73 data points from 10 LMICs in different years. Based on the data set, the estimated mean ratio on the log scale is $\hat{\mu_\theta} = -0.180$ (-0.250, -0.111) and variance across settings is estimated as $\hat{\sigma}_\theta^2 = 0.083$. The estimates of θ_i are shown in Appendix Figure A.1.

1.3.2 Exclusion procedure

If stillbirths are underreported relative to neonatal deaths for a specific observation, its associated observed log-ratio of SBR to NMR $\log(r_i)$ is biased downwards as compared to the true log-ratio θ_i. We calculate observed SBR to NMR ratios for all observations in data set $\mathcal{B}$ and use the fitted model described above to construct a predictive distribution for each log-ratio. We exclude an observation if its observed ratio is less than the 5% lower bound of its corresponding predictive distribution of the SBR to NMR ratio. Specifically, the predictive distribution of the SBR to NMR ratio for the i-th observation follows from Equations 1.2 and is given by

$$\log(r_i) \sim N(\hat{\mu}_\theta, \hat{\sigma}_\theta^2 + v_i^2),$$

where $\hat{\mu}_\theta$ and $\hat{\sigma}_\theta^2$ refer to point estimates for the mean and across-setting variance of θ and v_i^2 to the error variance of the log-ratio specific to that observation. Let Λ_i denote the lower 5% quantile of the predictive distribution for observation i, $\Lambda_i = \hat{\mu}_\theta + z_{.05}\sqrt{\hat{\sigma}_\theta^2 + v_i^2}$. We exclude observation i if its observed log ratio $\log(r_i) < \Lambda_i$. Based on the point estimates of μ_θ and σ_θ^2, the 5% lower bound of the predictive distribution of the SBR to NMR ratio is $\exp(\Lambda_i) = 0.52$ for observations with variance $v_i = 0$. For the data with alternative stillbirth definitions, we apply the exclusion procedure after definition adjustment (see Section 1.4.3).

1.4 Methods for SBR estimation

1.4.1 SBR estimation model summary

The SBR estimation model is summarized in Figure 1.4. We let $\Omega_{c,t}$ denote the main outcome of interest, which is the SBR for country c in year t using the standard definition. The process model specification, referring to the specification of $\Theta_{c,t} = \log(\Omega_{c,t})$, is explained in Section 1.4.4.

$\Omega_{c,t}$ is estimated using data set $\mathcal{B}^{-}$. Following earlier notation, observations are available across countries over time and are indexed by i; $c[i]$ refers to the country for which the i-th observation was recorded, $t[i]$ the calendar year of the observation, $j[i]$ the data source type of the observation, and $d[i]$ to its stillbirth definition. The index $r[c]$ refers to the region of country c. The data model is

$$\log(y_i)|\Theta_{c[i],t[i]}, \psi_{j[i]}, \sigma^2_{j[i]} \sim N(\Theta_{c[i],t[i]} + \psi_{j[i]} + \hat{\gamma}_{d[i]}, s_i^2 + \hat{\varphi}^2_{d[i]} + \sigma^2_{j[i]}), \tag{1.3}$$

where $\Theta_{c,t} = \log(\Omega_{c,t})$ refers to the log-transformed true SBR $\Omega_{c,t}$ for that country-year, s_i^2 to variance of $\log(y_i)$ (see Section 1.4.2.1), $\psi_{j[i]}$ and $\sigma^2_{j[i]}$ refer to its source type-specific bias and variance respectively (see Section 1.4.2.2), and $\hat{\gamma}_d$ and $\hat{\varphi}^2_d$ to definition-specific adjustment and variance for observations that are reported using alternative definitions.

Definition adjustment parameters are estimated prior to model fitting. As compared to the approach used previously in Blencowe et al. [2016], we have made two improvements. Firstly, we developed predictive models for the differences in SBRs that capture how stillbirths based on the alternative definition relate to stillbirths reported according to the standard definition. Secondly, we assess the variability in the relationship between standard and alternative SBRs and account for this uncertainty in the SBR estimation model. The approach is described in Section 1.4.3.

1.4.2 Estimation of data quality parameters

1.4.2.1 Variance of $\log(y_i)$

The term s_i^2 in the data model Eq. (2.1) refers to the variance of $\log(y_i)$. For observations administrative data, HMIS, and population studies, we assume a Poisson data-generating process to obtain s_i^2. Specifically, for SBR rate $y_i = z_i/g_i$, with stillbirth

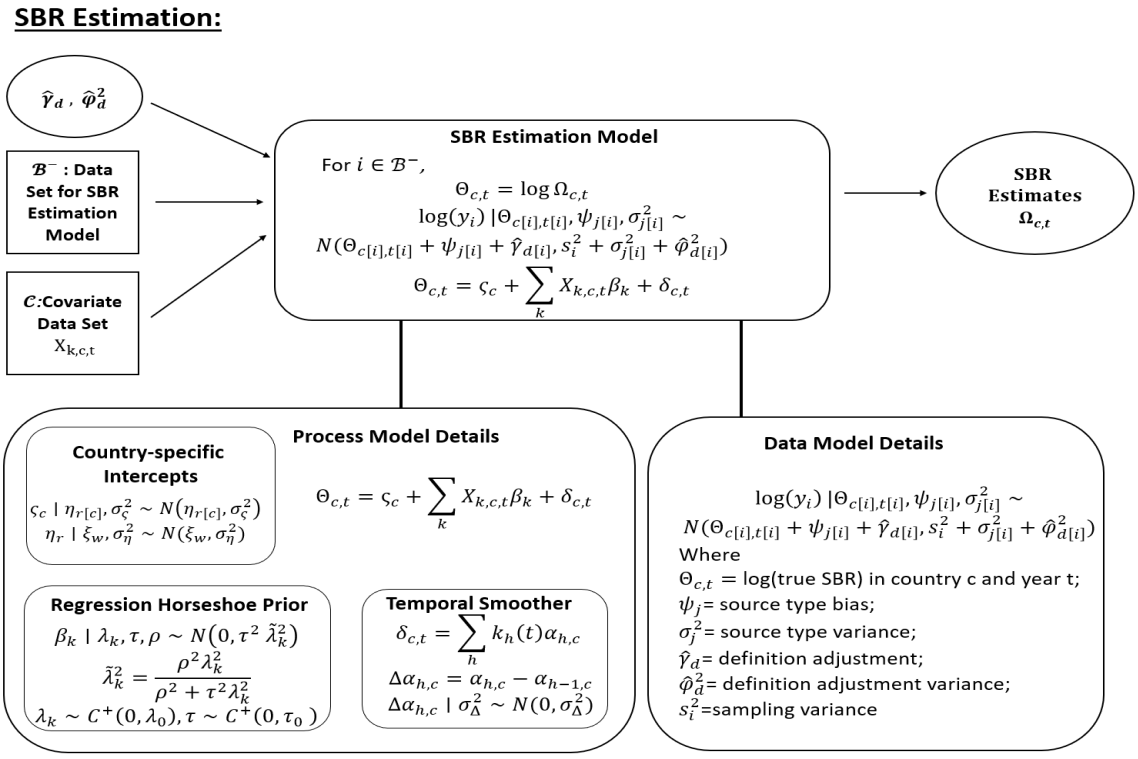


Figure 1.4. SBR estimation model overview. This chart summarizes the inputs and set up of the SBR estimation model. The input data includes SBR data set $\mathcal{B}^-$, covariate data set $\mathcal{C}$, and point estimates from the definition adjustment model (see Section 1.4.3).

z_i and total births g_i for the i-th observation, we assume $z_i \mid \Omega_i \sim Poisson(g_i \cdot \Omega_i)$. Then $var(y_i) = z_i/g_i^2$ and by using the delta method, we obtain:

$$\hat{var}(\log(y_i)) = \frac{1}{z_i \cdot y_i}. \tag{1.4}$$

Therefore, the variance s_i^2 for the i-th observation is set to $\frac{1}{z_i \cdot y_i}$. For observations from surveys, sampling error s_i is pre-calculated using a jackknife method (Pedersen and Liu [2012]), to reflect the survey sampling design.

1.4.2.2 Source type bias ψ_j and measurement error variance term σ_j^2

Source type bias terms ψ_j are included in model fitting to capture systematic biases associated with specific source types. We assume there is no source type biases for administrative, HMIS, and population-based studies, i.e. $\psi_j = 0$ for j referring to these three source types

$$\psi_{1,2,4} \quad = \quad 0.$$

Liu et al. [2016] and Bradley et al. [2015] suggest that stillbirths tend to be underreported in surveys, so we assume that data from surveys have a negative bias term and estimate this bias term, assigning a a half-normal vague prior to ψ_3:

$$\psi_3 \quad \sim \quad N^-(0, 5^2). \tag{1.5}$$

The measurement error variance term σ_j^2 captures non-systematic errors due to errors introduced in reporting. These variance parameters are estimated and assigned vague priors

$$\sigma_j \quad \sim \quad N^+(0, 1), j = 1, ..., 4$$

Definition	Income Group	Number of Countries	Number of Obs
22 weeks	Low	14	59
22 weeks	High	34	369
24 weeks	High	8	28
1000 grams	High	34	477
500 grams	High	30	355

Table 1.2. Data availability in definition adjustment data sets $\mathcal{D}^d$. "22 weeks" and "24 weeks" represent 22/24 weeks or more of gestation; "500 grams" and "1000" grams represent birthweight $\geq$ 500/1000 grams.

1.4.3 Definition adjustment

To estimate the definition-specific adjustment γ_d and variance φ_d^2 in Eq. (2.1), we use data sources that reported stillbirths using multiple definitions. Specifically, we construct definition adjustment data set $\mathcal{D}^d$ for each definition d, which contains all available paired observations of stillbirth counts $(z_i^{(d)}, z_i)$, where $z_i^{(d)}$ is the number of stillbirths under the alternative definition d, z_i is the number of stillbirths under standard definition, and the pair refers to the same source, country, and year. We use the paired counts to estimate γ_d and φ_d^2 for definition d without controlling for year and source, but separately for high-income countries (HICs) and LMICs. Due to lack of data, in LMICs we assume that 500 grams birthweight is equivalent to 22 weeks of gestational age, and 1000 grams birthweight is equivalent to 28 weeks of gestational age. Table 1.2 summarizes the data used for the analysis of the definition and income group combinations.

We define $\kappa_i^{(d)}$ as the log-ratio of the SBR as per alternative definition d to standard definition for observation i: $\kappa_i^{(d)} = \log(\left(\frac{\Omega_{c[i],t[i]}^{(d)}}{\Omega_{c[i],t[i]}}\right)$. With this definition of κ, the true log-transformed SBR for observation i under definition $d[i]$, $\Theta_{c[i],t[i]}^{(d)}$, can be written as

$$\Theta_{c[i],t[i]}^{(d)} = \Theta_{c[i],t[i]} + \kappa_i^{(d)}, \tag{1.6}$$

where $\Theta_{c,t}$ refers to the log-transformed SBR under the standard definition. We use this relation to define the adjustment term γ_d and variance φ_d^2 in Eq. (2.1): we set the adjustment γ_d and variance φ_d^2 in the SBR data model equal to the posterior median and variance of the predictive distribution for $\kappa_i^{(d)}$ for each alternative definition d.

In the derivation of the predictive density of κ, we approximate the log-ratio of SBRs $\kappa_i^{(d)}$ by the ratio of stillbirths, justified by the fact that the number of stillbirths are small relative to live births. Specifically, the true SBR for alternative definition d can be written as follows: $\Omega_{c,t}^{(d)} = \frac{\Upsilon_{c,t}^{(d)}}{q_{c,t}+\Upsilon_{c,t}^{(d)}}$, where $\Upsilon_{c,t}^{(d)}$ refers to the "true" stillbirth count associated with the true SBR under alternative definition d, and $q_{c,t}$ the number of live births. Given that $\Upsilon_{c,t}^{(d)} << q_{c,t}$, we approximate κ as follows:

$$\kappa_i^{(d)} = \log\left(\frac{\Omega_{c[i],t[i]}^{(d)}}{\Omega_{c[i],t[i]}}\right) \approx \log\left(\frac{\Upsilon_{c[i],t[i]}^{(d)}}{\Upsilon_{c[i],t[i]})}\right). \tag{1.7}$$

The assumptions made to obtain the predictive distribution for κ varies by the definition. Alternative definitions fall into two categories: definitions containing the standard definition and definitions overlapping with the standard definition. We consider each of these below. In each set up, we work towards providing a predictive distribution for $\kappa_i^{(d)}$ by introducing probabilities that relate the survival based on the alternative definition to that based on the standard definition.

Definitions containing the standard 28 weeks definition

Stillbirths z_i recorded using the 28 weeks definition are a subset of stillbirths recorded using the 22 or 24 weeks definitions, $z_i \leq z_i^{(d)}$ for d referring to 22 and 24 weeks. Given that stillbirths based on 22 or 24 weeks definitions contain those with 28 weeks definitions, we may assume

$$z_i|\omega_i^{(d)} \sim \text{Binomial}(z_i^{(d)}, \omega_i^{(d)}), \tag{1.8}$$

where $\omega^{(d)}$ is the definition-specific probability of a stillbirth with gestational age beyond 28 weeks conditional on being dead after 22 or 24 weeks. The probability $\omega_i^{(d)}$ relates to $\kappa_i^{(d)}$ as follows (as per Eq. (1.7) and the definition of $\omega_i^{(d)}$):

$$\kappa_i^{(d)} \approx \log(\Upsilon^{(d)}_{c[i],t[i]}/\Upsilon_{c[i],t[i]}) = -\log(\omega_i^{(d)}).$$

Based on this equation, we estimate the adjustment $\hat{\gamma}_d$ and variance $\hat{\varphi}_d^2$ in Eq. (2.1) by the median and variance of the predictive distribution for $-\log(\omega_i^{(d)})$. This predictive distribution is based on the following assumption:

$$\text{logit}(\omega_i^{(d)})|\mu_{\omega,d}, \sigma^2_{\omega,d} \sim N(\mu_{\omega,d}, \sigma^2_{\omega,d}), \tag{1.9}$$

where $\mu_{\omega,d}$ is the mean of the logit-transformed probabilities, and $\sigma_{\omega,d}$ the standard deviation. We use vague prior for the mean and variance parameters:

$$\begin{aligned} \sigma_{\omega,d} &\sim N^+(0,1), \\ \text{expit}(\mu_{\omega,d}) &\sim U(0,1). \end{aligned}$$

Definitions overlapping with the standard 28 weeks definition

Stillbirths $z_i^{(d)}$ recorded using the 1000 or 500 grams definitions are overlapping with the stillbirths z_i using the standard definition.

In this setting, let $N_i = n_i^{(r\&d)} + n_i^{(r)} + n_i^{(d)}$ refer to the total number of stillbirth based on the standard definition *or* an alternative definition, with $n_i^{(r\&d)}$ the count of stillbirths that satisfy the 28-week and alternative definition, $n_i^{(r)}$ the count of stillbirth with standard definition rather than alternative definition, and finally $n_i^{(d)}$

the counts of stillbirth with alternative definition rather than standard definition. We can assume

$$(n^{(r\&d)}, n^{(r)}, n^{(d)})|(\omega_i^{(r\&d)}, \omega_i^{(r)}, \omega_i^{(d)}) \sim Multinom\left(N_i, (\omega_i^{(r\&d)}, \omega_i^{(r)}, \omega_i^{(d)})\right),$$

where $\omega_i^{(r\&d)}, \omega_i^{(r)}$ and $\omega_i^{(d)}$ refer to the probabilities of a stillbirth satisfying both definitions, the standard definition only, and the alternative definition only, respectively.

Based on the expression for κ_i in Eq. (1.7), and the definitions of the $\omega_i^{(\cdot)}$s, we obtain the following relation:

$$\kappa_i^{(d)} \approx \log(\frac{\Upsilon_{c[i],t[i]}^{(d)}}{\Upsilon_{c[i],t[i]}}) = \log\left(\frac{\omega_i^{(r\&d)} + \omega_i^{(d)}}{\omega_i^{(r\&d)} + \omega_i^{(r)}}\right).$$

Based on this equation, we estimate the adjustment $\hat{\gamma}_d$ and variance $\hat{\varphi}_d^2$ in Eq. (2.1) by the median and variance of the predictive distribution for log-ratio of the definition-specific probabilities $\Gamma_i^{(d)} = \log\left(\frac{\omega_i^{(r\&d)}+\omega_i^{(d)}}{\omega_i^{(r\&d)}+\omega_i^{(r)}}\right)$.

We assume that the $\Gamma_i^{(d)}$s are normally distributed:

$$\Gamma_i^{(d)}|\mu_{\Gamma,d}, \sigma_{\Gamma,d}^2 \quad \sim \quad N(\mu_{\Gamma,d}, \sigma_{\Gamma,d}^2), \tag{1.10}$$

with $\mu_{\Gamma,d}$ and $\sigma_{\Gamma,d}^2$ referring to the across-setting mean and variance of the log-ratios. To guarantee that the estimation results in sets of $\omega_i^{(r\&d)}, \omega_i^{(r)}$ and $\omega_i^{(d)}$ that add up to one, we introduce the constraint $\frac{1}{1+\exp(\Gamma_i^{(d)})} < \omega_i^{(r)} + \omega_i^{(d)} < \frac{1}{\max\{1,\exp(\Gamma_i^{(d)})\}}$ and incorporate this constraint through a prior on the sum:

$$\left(\omega_i^{(r)} + \omega_i^{(d)}\right)|\Gamma_i^{(d)} \quad \sim \quad U\left(\frac{1}{1+\exp(\Gamma_i^{(d)})}, \frac{1}{\max\{1,\exp(\Gamma_i^{(d)})\}}\right).$$

Vague priors are used for the mean and variance parameters of $\Gamma_i^{(d)}$:

$$\begin{aligned} \sigma_{\Gamma,d} &\sim N^+(0,1), \\ \mu_{\Gamma,d} &\sim N(0,20). \end{aligned}$$

When fitting the model to the database $\mathcal{D}^d$ for the overlapping definition, we typically have available data pairs $(z_i, z_i^{(d)})$, as opposed to $n_i^{(r\&d)}, n_i^{(r)}$ and $n_i^{(d)}$. It follows that $z_i = n_i^{(r\&d)} + n_i^{(r)}$ and $z_i^{(d)} = n_i^{(r\&d)} + n_i^{(d)}$. We implemented an exact likelihood function to estimate the ωs from the overlapping data sets.

1.4.4 Bayesian hierarchical temporal sparse regression model

We developed a Bayesian hierarchical temporal regression model (BHTRM) to estimate the SBR for all country-years. It combines country-specific intercept ς_c, linear regression function $\sum_k X_{k,c,t}\beta_k$, and a temporal smoothing process $\delta_{c,t}$:

$$\Theta_{c,t} = \varsigma_c + \sum_k X_{k,c,t}\beta_k + \delta_{c,t}, \tag{1.11}$$

Country-specific intercepts ς_c are estimated hierarchically, with

$$\begin{aligned} \varsigma_c|\eta_{r[c]}, \sigma_\varsigma^2 &\sim N(\eta_{r[c]}, \sigma_\varsigma^2), \\ \eta_r|\xi_w, \sigma_\eta^2 &\sim N(\xi_w, \sigma_\eta^2), \end{aligned}$$

where $\eta_{r[c]}$ refers to the regional mean, σ_ς^2 to the across-country variance within regions, ξ_w to the global mean, and σ_η^2 to the across-region variance. Vague priors were used for the global mean and variances:

$$\begin{aligned} \xi_w &\sim N(2.5, 2^2), \\ \sigma_\varsigma, \sigma_\eta &\sim N^+(0,1). \end{aligned}$$

A penalized spline regression model is used for $\delta_{c,t}$,

$$\delta_{c,t} = \sum_{h=1}^{H} k_h(t)\alpha_{h,c}, \tag{1.12}$$

where $k_h(t)$ refers to the h-th spline function evaluated at time t and $\alpha_{h,c}$ to its regression coefficient for country c.

We use equally spaced quadratic B-splines, with knots spaced 1 year apart and placed at each integer year (Eilers and Marx [1996], Currie and Durban [2002]). The spline regression coefficients are modeled with a first-order random walk process with a sum-to-zero constraint $\frac{1}{H}\sum_h \alpha_{h,c} = 0$ to ensure identifiability. For each country, we define first order difference $\Delta\alpha_{h,c}$:

$$\Delta\alpha_{h,c} \quad = \quad \alpha_{h,c} - \alpha_{h-1,c}.$$

First-order differences are penalized as follows

$$\Delta\alpha_{h,c}|\sigma_\Delta^2 \sim N(0, \sigma_\Delta^2),$$

where the variance term σ_Δ^2 determines the smoothness of the fit. We address the sensitivity to these settings in Section 1.5.4.

Estimating regression coefficients using sparsity-inducing priors Blencowe et al. [2016] identified 16 candidate covariates for estimating SBR based on a conceptual framework and used a stepwise approach variable selection. In this study, we refrain from stepwise selection methods and instead use regularized horseshoe priors on regression coefficients (Piironen and Vehtari [2017a]) to impose sparsity by allowing shrinkage of coefficients to zero. We expand upon BHTRM by introducing sparsity-inducing priors for estimating regression coefficients β_k and refer to the

resulting model set up as a Bayesian hierarchical temporal sparse regression model (BHTSRM) which can be applied when the number of candidate covariates is large.

Regularized horseshoe priors for the regression coefficients are defined as follows:

$$\begin{aligned} \beta_k|\lambda_k, \tau, \rho &\sim N(0, \tau^2\tilde{\lambda}_k^2), \\ \tilde{\lambda}_k^2 &= \frac{\rho^2\lambda_k^2}{\rho^2 + \tau^2\lambda_k^2}, \end{aligned}$$

where τ and ρ are global shrinkage parameters, and the λ_ks are local (coefficient-specific) parameters. Priors are set as follows:

$$\begin{aligned} \lambda_k &\sim C^+(0, \lambda_0), \\ \tau &\sim C^+(0, \tau_0), \\ \rho^2 &\sim \text{Inv-Gamma}(\rho_1, \rho_2), \end{aligned}$$

where $C^+(0, s)$ refers to a half-Cauchy distribution with location parameter 0 and scale parameter s; λ_0, τ_0, ρ_1, and ρ_2 are fixed. The Cauchy distribution, which – compared to a normal distribution – has greater density around 0 and a heavier tail, allows the global hyperparameter τ to shrink all the parameters towards zero, while the heavy tail allows the coefficient-specific parameters λ_k's to make some coefficients escape from the global shrinkage. This set up allows for the inclusion of a larger set of candidate covariates and encourages sparseness by shrinking irrelevant covariates toward zero. It is not a variable selection method because it does not shrink all posterior samples to zero.

We set $\lambda_0 = \tau_0 = 1$, $\rho_1 = 2$, and $\rho_2 = 8$, as per the recommended defaults in Piironen and Vehtari [2017b], Carvalho et al. [2009b], and Gelman [2006]. We address the sensitivity to these settings in Section 1.5.4.

1.4.5 Computation

A Hamiltonian Monte Carlo (HMC) algorithm is employed to sample from the posterior distribution of the parameters of the SBR estimation model with the use of Stan (Carpenter et al. [2017]) and R package Rstan (Stan Development Team [2018]). Six parallel chains are run with a total of 6,000 iterations in each chain. The first 2,000 iterations in each chain are discarded as burn-in so that the resulting chains contain 4,000 samples each. Point estimates are given by medians of the posterior samples. Standard diagnostic checks are used to check convergence and sampling efficiency. These checks are based on trace plots, the improved Rhat diagnostic using rank-normalized draws (Gelman and Rubin [1992], Vehtari et al. [2020b]), and various calculations of effective sample size (ESS), including the bulk ESS and the tail ESS - giving the minimum of the effective sample sizes of the 5% and 95% quantiles.

1.4.6 Model validation and comparison

Performance of the SBR estimation model is assessed through two out-of-sample validation exercises. In the first exercise, we randomly leave out 20% of the observations and repeat this exercise 20 times (leaving out 306 observations each time). In the second exercise, we leave out the last observation for each country to check the predictive performance. To evaluate model performance, we calculate various measures based on a comparison between left-out observations and their predictive distributions. We define prediction errors e_i as the difference between the left-out observation and the median of its predictive posterior distribution based on the training set:

$$e_i = (\log(y_i) - \log(\tilde{y}_i))/S_i,$$

where y_i is the left-out observations, and $\tilde{y}_i$ and S_i refer to the estimated median and standard deviation of the predictive distribution for y_i based on the training set. Coverage of prediction intervals is given by $N^{-1}\sum_{i=1}^{N} 1[l_i \leq y_i \leq u_i]$, where N denotes

the total number of left-out observations considered, and l_i and u_i are the lower and upper bounds of the prediction interval for the ith observation. We also carry out approximate leave-one-out cross-validation (LOO), which is implemented in the loo package in R (Vehtari et al. [2019]).

For comparing models, we consider the expected log pointwise predictive density (ELPD) and Pareto K diagnostic (Vehtari et al. [2017]). The ELPD is the log pointwise predictive density for a new data set, which can be used to evaluate the performance of the model to predict the future data. The Pareto K diagnostic refers to the estimates of the shape parameter k of the generalized Pareto distribution. Values larger than one may indicate that the observation is outlying and influential.

1.5 Results

1.5.1 Data quality and data adjustments

Adjustments $\hat{\gamma}_d$ and standard deviations $\hat{\varphi}_d$ associated with alternative definitions are given in Table 1.3. For example, adjustments on the log-scale for 1000 grams definition is -0.065 (-0.074, -0.056), suggesting that the 1000 grams definition data are on average 0.937 (0.929, 0.946) times lower than the standard definition.

Definition	Income group	$\hat{\gamma}_d$ (95% CI)	$e^{\hat{\gamma}_d}$(95% CI)	$\hat{\varphi}_d$
22 weeks	Low	0.214 (0.101, 0.426)	1.239 (1.106, 2.031)	0.084
22 weeks	High	0.389 (0.175, 0.777)	1.476 (1.192, 2.175)	0.156
24 weeks	High	0.222 (0.058, 0.709)	1.248 (1.060, 2.031)	0.172
1000 grams	High	-0.065 (-0.074, -0.056)	0.937 (0.929, 0.946)	0.073
500 grams	High	0.244 (0.232, 0.257)	1.277 (1.261, 1.293)	0.087

Table 1.3. Adjustments and standard deviation of alternative definition versus the 28 week definition.

Table 1.4 summarizes the differences in error standard deviation σ_j associated with the different source types, ranging from a standard deviation of 0.017 for national administrative data to 0.239 for population study data. The bias ψ_j for survey data

is estimated at -0.165 (-0.229, -0.100) on the log-transformed scale, suggesting that survey data are on average 0.848 (0.795, 0.905) times lower than the truth.

Source type	$\hat{\psi}_j$	$\hat{\sigma}_j$
Administrative	-	0.017
HMIS	-	0.045
Population study	-	0.239
Survey	-0.165 (-0.229, -0.100)	0.135

Table 1.4. Source type bias and source type standard deviation.

1.5.2 Illustrative findings

Estimates for selected countries[1] are given in Figure 1.2, with final estimates displayed in red and underlying covariate-based estimates (obtained by removing the smoother term $\delta_{c,t}$ from the model) in green. As highlighted earlier in the paper, data availability ranges in the selected countries from no data (Afghanistan) to an annual time series of national administrative data based on the standard definition for Ireland. The BHTSRM produces estimates for both countries. Point estimates for Ireland track the observed SBR from administrative system closely and credible intervals are close to the uncertainty associated with each observed SBR. Estimates for Afghanistan are driven by covariates and the estimates are uncertain due to the absence of data.

Botswana, Malawi, Ukraine, and Uganda are examples of countries with SBR data that are either subject to bias, substantial error variance, or missing for periods of interest. In Ukraine, SBR data are available from 2007 to 2017 from administrative systems but recorded using a 22 week definition. SBR estimates are informed by the adjusted observations and uncertainty increases in extrapolations past the observation period. The survey data point has a large associated uncertainty and has

[1]Estimates for all countries see childmortality.org

little influence on the resulting model fit. In Uganda, the only available data come from HMIS, surveys, and population-based studies. There is substantial uncertainty associated with survey and population-based study data and resulting SBR estimates reflect this. There are four different data sources in Botswana and Malawi. Resulting estimates are more certain in years with administrative or HMIS data as compared to population-based, survey, or no data.

The effect of adding the smoother to the regression model on point estimates is most visible in Ireland where final point estimates differ from the covariate-driven ones. In general, credible intervals are wider for the model that includes the smoother as shown in Figure 1.2. Exceptions include countries where data are limited except for a short period with low-variance data such as Malawi: in such countries, the addition of the smoother results in reduced uncertainty in the period with low-variance data (when the estimates are data-driven).

1.5.3 Covariates

Table 1.5 summarizes the estimates for regression coefficients, ordered by absolute point estimates of the coefficients. Given that covariates were standardized, the coefficients are measured in units of standard deviation of the covariate, which are added to the table for reference. In the analysis by Blencowe et al. [2016], NMR, low birthweight, gross national income, mean years of female education, and coverage of 4 antenatal care visits (log(nmr), log(lbw), log(gni), edu, and anc4 in Table 1.5) were selected for inclusion in the regression model. Here we find that these covariates are ranked among the top in terms of their absolute regression coefficient along with C-section (csec). Comparisons between the model with horseshoe priors and additional models for sensitive checks are given in Section 1.5.4.

Covariates	Estimate $\hat{\beta}$	2.5%	97.5%	SD(covariate)
log(nmr)	0.414	0.336	0.492	0.999
log(gni)	-0.102	-0.212	0.001	1.20
log(lbw)	0.078	0.009	0.141	0.439
edu	-0.037	-0.104	0.007	3.41
csec	-0.027	-0.082	0.008	11.9
anc4	-0.025	-0.094	0.014	21.8
pab	-0.018	-0.050	0.006	11.6
abr	-0.017	-0.109	0.023	46.5
urban	-0.012	-0.087	0.024	23.1
gini	0.010	-0.017	0.061	8.17
sab	-0.010	-0.083	0.026	0.215
anc1	-0.009	-0.067	0.021	14.7
mmr	0.003	-0.057	0.109	288.5
pfpr	-0.002	-0.045	0.030	0.118
gdp	0.001	-0.047	0.063	$207 \cdot 10^2$
gfr	0.000	-0.057	0.054	0.049

Table 1.5. Estimates for regression coefficients under BHTSRM fit. Point estimates of regression coefficients, 95% credible interval given by the 2.5th and 97.5th percentiles of the posterior, and the standard deviation of the covariate prior to standardization. Details on covariates are given in Table A.3.

1.5.4 Model validation, comparison and sensitivity analyses

Validation results for the BHTSRM are given in Table 1.6. For all scenarios, mean residuals are close to zero, and the mean of the absolute residuals are around 0.1. The approximate leave-one-out validation exercise suggests that predictive distributions are overdispersed as compared to the left-out observations, with the percentages outside of 80% and 90% prediction intervals being lower than expected. The out-of-sample exercises suggest that the model is reasonably well calibrated, with slightly less left-out observations falling below their respective predictive intervals than expected.

We compare the performance of the BHTSRM, using sparse priors, to that of a model with vague priors on the regression coefficient, labeled BHTRM. Regression coefficients estimates for both the BHTSRM and BHTRM are given in Appendix Figure A.2 and Table A.1. Some of the coefficients are closer to zero in the BHTSRM

Validation Desirable*	N.test -	Mean err. 0	Mean abs. err. N/A	below 5% 5%	below 10% 10%	above 90% 10%	above 95% 5%
BHTSRM Recent	112	-0.001	0.091	3.5%	6.2%	2.7%	1.8%
BHTSRM In-sample	1531	-0.002	0.090	1.8%	3.5%	4.2%	1.6%
BHTSRM Random	306	-0.001	0.090	1.8%	3.3%	4.3%	1.8%

Table 1.6. Validation results for SBR estimates. "N.test" represent the number of observations in the test set. Validation exercises "Recent", "Random", and "In-sample" represent leaving out recent observations, randomly leaving out 20% of all observations, and approximate leave-one-out validation, respectively. The outcome measures are as follows: mean of error, mean absolute error, and % of left-out observations below and above their respective 90% and 80% prediction intervals. Desirable* refers to outcomes for models that are unbiased and well calibrated.

as compared to in the BHTRM, due to the shrinkage by the regularized horseshoe prior. We compare predictive performance between the BHTSRM and the BHTRM in Table A.2 and find that the mean error and mean absolute error are close to each other. Validation results by income group do not suggest difference in model performance either. The ELPD is higher for our reference BHTSRM as compared to the BHTRM, the 95% CI for the difference is (-12.6, -0.06) (see Table 1.7), suggesting improved predictive performance due to the horseshoe prior.

We also compare the reference model to another BHTSRM that is fitted using an alternative choice of hyperparameters for the horseshoe prior based on Piironen and Vehtari [2017b]. For standard regression models with $y_i \sim N((\boldsymbol{X}_i\boldsymbol{\beta}, \sigma^2)$, Piironen and Vehtari propose to set the scale parameter τ_0 in the prior for τ as follows:

$$\tau \sim C^+(0, \tau_0),$$
$$\tau_0 = \frac{p_0}{D - p_0}\frac{\sigma}{\sqrt{n}},$$

where p_0 is the guess of number of relevant predictors, D is the total number of predictors, σ is the standard deviation of observation $\log(y)$, and n is the number of observations. We cannot directly follow this recommendation because our modeling context differs from the one where this setting was explored, i.e. our setting includes heteroskedasticity of observations and the regression model is combined with a temporal smoothing term. We obtain a model fit based on the recommendation as a sensitivity test. Specifically, we obtain the fit for $p_0 = 5$, $D = 16$, $\sigma = 0.094$ (the median standard deviation across observations), and $n = 1531$, corresponding to $\tau_0 = 0.001$. Its ELPD is lower than the reference BHTSRM, but the difference is not significant according to the 95% CI (-6.62, 3.07).

We checked the sensitivity of the choice of the splines in the smoothing term $\delta_{c,t}$ in Eq. (1.12) by comparing the reference model fit to the fits obtained from two

alternative models. Model specifications were the same for the three models except for the specification of the smoothing term. In the reference model, a quadratic B-spline model was used, with knots spaced 1 year apart and placed at each integer year. In model "smooth1", a cubic B-splines model was used while in "smooth2" the knots were spaced 2 years apart. Table 1.7 summarizes the differences in ELPD and Pareto K values for different models. There are no improvements when comparing the alternative smoothers with our reference SBR model.

Model	ELPD			Pareto K diag.
	estimate	SE	95% CI for difference	
BHTSRM	1194.5	40.6	reference model	0.5%
HS $\tau_0 = 0.001$	1192.7	40.6	(-6.62, 3.07)	0.7%
BHTRM	1188.2	40.7	(-12.59 , -0.06)	0.8%
smooth1	1185.9	40.5	(-13.08, -4.15)	0.5%
smooth2	1176.7	40.9	(-11.1, 3.0)	0.3%

Table 1.7. Model comparison based on expected log pointwise predictive density and Pareto K diagnostic values. BHTSRM is our proposed model, and BHTRM is model with vague prior on covariates. The HS $\tau_0 = 0.001$ model stands for BHTSRM with $\tau_0 = 0.001$. Smooth1 and Smooth2 are two models with different settings of smoothers described in the text. When comparing models, larger ELPD value suggests better model performance. The percentage of high influential points (Pareto K values ¿ 1) for all models are presented in the "Pareto K diag." column, lower outcomes are preferred.

1.6 Discussion

We developed a Bayesian hierarchical temporal sparse regression model (BHTSRM) for estimating SBRs for all countries from 2000 until 2019. Estimating SBRs is challenging because of data paucity, especially for many LMICs where most stillbirths occur, and the substantial uncertainty associated with observations due to reporting issues and errors associated with the observations. Our BHTSRM extends the approach previously proposed by Blencowe et al. [2016] to produce estimates that are informed by a covariate model and available data, accounting for different definitions

and uncertainty associated with the available data. Model validation exercises suggest that the model is reasonably well calibrated.

The BHTSRM extends upon previous applications of Bayesian hierarchical temporal regression models through the introduction of sparsity-inducing priors and new statistical approaches to addresses data quality issues. Sparsity-inducing priors allow for the inclusion of larger sets of (potentially correlated) candidate covariates into the model. While validation exercises do not indicate improved performance of the model with the horseshoe prior over a model with vague priors, improved predictive performance was suggested by higher ELPD in our application.

To address data quality issues, we developed a statistical procedure for data exclusion based on comparing observed ratios of SBR to NMR for the population of interest to a reference distribution of such ratios. This approach improves upon the previously used approach for data exclusion by defining a predictive distribution for the ratio and a decision rule that makes explicit the probability of a false exclusion. Secondly, we developed a new approach to adjust and estimate additional uncertainty associated with observations using a different definition of stillbirths. In the model fitting, we used a data model that accounts for bias and varying sources of random error associated with the observations.

While our approach to estimating the SBR improves upon existing approaches, there are limitations related to the model and data availability. Limited data availability restricted the analyses we are able to carry out and result in stricter modeling assumptions. For example, we excluded data based on observed SBR to NMR ratios. In this analysis, we combined data across settings when constructing a predictive distribution for the expected ratio and chose 5% as the threshold for data exclusion. We acknowledge that the choice of a higher (or lower) threshold would have resulted in the exclusion of more (or less) data. Additional data related to the quality of reporting would allow for more detailed analyses and may allow for avoiding having

to set a threshold. Relative differences in SBRs associated with the use of different definitions, i.e., gestational age, may vary across settings. Data limitations resulted in the use of a simple dichotomy of high income and low income countries to capture this difference. With additional data, this relationship can be studied in more detail. Lastly, although the horseshoe prior allows for the inclusion of a larger set of candidate covariates and shrinkage toward zero of irrelevant covariates, it is not a variable selection method because it does not shrink all posterior samples to zero.

The BHTSRM as described in this paper is used by the UN IGME to generate estimates for the SBR globally (UN Inter-agency Group for Child Mortality Estimation [2020], Hug et al. [2021]). While the modeling approach allows for the construction of estimates for all countries, we find that uncertainty associated with the estimates is substantial in many settings, including countries with high SBRs. This highlights the need for additional data collection to produce more precise information for monitoring and program planning, especially in high-burden settings.

CHAPTER 2

SELECTION OF PREDICTORS BASED ON HORSESHOE PRIOR

Bayesian hierarchical temporal regression models (BHTRMs) are used in various applications, including the estimation of national-level stillbirth rates from 2000 to 2019 for countries around the world. In a BHTRM, a regression model is combined with a temporal smoothing process so that estimates are data-driven in periods with high-quality data and determined by covariates for periods with limited or no data. Horseshoe priors can be used to encourage sparseness - shrinkage towards zero of regression coefficients of irrelevant predictors - in settings where the number of candidate covariates is large. Although the use of horseshoe priors can result in models with good predictive performance, posterior samples of regression coefficients are not shrunk to zero exactly to inform the exclusion of irrelevant covariates. The performance of existing approaches to variable selection have not been validated for BHTRMs.

Motivated by the demand for a parsimonious model for estimating stillbirth rates, we explored the use of existing methods for variable selection in a BHTRM. Through simulation exercises, we find that existing approaches may be subject to high false exclusion rates in settings that mimic that of the stillbirth estimation problem, where weak signals may be present.

To improve upon existing methods, we propose a new computationally-efficient approach to variable selection. The new approach for variable selection is based on a model fit using horseshoe priors on regression coefficients. In the proposed set up,

we introduce noise covariates in the model of interest to obtain a distribution of a shrinkage parameter for irrelevant covariates. For each candidate covariate in the model of interest, we propose an exclusion rule based on comparing the estimate shrinkage for the candidate covariate to the distribution obtained from the noise covariates.

We use a simulation study to assess the predictive performance of the proposed approach and compare its performance to existing methods. We find that the new method improves upon the false exclusion rates of other methods in BHTRMs with weak signals, while giving comparable performance in terms of prediction errors and coverage. We apply the new approach to stillbirth estimation problem and compare its estimates and goodness of fit to a reference BHTRM that is fitted with horseshoe priors.

2.1 Introduction

In the previous chapter (Wang et al. [2020]), we develop a Bayesian hierarchical temporal regression model (BHTRM) to estimate stillbirth rates (SBR) globally. This type of model combines a temporal smoothing process to produce estimates that track high quality, and a hierarchical regression model to produce covariate-driven trend estimates for countries with limited data. In this particular model, horseshoe priors were used to shrink the coefficients of the irrelevant covariates to zero to encourage sparseness. Although the use of horseshoe priors results in models with good predictive performance (Piironen and Vehtari [2017b]), posterior samples in the resulting model fit are not shrunk to zero exactly to identify the irrelevant covariates. Hence it cannot identify the irrelevant covariates. Given the demand for a parsimonious model for estimating stillbirth rates, we explored variable selection approaches for BHTRMs

Numerous methods for Bayesian model selection have been proposed but no standard variable selection methods have been designed or even tested for a semi-parametric model like the BHTRM. Various approaches with their theoretical properties have been reviewed by Piironen and Vehtari [2017a]. However, given that no standard variable selection methods have been designed or even tested for a semi-parametric model like the BHTRM, their performance in these settings is unknown. Obviously, the temporal smoother and the hierarchical intercepts could introduce difficulties in selecting relevant covariates or optimizing predictive performance. Furthermore, there are two additional features of the SBR application which may affect the selection accuracy: high correlation among covariates and possibly weak signals. The candidate covariates used in SBR project consist many socio-economic factors, demographic and biomedical factors, which are highly correlated. Based on estimated regression coefficients obtained in Wang et al. [2020], the great majority of signals may be very weak.

Motivated by the demand for a parsimonious model for estimating stillbirth rates, we explored the use of existing methods for variable selection in a BHTRM and propose a new approach. The new approach is referred to as the horseshoe shrinkage parameter reference distribution variable selection method (HSS-VS). As the name suggested, the HSS-VS is developed based on horseshoe prior. The basic idea of our approach is summarized as follow: we introduce dummy covariates in the model of interest and use horseshoe priors for all coefficients to obtain a distribution of shrinkage parameters for irrelevant covariates. For each candidate covariate in the model of interest, we propose an exclusion rule based on comparing the estimate shrinkage for the candidate covariate to the distribution obtained from the dummy covariates. We use a simulation study to assess the predictive performance of the proposed approach and compare its performance to existing methods.

This paper is organized as follows. In Section 2.2, we summarize the BHTRM using for estimation the SBR. In Section 2.3, we summarize some existing variable selection methods, and discuss the feasibility of using each method for BHTRMs. We introduce our proposed HSS-VS method in Section 2.4. In Section 2.5, we explain how we compare predictive performance and selection accuracy across different variable selection methods. A simulation study is presented in Section 2.6 and in Section 2.7, we apply the variable selection methods to the SBR estimation problem. We conclude with a summary of findings and a discussion of limitations and future research directions in Section 2.8.

2.2 Summary of the BHTRM used for SBR estimation

This section gives a simplified introduction to the BHTRM used for SBR estimation, focusing on the presenting the details relevant for the variable selection problem. A full model description is given in Wang et al. [2020].

We denote y_i as an observed SBR outcome for country $c[i]$ in year $t[i]$ from source $d[i]$, $\Theta_{c,t}$ as the true outcome for that country-year, and σ_d^2 as the source type variance. We have following data model,

$$y_i|\Theta_{c[i],t[i]}, \sigma^2_{d[i]} \sim N(\Theta_{c[i],t[i]}, \sigma^2_{d[i]}), \tag{2.1}$$

with following BHTSRM specification for $\Theta_{c,t}$:

$$\Theta_{c,t} = \alpha_c + \sum_{k=1}^{K} x_{k,c,t}\beta_k + \delta_{c,t}, \tag{2.2}$$

where α_c refers to the country-specific intercept, $\delta_{c,t}$ refers to a temporal smoothing process, and $\sum_k x_{k,c,t}\beta_k$ refers to the linear regression part. In the regression part, covariates x_k are standardized. The shrinkage of regression coefficients toward zero is encouraged by regularized horseshoe priors, which are defined as follow:

$$\beta_k|\lambda_k, \tau, \rho \sim N(0, \tau^2\tilde{\lambda}_k^2), \tag{2.3}$$

$$\tilde{\lambda}_k^2 = \frac{\rho^2\lambda_k^2}{\rho^2 + \tau^2\lambda_k^2}, \tag{2.4}$$

where τ and ρ are global shrinkage parameters, and the λ_ks are local (coefficient-specific) parameters. Priors are set as follows:

$$\lambda_k \sim C^+(0, \lambda_0), \tag{2.5}$$

$$\tau \sim C^+(0, \tau_0), \tag{2.6}$$

where $C^+(0, s)$ refers to a half-Cauchy distribution with location parameter 0 and scale parameter s. λ_0, τ_0, q and g are fixed.

Country-specific intercepts α_c are estimated hierarchically, with

$$\alpha_c|\eta_{r[c]}, \sigma_\alpha^2 \sim N(\eta_{r[c]}, \sigma_\alpha^2), \tag{2.7}$$

$$\eta_r|\xi_w, \sigma_\eta^2 \sim N(\xi_w, \sigma_\eta^2), \tag{2.8}$$

where $\eta_{r[c]}$ refers to the regional mean, σ_ς^2 to the across-country variance within regions, ξ_w to the global mean, and σ_η^2 to the across-region variance.

A penalized spline regression model is used for $\delta_{c,t}$,

$$\delta_{c,t} = \sum_{h=1}^{H} k_h(t)\varsigma_{h,c}, \tag{2.9}$$

where $k_h(t)$ refers to the h-th spline function evaluated at time t and $\varsigma_{h,c}$ to its regression coefficient for country c. We use equally spaced quadratic B-splines, with knots spaced 1 year apart and placed at each integer year. The spline regression coefficients are modeled with a first-order random walk process with a sum-to-zero constraint $\frac{1}{H}\sum_h \varsigma_{h,c} = 0$ to ensure identifiability. For each country, we define first order difference $\Delta\varsigma_{h,c}$:

$$\Delta\varsigma_{h,c} = \varsigma_{h,c} - \varsigma_{h-1,c}. \tag{2.10}$$

First-order differences are penalized as follows

$$\Delta\varsigma_{h,c}|\varsigma_\Delta^2 \sim N(0, \varsigma_\Delta^2), \tag{2.11}$$

where the variance term σ_Δ^2 determines the smoothness of the fit.

2.3 Existing approaches for Bayesian variable selection

Numerous methods for Bayesian model selection and assessment have been proposed. Various approaches and their theoretical properties have been reviewed in

prior work, for example by Bernardo and Smith [1994], Vehtari and Ojanen [2012], and Piironen and Vehtari [2017a]. Drawing from these prior studies, we summarize some widely used Bayesian variable selection methods in section B.1 in the Appendix.

While various approaches to variable selection exist, not all of them are feasible to apply in a BHTRM model as given in Eq. (2.2), where the outcome of interest is the sum of a linear regression model, group-specific intercepts α_c, and temporal smoother terms $\delta_{c,t}$. Moreover, given the run time associated with the SBR estimation model and its implementation in Stan (Stan Development Team [2018], Carpenter et al. [2017]), methods that require repetitive model fitting or are based on a discrete parameter (such as spike and slab priors Mitchell and Beauchamp [1988]) are not an option. This is explained in more detail in Appendix section B.2.

In the remainder of this section, we summarize existing methods that we explored for variable selection in BHTRMs in general, and the SBR estimation model in particular.

2.3.1 Expected log posterior predictive: approximate leave-one-out CV (loo)

Information criteria offer a computationally appealing way of estimating the predictive performance of the model. The ELPD is the log pointwise predictive density for a new dataset which can be used to evaluate the performance of the model to predict the future data. Vehtari et al. [2017] proposed the Pareto Smoothed Importance Sampling (PSIS) for approximate leave-one-out cross validation (LOO) implemented in the loo package (Vehtari et al. [2020a]). It is computationally efficient as it does not require completely re-fitting the model, unlike cross-validation, and it is more robust than widely applicable information criterion (WAIC) (Watanabe [2010]) in finite case with weak priors or influential observations.

We will use an iterative forward method to test the performance of a variable selection method using ELPDs in the BHTRM setting:

1. Start from the empty model and compute ELPD.

2. Add covariate which increases the ELPD by the largest amount.

3. Repeat step 2 until no improvement of ELPD is observed.

2.3.2 Variable selection based on 95% credible intervals (CI)

One of the most simple methods to select variable is to use credible intervals based on the posterior distributions of regression coefficients. However, models with non-informative priors on the covariates' coefficient usually results in wide credible intervals which make it hard to detect the weak signals. van der Pas et al. [2017] show that the marginal credible intervals given by horseshoe priors are narrow enough to be informative for variable selection in the normal means problem. Bhattacharya et al. [2016] compare various aspects of horseshoe prior to frequentist procedures in the linear regression problem, and obtain highly promising results.

We will consider the performance of using 95% CIs for variable selection. The procedure is as follows: Denote (l_k, u_k) as the 95% CI for β_k from a model using regularized horseshoe priors. We select covariate k if $l_k < 0 < u_k$.

2.3.3 Variable selection based on Projection method (proj)

The key characteristic of a projection approach (Goutis and Robert [1998], Piironen et al. [2020]) is to find an optimal trade-off between sparsity and predictive accuracy. The goal is to simplify the full model by projecting the information in the posterior onto a candidate submodel so that the predictive distribution of the submodel is as close to the reference model as possible. However, the projection approach as proposed by Goutis and Robert [1998],Dupuis and Robert [2003a],Piironen et al. [2020] cannot be applied to the BHTRM directly.

We revised the approach as follows: First, we project the full model M^* onto a submodel M^b (with parameters indicated with superscript b) by setting the smoothing term equal to 0, i.e., $\delta^b_{c,t} = 0$. Then the projected parameters β^p in the parameter space of sub-submodel M^p can be defined as

$$\beta^p = \underset{\beta}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} D(p(\tilde{y}^b \mid x_i, \beta^*, \alpha, \sigma, \delta^b, M^b) \mid\mid p(\tilde{y}^p \mid x_i, \beta, \alpha, \sigma, \delta^b, M^p)), \quad (2.12)$$

where D refers to the divergence between two posterior distributions, and $\tilde{y}^b$ refers to the data without smoothing term, i.e.,

$$\tilde{y}^b_i \sim N(\alpha_{c[i]} + \sum_k x_{ki}\beta_k, \sigma^2_{d[i]}) \quad (2.13)$$

The discrepancy between the first projection submodel, M^b, and the submodel of the projection submodel, M^p, is then defined to be the expectation of this divergence over the posterior of the first order projection model. Dupuis and Robert [2003b] introduce the notation of explanatory power to measure the distance between models. In BHTSRM setting, we use the mean square error to obtain the best sub-submodel.

We calculate the discrepancy $\Delta(M^b||M^p)$ based on samples $\{\beta^*, \alpha, \sigma\}^S_{s=1}$ from the posterior of the reference model, the samples of the projected parameters $\{\beta^p\}^S_{s=1}$, as follows:

$$\Delta(M^b||M^p) = \frac{1}{n} \sum_{i=1}^{n} (\tilde{y}^b_i - \tilde{y}^p_i)^2, \quad (2.14)$$

where $\tilde{y}^b_i = \frac{1}{S} \sum \tilde{y}^{b(s)}_i$, and $\tilde{y}^p_i = \frac{1}{S} \sum \tilde{y}^{p(s)}_i$, with

$$\tilde{y}^{b(s)}_i \sim N(\alpha^{(s)}_{c[i]} + \sum_k x_{ik}\beta^{*(s)}_k, \sigma^{2(s)}_{d[i]}). \quad (2.15)$$

$$\tilde{y}^{p(s)}_i \sim N(\alpha^{(s)}_{c[i]} + \sum_k x_{ik}\beta^{p(s)}_k, \sigma^{2(s)}_{d[i]}). \quad (2.16)$$

After fitting the model with all the variables using regularized horseshoe prior, we use the projection predictive variable selection strategy above. As a search heuristic, we use forward searching, that is, starting from the empty model, we add variables one at a time, each time choosing the variable that decreases the mean square error the most.

2.4 HSS-VS method

In this section, we introduce the HSS-VS method to carry out variable selection for a BHTRM with process model as specified in Eq. (2.2). The new approach for variable selection is based on a model fit using horseshoe priors on regression coefficients. In the proposed set up, we introduce noise covariates in the model of interest to obtain a distribution of a shrinkage parameter for irrelevant covariates. For each candidate covariate in the model of interest, we propose an exclusion rule based on comparing the estimate shrinkage for the candidate covariate to the distribution obtained from the noise covariates.

This section is organized as follows: we first introduce a shrinkage measure associated with each regression coefficient. We then explain how we obtain a reference distribution of the shrinkage parameter for irrelevant covariates, and how we use that reference distribution to carry out variable selection. Finally, we summarize the approach.

2.4.1 Shrinkage measure ω_k for regression coefficient k in a sparse regression model

When fitting a regression model with horseshoe priors on regression coefficients β_k, we set (as explained in Section 2.2):

$$\beta_k|\lambda_k, \tau, \rho \quad \sim \quad N(0, \tau^2\tilde{\lambda}_k^2), \tag{2.17}$$

where τ is a global shrinkage parameter, and the λ_ks are local (coefficient-specific) parameters. We denote ψ_k as the standard deviation of the prior on β_K:

$$\psi_k = \tau\tilde{\lambda}_k. \tag{2.18}$$

When using the horseshoe prior, shrinkage toward zero is achieved when the variance $\psi_k^2 \to 0$, such that more weight is assigned around 0 for irrelevant covariates. The smaller (larger) ψ_k^2, the more (less) shrinkage is encouraged. To obtain a standardized measure of shrinkage, we introduce shrinkage measure $0 \leq \omega_k \leq 1$, given by

$$\omega_k = \frac{1}{1+\hat{\psi}_k}, \tag{2.19}$$

where $\hat{\psi}_k$ denotes the posterior median of ψ_k, such that ω_k closer to 1 indicates shrinkage while values closer to 0 suggest that the regression coefficient is more likely to escape from shrinkage. Thus, the ω_k^2 will differ between relevant covariates and irrelevant covariates. This property motivates the new proposed approach for variable selection.

2.4.2 HSS-VS method details

In the HSS-VS approach, we aim to compare the shrinkage measures ω_k to a reference distribution that quantifies shrinkage for coefficients that are equal to zero, to then exclude covariates that are likely to be irrelevant.

We obtain a reference distribution by adding dummy noise covariates o_ms in the model as described in Eq. (2.2) as follows:

$$\Theta_{c,t} = \alpha_c + \sum_{k=1}^{K} x_{k,c,t}\beta_k + \delta_{c,t} \tag{2.20}$$

$$+ \sum_{m=1}^{M} o_{m,c,t}\nu_m, \tag{2.21}$$

where ν_m is the coefficient of the dummy covariate o_m, with ν_m set to be 0 and dummy covariate $o_{m,c,t}$ sampled from a standard normal distribution. We also assign the same regularized horseshoe prior on ν_m as

$$\begin{aligned} \nu_m \mid \psi'_m &\sim N(0, {\psi'_m}^2), && (2.22) \\ \psi'_m &= \tau\lambda'_m, && (2.23) \end{aligned}$$

where priors for τ and λ'_m are defined as in Eq. (2.3). Based on the estimates of the shrinkage parameters $\hat{\omega}'_m$ for the coefficients for the dummy covariates o_m, we obtain a shrinkage parameter reference distribution Ω. This distribution is typically skewed towards one, given the irrelevant coefficients are shrunk toward zero. Figure 2.1 includes a illustration.

We define the variable selection procedure using the reference distribution Ω. This reference distribution quantifies the range of outcomes for shrinkage parameters when the true regression coefficient equals zero. We assume that the shrinkage parameter ω_k for candidate covariate k that equals zero follows distribution Ω. For significance level α, we carry out variable selection as follows: the k-th covariate x_k is included in the model if its shrinkage parameter $\omega_k < \Omega_\alpha$, where Ω_α denotes the $(\alpha) \cdot 100$-th percentile of reference distribution Ω. This rule results in exclusion of covariates with shrinkage that is in the lower tail of the reference distribution, see Figure 2.1 for an illustration of the approach. If the shrinkage parameters ω_k are distributed according to Ω, the false inclusion rate is 5%.

2.4.2.1 Summary of implementation of HSS-VS method

Variable selection is carried out with the HSS-VS method in two steps:

1. Fit the BHTRM that includes dummy covariates, using horseshoe priors for all regression coefficients, and obtain the reference distribution of the shrinkage

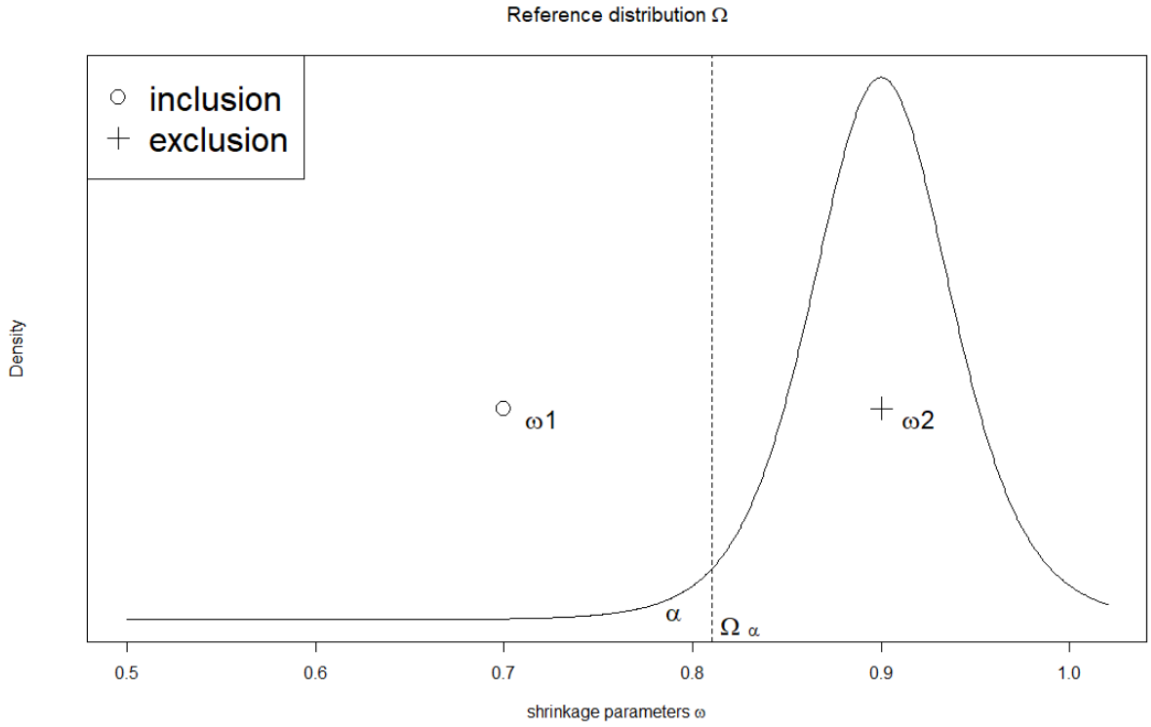


Figure 2.1. Illustration of reference distribution Ω. The black line is the density and the dashed line is the threshold Ω_α. The estimated shrinkage parameters for two example candidate covariates are given by the circle with shrinkage parameter ω_1 and cross ω_2. Because $\omega_1 < \Omega_\alpha$ and $\omega_2 > \Omega_\alpha$, the two covariates are selected as signal and noise respectively.

parameter for the dummy covariates. In simulations and application in this study, we set $M = 30$ and run the model 10 times to estimate Ω based on 300 estimates of ω'_m. It may be possible to improve upon this approach, see discussion.

2. Select candidate covariates based on comparing their shrinkage estimates to the distribution obtained from the dummy covariates: include covariate k if $\omega_k < \Omega_\alpha$, exclude otherwise.

2.5 Assessing performance of variable selection methods

We calculate various measures to asses the performance of our proposed variable selection method and compare our approach with other existing approaches. We first present measures to use in simulation exercises, where the truth is known, followed by measures used in real data applications.

2.5.1 Measure of prediction accuracy

We quantify a model's predictive performance in terms of estimating the outcome of interest Θ_i using mean error (ME), mean absolute error (MAE) and coverage rate (P_ϕ). The ME and MAE are defined as follows:

$$\text{ME} = \frac{1}{N}\sum_{i=1}^{N}(\widehat{\Theta}_i - \Theta_i), \tag{2.24}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\widehat{\Theta}_i - \Theta_i|, \tag{2.25}$$

where $\widehat{\Theta}_i$ it the median of the posterior sample of Θ_i. Probability integral transform (PIT) values are defined as,

$$P_\phi = \frac{1}{N}\sum_{i=1}^{N}1[\Theta_i \leq u_i] \tag{2.26}$$

where u_i refers to the ϕ percentile of the posterior distribution of Θ_i from the model fit.

We calculate the selection accuracy T, false exclusion rate F^-, and false inclusion rate F^+. Let $K = K_1 + K_2$ where K_1 equals to the number of relevant covariates with non-zero coefficients and K_2 denotes the number of irrelevant covariates. Accuracy T is defined as the number of correct excluded or included covariates over total number of covariates K:

$$T = \frac{1}{K} \sum_{k=1}^{K} 1_T(\hat{\beta}_k), \tag{2.27}$$

where $1_T(\hat{\beta}_k)$ is defined as

$$1_T(\hat{\beta}_k) : \quad = \quad \begin{cases} 1 & \text{If correct exclude/include } k\text{th covariate,} \\ 0 & \text{Otherwise.} \end{cases} \tag{2.28}$$

False exclusion rate F^- is defined as the number of false excluded covariates over total number of relevant covariates K_1:

$$F^- \quad = \quad \frac{1}{K_1} \sum_{k=1}^{K_1} 1_{F^-}(\hat{\beta}_k^{(\neq 0)}) \tag{2.29}$$

where $1_{F^-}(\hat{\beta}_k^{(\neq 0)})$ is defined as

$$1_{F^-}(\hat{\beta}_k^{(\neq 0)}) : \quad = \quad \begin{cases} 1 & \text{If false exclude } x_k \\ 0 & \text{otherwise} \end{cases} \tag{2.30}$$

False inclusion rate F^+ is defined as the number of false included covariates over total number of irrelevant covariates K_2:

$$F^+ \quad = \quad \frac{1}{K_2} \sum_{k=1}^{K_2} 1_{F^+}(\hat{\beta}_k^{(=0)}) \tag{2.31}$$

where $1_{F^+}(\hat{\beta}_k^{(=0)})$ is defined as

$$1_{F^+}(\hat{\beta}_k^{(=0)}) : \; = \; \begin{cases} 1 & \text{If false include } x_k \\ 0 & \text{Otherwise} \end{cases} \tag{2.32}$$

2.6 Simulation study

We evaluate our proposed approach in the multiple regression setting in Section 2.6.1, and then compare its performance with other existing approaches in the BHTRM setting in Section 2.6.2.

2.6.1 Multiple regression

In this simulation, we generate data from a multiple regression model:

$$\Theta_i = \sum_{k=1}^{K} \beta_k x_{ki}, \; y_i \sim N(\Theta_i, \sigma^2), \tag{2.33}$$

with error variance $\sigma^2 = 1$ and sample size $N = 2000$.

We set the total number of covariates to $K = 20$. The covariates are divided into four groups of five covariates. Each covariate x_k has a zero mean and unit variance and is correlated with the same variables across the groups with coefficient $\rho = 0.5$. Covariates in one group are not correlated in other groups.

The regression coefficients of covariates for the first group are $\beta_k = k \cdot \sigma$ for $k = 1, \ldots, 5$ and for the second group $\beta_{5+k} = k/10 \cdot \sigma$, corresponding to coefficients increasing from 0.1 to 0.5. Regression coefficients are set to 0 for $k > 10$.

We generated 30 datasets and added 30 dummy covariates each dataset and run each model 10 times to estimate reference distribution with regularized horseshoe prior as per Eq. (2.3).

We summarize findings from the multiple linear regression setting with uncorrelated and correlated covariates in Table 2.1. The PIT value is as expected. Selection

accuracy (T), false exclusion rate (F^-) and false inclusion rate (F^+) are (0.968, 0.01, 0.053) for uncorrelated covariates and (0.958, 0, 0.083) for correlated covariates. Note that the default threshold value $\alpha = 0.05$, which means that the expected false inclusion rate is 5% in both settings. We find that the false inclusion rate in the correlated setting is slightly greater than our expected α level.

Simulation	ME	MAE	PIT(5%,10%,90%,95%)	ACC (T)	FER (F^-)	FIR (F^+)
uncorrelated	-.0006	.101	(6%, 10%, 91%, 95%)	0.968	0.01	0.053
correlated	-.0003	.090	(4%, 8%, 92%, 97%)	0.958	0	0.083

Table 2.1. Linear regression simulation results. "ME" and "MAE" represent the mean error and mean absolute error. The values in "PIT(5%,10%,90%,95%)" are the desirable PIT values. "ACC", "FER" and "FIR" represent the accuracy rate, false exclusion rate, and false inclusion rate, respectively.

2.6.2 BHTRM

We compare the performance of the HSS-VS approach for variable selection in BHTRM settings with 4 existing approaches. The existing approaches are information criteria using ELPD LOO (loo), credible intervals (95%CI), and projection method (proj).

Simulation setting The simulations are based on the following BHTRM model,

$$\begin{aligned} y_i|\Theta_{c[i],t[i]}, \sigma^2_{d[i]} \sim & N(\Theta_{c[i],t[i]}, \sigma^2_{d[i]}), i = 1, ..., N \\ \Theta_{c,t} = & \alpha_c + \sum_{k=1}^{K_1} x_{k,c,t}\beta_k^{(\neq 0)} + \delta_{c,t}, \\ \delta_{c,t} = & \sum_{h=1}^{H} k_h(t)\varsigma_{h,c}, \end{aligned}$$

with - based on the SBR application - α_c generated hierarchically and $\delta_{c,t}$ generated based on the splines setup as explained in Section 2.2. Mean and variance parameters

associated with α_c and $\delta_{c,t}$ were fixed at estimates from the SBR model fit (Wang et al. [2020]).

We simulate $N = 2000$ observations and introduce varying error variance, as observed in the SBR estimation. The observations are divided into three groups with error standard deviation $\sigma_d^2 = \{0.1, 0.2, 0.3\}$.

The simulation settings for the covariates and regression coefficients are also motivated by the stillbirth estimation problem. We use $K = 16$ candidate covariates. Given that the candidate covariates for predicting the SBR are highly correlated, and the signals are relatively weak compared to the outcome variance, it is necessary to check the impact of correlation and weak signals on the selection accuracy of the approaches. We use four scenarios: (1) strong signals and median correlation; (2) weak signals and median correlation; (3) weak signals and high correlation; (4) strong signals and high correlation. In scenarios 1 and 2, the covariates and their correlation structure are taken from the correlated linear regression setting from the previous exercise with only first 16 covariates are used as candidate covariates (to make it comparable to true SBR covariates), thus the correlation with the same variables across the group is still $\rho = 0.5$. In scenarios 3 and 4, the SBR covariates are used. The correlation between those covariates is around 0.8. The $\beta^{(\neq 0)}$s are summarized in Table 2.2.

Scenario	N	σ_d^2	ρ	(K, K_1)	$\beta^{(\neq 0)}$
1	2000	(.1,.2,.3)	0.5	(16, 5)	(5,5,5,5,5)
2	2000	(.1,.2,.3)	0.5	(16, 9)	(.4,-.2,.2,-.15,.15,0.8,.06,.04,.02)
3	2000	(.1,.2,.3)	-	(16, 9)	(.4,-.2,.2,-.15,.15,0.8,.06,.04,.02)
4	2000	(.1,.2,.3)	-	(16, 5)	(5,5,5,5,5)

Table 2.2. Summary of BHTRM simulation settings. "N", "σ_d^2", and ρ are number of observations, the three-group error variance, and correlation between covariates. "(K, K_1)" represents the number of candidate covariates K and relevant covariates K_1. $\beta^{(\neq 0)}$ represents the true coefficients.

BHTRM Results The results for each of the four scenarios are summarized in Tables 2.3-2.6.

In scenarios 1 and 4 (strong signals and median/high correlation), all approaches have a prediction accuracy of around 0.95 or greater. There is no false decision for loo and projection methods. In the strong signal and increased correlation setting, there are some false inclusions when using HSS and 95% CI. For HSS, the false inclusion rate of 0.086 is slightly greater than our expected α level. Performance in terms of prediction errors and coverage is comparable.

In the weak signal settings in scenario 2 and 3, the selection accuracy of all methods becomes worse. In these settings, HSS has the smallest false exclusion rate among the four methods, which suggests that HSS is more sensitive to the weak signals. However, its false inclusion rate increases from 0.152 (scenario 2 with median correlation) to 0.305 (scenario 3 with high correlation), suggesting that HSS is more likely to include unrelated predictors. Performance in terms of prediction errors and coverage is again comparable.

The simulation results suggest that the HSS method is a promising approach for variable selection in settings similar to the SBR estimation problem, if the goal is to obtain a parsimonious model with predictive performance that is comparable to the full model, while minimizing false exclusion of covariates with weak signals.

Simu	ME	MAE	PIT(5%,10%,90%,95%)	ACC (T)	FER (F^-)	FIR (F^+)
HSS	.0009	.101	(3%, 6%, 94%, 98%)	.965	0	.082
loo	.0006	.109	(2%, 6%, 95%, 98%)	1	0	0
proj	.0006	.109	(2%, 6%, 95%, 98%)	1	0	0
95%CI	-.0003	.090	(2%, 6%, 95%, 97%)	.975	0	.036

Table 2.3. Scenario 1: strong signals and median correlation

Simu	ME	MAE	PIT(5%,10%,90%,95%)	ACC (T)	FER (F^-)	FIR (F^+)
HSS	-.0005	.100	(3%, 5%,95%,98%)	.894	.070	.152
loo	.0004	.099	(3%,4%, 95%, 97%)	.823	.156	.205
proj	.0004	.099	(3% 4%, 96%, 98%)	.817	.159	.210
95%CI	.0003	.101	(3%, 5%, 95%, 98%)	.894	.185	.005

Table 2.4. Scenario 2: weak signals and median correlation

Simu	ME	MAE	PIT(5%,10%,90%,95%)	ACC (T)	FER (F^-)	FIR (F^+)
HSS	-.0002	.098	(2%, 5%,96%,98%)	.821	.081	.305
loo	-.0002	.097	(2%,4%, 96%, 98%)	.829	.207	.224
proj	-.0005	.099	(2% 5%, 95%, 98%)	.813	.167	.214
95%CI	-.0002	.098	(2%, 5%, 95%, 98%)	.896	.181	.005

Table 2.5. Scenario 3: weak signals and high correlation

Simu	ME	MAE	PIT(5%,10%,90%,95%)	ACC (T)	FER (F^-)	FIR (F^+)
HSS	-.0004	.103	(3%, 6%,96%,97%)	.942	0	.086
loo	.0002	.101	(2%,5%, 96%, 98%)	1	0	0
proj	-.0001	.104	(2% 6%, 96%, 98%)	1	0	0
95%CI	.0002	.100	(3%, 5%, 95%, 98%)	.965	0	.082

Table 2.6. Scenario 4: strong signals and high correlation

2.7 Case study: SBR estimation

We use the four variable selection methods, HSS, proj, and 95% CI, to carry out variable selection for the SBR estimation problem. In addition to the four methods, we also include variable selection as carried out by UN IGME for producing SBR estimates, which is to select variables based on an absolute cut-off value 0.02 (i.e.,select covariate k if $|\hat{\beta}_k| > 0.02$).

Selected variables and their estimated coefficients are illustrated in Figure A.2. Additional information about the covariates is given in the previous chapter. Estimates of coefficients for subsetted models were obtained from model fits using the selected covariates, with vague priors on the regression coefficients. The selection of covariates varies across the different methods. Five covariates are selected by the HSS method: log NMR, log GNI, log LBW, edu and csec. The 95% CI method selects the smallest number of two covariates: log NMR and log LBW. The log NMR, log GNI, and ANC1 are selected by the projection method. Using an absolute cut-off value of 0.02 results in the selection of log NMR, log GNI, log LBW, edu, csec and anc4. Focusing on HSS, we find that HSS selects all covariates included in the union of covariates selected by other selection methods, with the exception of ANC1/ANC4.

We compare SBR estimates as obtained from the different (subsetted) models and the full model in Figure 2.3. We use the full model, fitted with horseshoe priors as the reference model, and plot the standardized differences between the estimates from the reference model and subsetted model, where the standardized differences is defined as the median of the estimated true SBR from the plotted model divided by the standard deviation of this estimates of the posterior samples from the full model. The differences based on the comparison between the reference model and the CI model have greater variance.

We compare SBR estimates across models for four countries Afghanistan, Ireland, Malawi and Ukraine in Figure 2.4. The estimates are similar across model in most

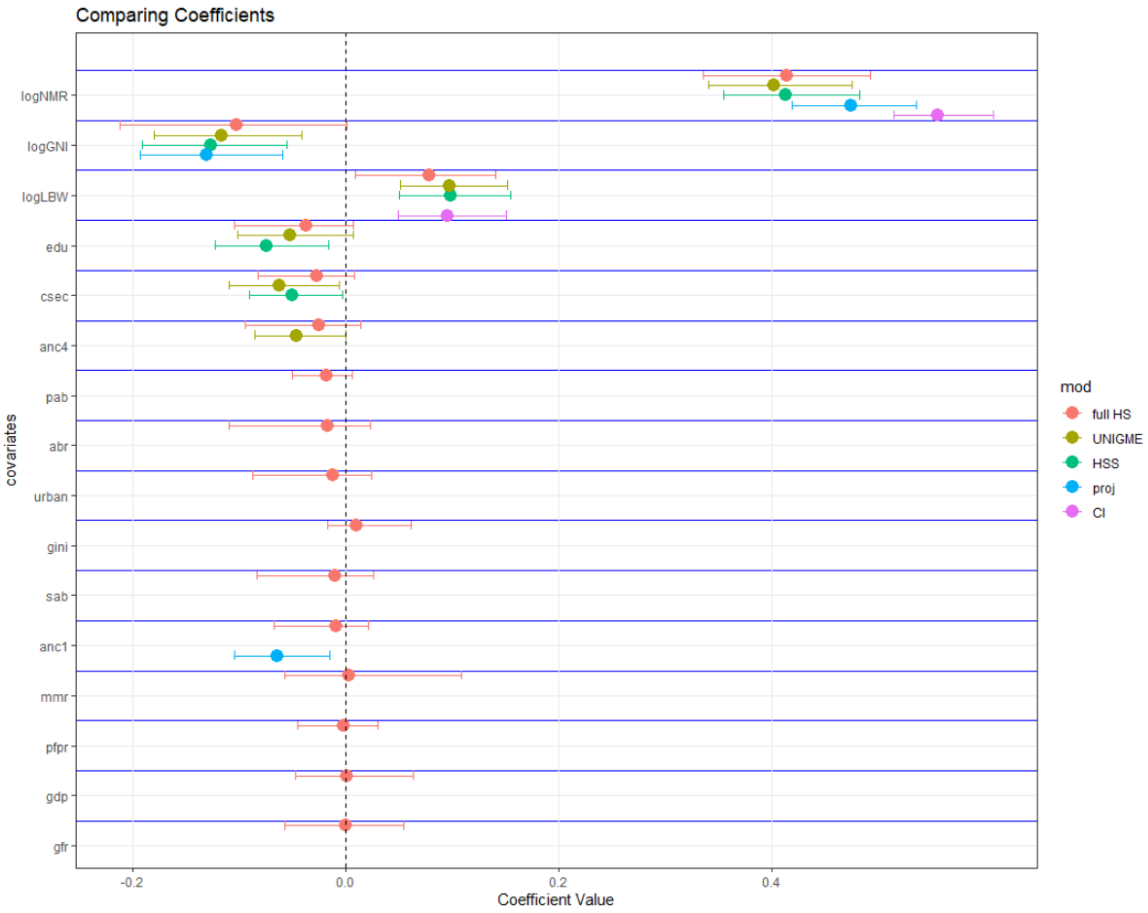


Figure 2.2. Estimates of regression coefficients for full HS model, and for subsetted models obtained by different variable selection methods.

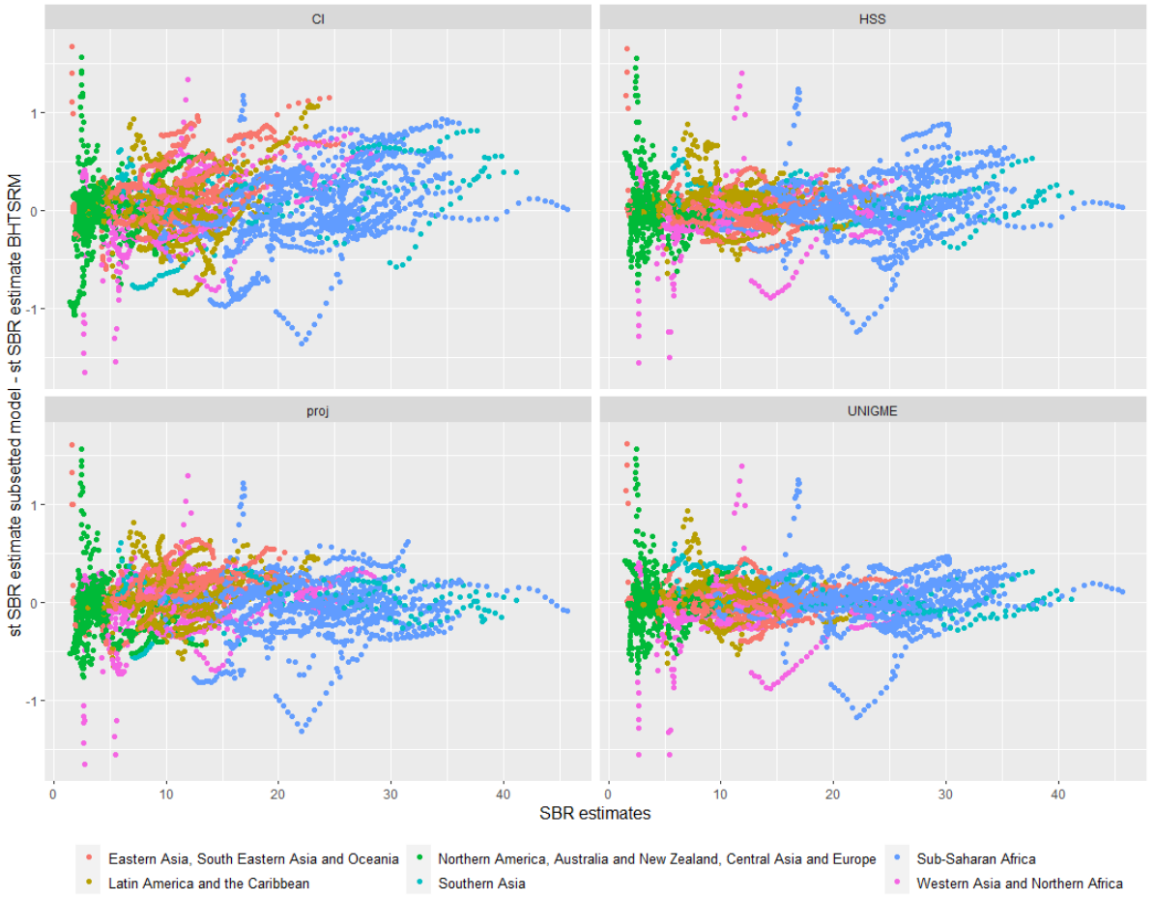


Figure 2.3. Standardized differences between SBR estimates for all countries obtained from the BHTSRM and subsetted models (y-axis), plotted against the SBR estimates from the BHTSRM. The variable selection method is given in the title of each plot. The regions are represented by colors.

country-years and the median are covered by the uncertainty estimated from the BHTSRM, but all estimates are slightly different in Afghanistan where do not have any observed SBR data.

To compare the predictive performance, We use the expected log pointwise predictive density (ELPD) and PSIS diagnostics (See Table 2.7). We find that the CI model has the largest ELPD but there is no significant difference among the models.

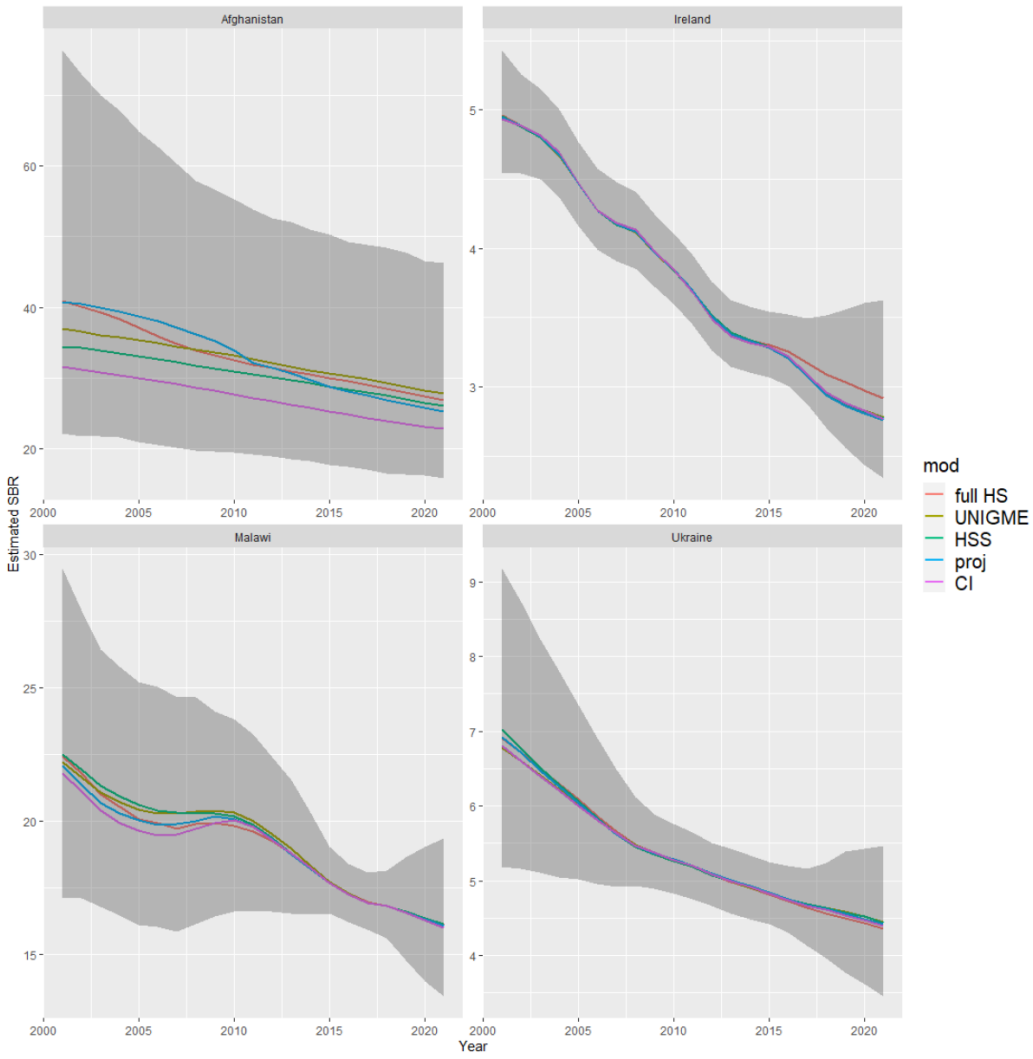


Figure 2.4. SBR estimates (per 1,000 births) of different models for Afghanistan, Ireland, Malawi and Ukraine. The 95% credible interval of full HS model is shown by grey.

Model	ELPD			Pareto K diag.
	estimate	SE	95% CI for difference	
HSS	1195.3	40.3	reference model	0.5%
CI	1198.1	40.6	(-3.8, 9.4)	0.7%
proj	1194.8	40.5	(-6.7 , 5.7)	0.7%
UNIGME	1187.7	40.7	(-124.2, 109.0)	0.6%

Table 2.7. Model comparison based on expected log pointwise predictive density and Pareto K diagnostic values. When comparing models, larger ELPD value suggests better model performance. If 95% CI for difference of ELPD contains 0, it suggests that the difference between two model is not significant. The percentage of high influential points (Pareto K values ¿ 1) for all models are presented in the "Pareto K diag." column, lower outcomes are preferred.

2.8 Discussion

Motivated by the demand for a parsimonious model for estimating stillbirth rates, we explored variable selection methods for Bayesian hierarchical temporal regression models (BHTRMs) and developed a new method. Through simulation exercises, we find that existing approaches may be subject to high false exclusion rates in settings that mimic that of the stillbirth estimation problem, where weak signals may be present. As an alternative option, we developed a horseshoe shrinkage parameter reference distribution variable selection method (HSS-VS) to select predictors. In a simulation study, we find that the new method improves upon the false exclusion rates of other methods in BHTRMs with weak signals, while giving comparable performance in terms of prediction errors and coverage.

We apply the new approach to stillbirth estimation problem and compare its estimates and goodness of fit to a reference BHTRM that is fitted with horseshoe priors. The subsetted model produces estimates that are generally comparable to the full model and has comparable predictive performance. We also show the subsetted models that would be obtained by other variable selection methods that may be subject to increased false exclusion of covariates. As expected, we find that these methods typically select fewer covariates.

The simulation study suggest that the HSS method is a promising approach for variable selection in settings similar to the SBR estimation problem, if the goal is to obtain a parsimonious model with predictive performance that is comparable to the full model, while minimizing false exclusion of covariates with weak signals. We also found that predictive performance of the HSS-VS method was comparable to other methods.

While the HSS-VS method was designed to bound the false inclusion rate, the method may results in higher-than-expected exclusion rates. The HSS-VS method is based on setting a threshold for the expected false inclusion rate. However, the simulation study suggested that the method may result in higher-than-expected rates of false inclusion, especially in settings where covariates are strongly correlated. Further research can focus on assessing in greater detail the properties of the proposed method.

The HSS-VS method is computationally efficient and convenient method to implement. Although the HSS-VS approach requires fitting the full model, it does not require refitting of models, as is needed in other approaches. Moreover, the method does not depend on discrete parameters, like the binary parameter used in the spike-and-slab prior, which facilitates easy use in software like STAN (Stan Development Team [2018],Carpenter et al. [2017]).

Other areas of future work include the assessment of the performance of HSS-VS in setting with larger numbers of covariates, relative to the data availability, and the choice of dummies. Our simulation study was limited to settings that mimicked the SBR application. The performance of HSS has not been tested yet for settings where the number of covariates is close to (or greater than) the number of observations. Another question that can be addressed in more detail is the effect of the choice of the number of dummy variables on predictive performance and selection accuracy, varying the number of dummies and possibly, their correlation with the covariates.

CHAPTER 3

ESTIMATING SURVIVAL PROBABILITIES FROM SSH DATA

In countries with limited civil registration vital statistics data, information on adult mortality can be obtained through asking survey respondents about survival status of their siblings, current age, or age at death and time since death if deceased. These data are referred to as sibling survival history (SSH) data and can be used to estimate survival probabilities. However, SSH data are subject to various reporting errors, i.e. a respondent may misreport the age of their sibling or age at death, resulting in reported survival probabilities that are subject to error.

To estimate true age-cohort specific survival probabilities accounting for related reporting errors, in populations with only SSH data available, we developed a two-stage approach. In stage I, we predict reporting errors obtained from SSH data using covariates related to the respondent and their sibling, and the the error-free validation collected in health and demographic surveillance system (HDSS). Given data sparsity, we explore the use of regression models with horseshoe prior and the approach as proposed in chapter 2 for variable selection in this setting. In stage II, we propose a 2-Dimensional B-spline conditional auto-regression survival (2D-BCAR-surv) model to estimate survival probabilities from SSH data while accounting for reporting errors using the estimates from stage I.

3.1 Introduction

Sibling survival history (SSH) data is a commonly used indirect method to obtain demographic information in countries with limited civil registration vital statistics

(CRVS) systemsGraham et al. [1989]. Many low- and middle income countries suffer from incomplete and limited data from civil registration and vital statistics (CRVS) systems Mikkelsen et al. [2015]. In countries with limited CRVS data, national estimates of adult mortality, for age groups 15-49 by sex, can be derived from information obtained by SSH respondents on their maternal siblings, i.e. siblings which share the same biological mother Helleringer et al. [2014b], Masquelier et al. [2021]. The information includes their siblings' vital status, current age or age at death and time since death for deceased siblings. Age-cohort specific adult mortality and survival rates can be estimated directly using information on current age, age at death and time since death as reported by the surveyed respondents Helleringer et al. [2014b], Masquelier et al. [2021].

A main concern with using SSH data to estimate age-specific mortality rates is the effect of reporting errors on mortality estimates. Reporting errors include misreporting of age, and/or date errors by respondents, vital status, as well as omissions or additions of maternal siblings Pison et al. [2014], Helleringer et al. [2014b], Masquelier [2012]. To assess errors, Helleringer et al. (2014) and Masquelier et al. (2021) evaluated the completeness and accuracy of SSH data in a health and demographic surveillance (HDSS) systems in Bandafassi and Niakhar, in Senegal. HDSS systems monitor over time the entire population located in the geographic area. Taking the data collected in the HDSS system as gold-standard data, and matching SSH data with HDSS data, the studies that compared SSH data with HDSS data identified several kinds of errors in SSH data. List errors occur when a maternal sibling was not reported on by the respondent. Vital status errors occur when respondents reported an incorrect survival status of a sibling at the time of the SSH survey. Finally, age/date errors occur when age at death and/or birth date of sibling were misreported.

The objective of this study is to estimate true age-cohort specific survival probabilities, accounting for SSH related reporting errors, in populations with only SSH data

available. We focus on errors related to SSH reported birth year, and time since death errors. To estimate age-cohort specific adult true mortality and survival rates, we use a two-stage approach. In the first stage, we estimate the distribution of errors in birth year and time since death based on matched SSH and HDSS data from Senegal, In the second stage, we propose a survival model (B-Surv) to estimate age-cohort survival probabilities from SSH data, while accounting for reporting errors. In B-Surv, we parametrize the age and-cohort specific cumulative hazard using a 2-dimensional B-spline regression model. We verify the predictive performance of B-Surv in terms of estimating HDSS-based survival when only SSH data are used for fitting. The contribution of this work includes (1) the two-stage approach to incorporate errors into the estimation of survival rates and the checking of predictive performance of the survival model, and (2) parametrization of the survival function and associated likelihood function for SSH data in terms of cumulative hazard function with the 2 dimensional B-spline setting, to allow for the estimation of survival functions while accounting for errors.

This paper is organized as follow: Section 3.2 summarizes the data used in our analysis of reporting errors. Section 3.3 outlines our proposed method to model misreporting errors in birth year and time since death, and approach to estimate survival probabilities given only SSH data. We summarize results in Section 3.4. Section 3.5 summarizes the limitations of our approach and future work.

3.2 Data

The data used in this study comes from Niakhar, Senegal. In 2013, a SSH survey was conducted among the population of Niakhar HDSS. This section discusses both types of data, as well as covariate data used to inform analysis.

3.2.1 Health and Demographic Surveillance Data

HDSS data are obtained by monitoring an entire population located in a given small geographic area, over time Pison et al. [2014]. These data typically include a baseline census, followed by continuous registration of demographic events including births, deaths, marriages, and migrations. Registration of events occurs yearly (or sometimes more frequently). HDSS data are collected by in-person interviewers who visit every household and ask household informants to provide information on recent events among household membersHelleringer et al. [2014b], Masquelier [2012], Masquelier et al. [2021], Obermeyer et al. [2010].

In this study we used data from the HDSS in Niakhar. This HDSS data set covers the period 1962-2013, in which the population was consistently monitored Delaunay et al. [2013].

3.2.2 Sibling Survival Histories

In SSH data collection, respondents are asked to list all siblings born to their biological mother (maternal siblings) by birth order and then provide information about each sibling's survival status, current age if sibling is reported to be alive, or age at death, as well as time since death if sibling is reported to be deceased.

In 2013, a SSH survey was conducted among the population of Niakhar HDSS. Siblings were matched between SSH and HDSS databases using record linkage, i.e., matching the report of a particular sibling's survival obtained through SSH to the record of that same sibling's survival in the HDSS dataset Masquelier et al. [2021].

3.2.3 Data Summary

Information on SSH birth and time since death errors is obtained by comparing SSH data to HDSS data, which refers to the same (matched) sibling. Matched data are available for 3,974 maternal siblings while 3,046 SSH reported observations (siblings) could not be matched with HDSS Masquelier et al. [2021]. Unmatched siblings may

occur due to respondents including others who are not true maternal siblings, siblings migrated outside of follow-up region, and/or absence of unique identifier information that would allow for linkage to HDSS.

Matched siblings are excluded from the data if they are not alive at the start of the period of interest (1997) or older than age 60, corresponding to exclusion of sibling born before 1947 and those who died before 1997. In addition, we exclude siblings with survival status errors at the start or the end of the reference period: we exclude (i) siblings who are alive at the end of the reference period but reported deceased in SSH data, and (ii) siblings for whom SSH-reported vital status is incorrect at the start of the reference period. The exclusion based on vital status results in a small number of exclusions only (29 siblings excluded).

After data exclusions, the matched data set contains 1,294 observations (siblings) matched between SSH and HDSS data. Table 3.1 and Table 3.2 gives demographic information broken down by data source. The truly living siblings at the end of the reference period, but reported deceased by SSH, are defined as false deceased in Table 3.2.

	SSH	HDSS
Vital Status $\%(n)$		
Living at end of reference period	80 (1029)	79 (1021)
Deceased at end of reference period	20 (265)	21 (273)
Sex $\%(n)$		
Males	64 (842)	64 (842)
Females	36 (452)	36 (452)
Age		
Mean (SD)	31.02(11.09)	31.58 (11.79)
Age at death		
Mean (SD)	28.81 (13.56)	29.53 (13.31)
Years since death		
Mean (SD)	6.53 (4.31)	6.52 (4.47)

Table 3.1. *Demographic characteristics of maternal siblings matched between SSH and HDSS data* $N = 1,294$.

Vital status error	
False deceased (% (n))	0.6% ($n = 8$)
Birth year error	
Mean (SD)	-0.61 (4.17)
Age at death error	
Mean (SD)	-0.74 (4.15)
Time since death error	
Mean (SD)	-0.06 (2.56)

Table 3.2. *Summary statistics of reporting errors of matched siblings*

3.3 Methods

3.3.1 Objective and notation

We develop a 2-stage approach to estimate age-cohort survival probabilities from SSH data collected at time t^*, for some period (t_0, t^*) prior to the time of the survey. In stage I, we use a regression model to estimate the distribution of reporting errors in SSH data, based on linked SSH and HDSS data. In Stage II, we propose a survival model to estimate survival probabilities while accounting for reporting errors, using the estimated joint densities of such errors from Stage I. As such, the second stage model can produce estimates of true survival probabilities for populations with only SSH data available, while accounting for reporting errors.

3.3.1.0.1 Notation We use lowercase Greek letters for unknown parameters and uppercase Greek letters for variables which are functions of unknown parameters (modeled estimates). Roman lower case letters indicate variables that are known or fixed, including data. Roman letters are also used to denote functions.

The period of interest is denoted by (t_0, t^*). For our application in Senegal, $t^* = 2013$ and we set $t_0 = t^* - 15$. The ages of interest are 15 to 60. We are interested in estimating the conditional survival function denoted by $S(b, t|\boldsymbol{\alpha})$, where b refers to time of birth, t to time since t_0, and $\boldsymbol{\alpha}$ to the parameters of the survival function.

For matched sibling i, HDSS-reported data are as given by: (1) vital status, with vital status at time t denoted by $s_i(t)$, with $s_i(t) = 1$, if the sibling is alive at time t; 0 otherwise, (2) birth year b_i, (3) and time since death d_i for deceased siblings. We add superscript $^{(ssh)}$ for data reported in the SSH, e.g., $s_i^{(\text{ssh})}(t)$, $b_i^{(\text{ssh})}$ and $d_i^{(\text{ssh})}$. An example of SSH data is illustrated in the Lexis diagram Demeny et al. [2003] in Figure 3.1. The diagram shows sibling j who is still alive at the end of the reference period and a deceased sibling i.

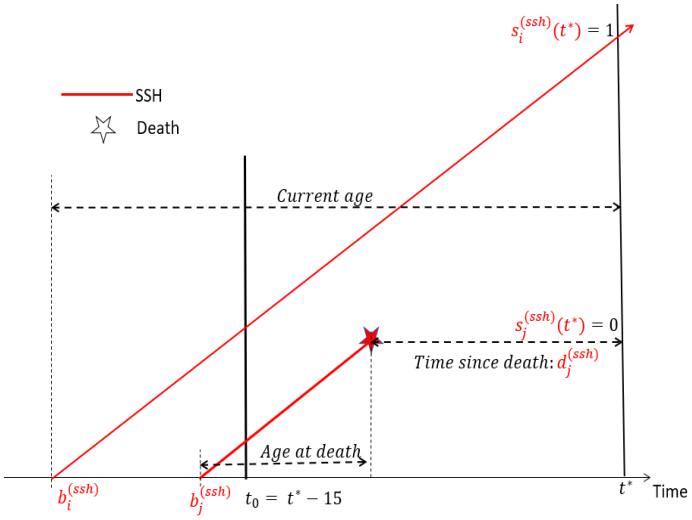


Figure 3.1. Lexis diagram illustration information collected from SSH. x-axis represents the calendar year. For alive sibling i where $s_i^{(ssh)}(t^*) = 1$, the current age is reported. For deceased sibling j where $s_j^{(ssh)}(t^*) = 0$, the age at death and time since death $d_j^{(ssh)}$ are collected in the survey.

3.3.1.0.2 Reporting errors In this study, SSH reporting errors refer to errors in the reporting of the birth year and time since death of a deceased sibling. Information on reporting errors is obtained by comparing SSH reported birth year $b_i^{(ssh)}$ and time since death $d_i^{(ssh)}$ to those reported by HDSS. We defined the error in sibling birth year $e_i^{(birth)}$ as the difference in the SSH reported birth year $b_i^{(ssh)}$ and the HDSS reported birth year b_i for sibling i:

$$e_i^{(birth)} = b_i^{(ssh)} - b_i.$$

Error in time since death applies to deceased siblings only, which is similarly defined as, the difference in the SSH reported time since death $d_i^{(ssh)}$ and the HDSS reported time since death d_i for deceased sibling i:

$$e_i^{(death)} = d_i^{(ssh)} - d_i.$$

Figure 3.2 illustrates differences in reported survival experiences between HDSS (blue) and SSH (red) data. In the case of the living sibling (sib i), error in birth year is captured by the difference between the SSH and HDSS reported birth years shown on the x-axis, $b_i^{(\text{ssh})}$ and b_i, respectively. In the case of the deceased sibling (sib j), in addition to birth year errors, time since death error is captured by the difference between the SSH and HDSS reported time since death shown on the x-axis, $d_j^{(\text{ssh})}$ and d_j, respectively.

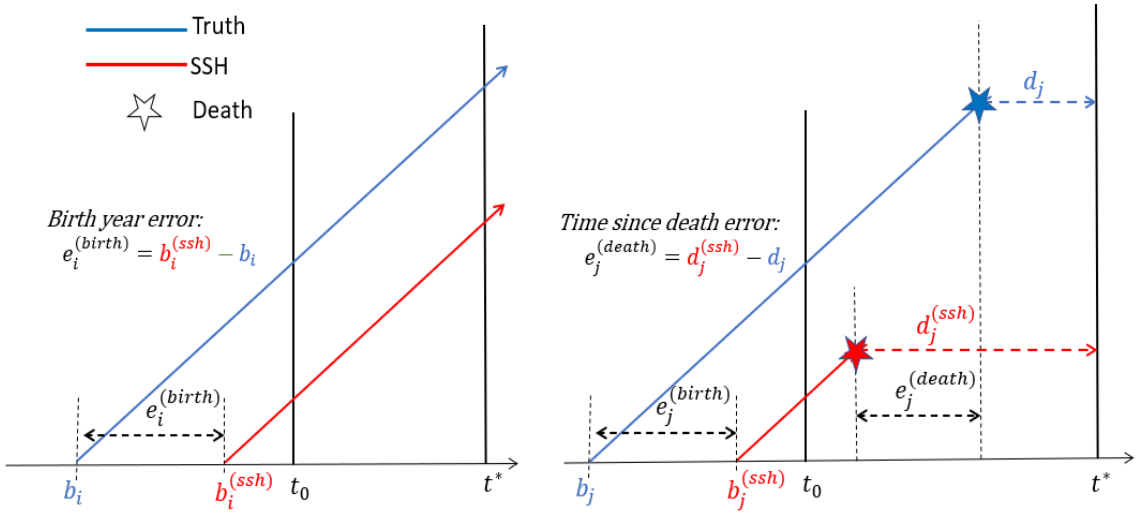


Figure 3.2. Information on reporting errors, alive sibling i (Left) and deceased sibling j (Right). x-axis represents the calendar year. For alive sibling i, $s_i^{(ssh)}(t^*) = s_i(t^*) = 1$, and birth year error $e_i^{(birth)} = b_i^{(ssh)} - b_i$. For deceased sibling j, $s_j^{(ssh)}(t^*) = s_j(t^*) = 0$, the time since death error is $e_j^{(death)} = d_j^{(ssh)} - d_j$.

Summary of 2-stage approach Figure 3.3 illustrates the process by which age-cohort specific probabilities are estimated by the two-stage approach. In the first stage, referred to as the SSH Error Model, we use a regression model, summarized in Section 3.3.2, to estimate the distribution of reporting errors in SSH data, based on linked SSH and HDSS data. In stage II, the SSH Survival (B-Surv) model uses posterior median estimates, obtained from the regression model in stage I, as fixed inputs. In this stage, we estimate the survival function $S(b, t|\boldsymbol{\alpha})$.

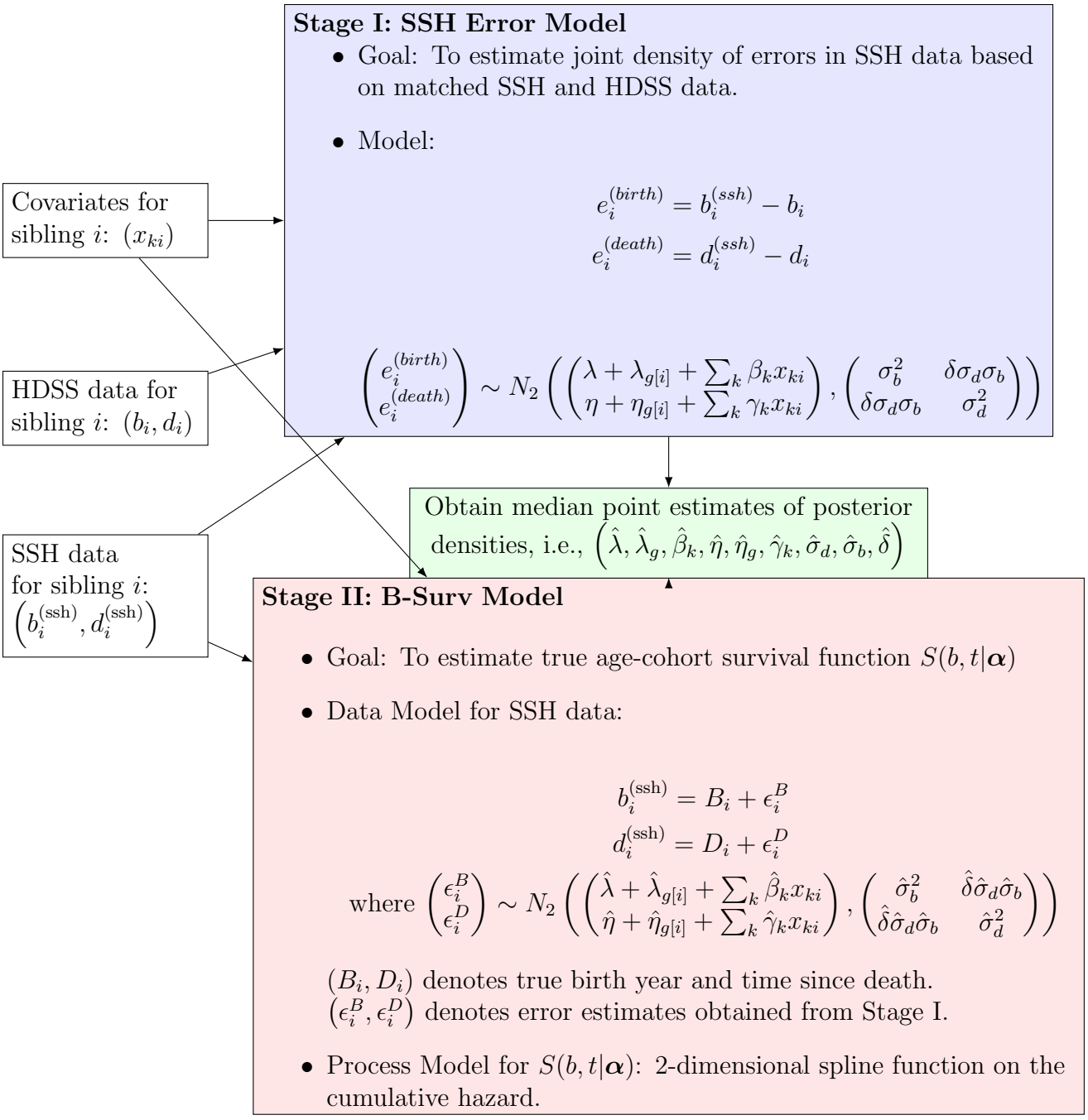


Figure 3.3. Graphical representation of the Bayesian hierarchical two-stage SSH error and survival models. Blue rectangle denotes SSH error model, green rectangle denotes outputs obtained from Stage I, used as inputs in Stage II, and red rectangle denotes SSH survival model). Solid arrows denote stochastic dependency. Hollow boxes contain observed data.

3.3.2 Model for reporting errors

Previous studies Masquelier et al. [2021] found associations between SSH-based reporting errors and several demographic variables. We developed a bivariate hierarchical regression model to predict sibling-level reporting errors for birth year and time since death, using sibling-level covariate information $\boldsymbol{x}_i$.

Eight interviewers are employed to conduct the survey. Because of differences of the personal characterstic, prior data collection experience and language skill, the mean and variance of reporting errors could vary. Thus, we introduce the interviewer effects and estimated it hierarchically centered around the overall intercepts.

Available covariates consisted of birth year, time since death, absolute age difference between sibling and respondent, siblings' sex, indicator of same sex between respondent and sibling, and respondents' sex, which were selected based on a combination of expert knowledge of possible associations, and prior study results Helleringer et al. [2014a], Masquelier et al. [2021]. The relationship between the observed reporting errors and covariates are shown in Figure C.2 in Appendix Section C.1.1.

The model for reporting errors is as follows:

$$\begin{pmatrix} e_i^{(birth)} \\ e_i^{(death)} \end{pmatrix} \sim N_2 \left(\begin{pmatrix} \lambda + \lambda_{g[i]} + \sum_k \beta_k \cdot x_{ki} \\ \eta + \eta_{g[i]} + \sum_k \gamma_k \cdot x_{ki} \end{pmatrix}, \begin{pmatrix} \sigma_b^2 & \delta\sigma_d\sigma_b \\ \delta\sigma_d\sigma_b & \sigma_d^2 \end{pmatrix} \right), \tag{3.1}$$

where this distribution simplifies into a univariate normal for siblings that are alive at the end of the reference period, i.e with $s_i(t^*) = 1$.

We use a regularized horseshoe prior Piironen and Vehtari [2017b] for β_k and γ_k to impose shrinkage of the irrelevant covariates,

$$\beta_k, \gamma_k \sim N(0, \tau^2\tilde{\omega}_k^2), \tag{3.2}$$

$$\tilde{\omega}_k^2 = \frac{\vartheta^2\omega_k^2}{\vartheta^2 + \tau^2\omega_k^2}, \tag{3.3}$$

where τ and ϑ are global shrinkage parameters, and the ω_ks are coefficient-specific parameters with priors,

$$\omega_k \sim C^+(0, \omega_0), \tag{3.4}$$

$$\tau \sim C^+(0, \omega_0), \tag{3.5}$$

$$\vartheta \sim \text{Inv-Gamma}(\vartheta_1, \vartheta_2), \tag{3.6}$$

in which $C^+(0, c)$ refers to a half-Cauchy distribution with location parameter 0 and scale parameter c. We set $\omega_0 = \tau_0 = 1$, $\vartheta_1 = 2$, and $\vartheta_2 = 8$ as per the recommended defaults in Piironen and Vehtari [2017b] and Carvalho et al. [2009c].

From Eq. (3.1) we obtain posterior median estimates of mean, variance, and correlation parameters, which are used as fixed inputs in the B-Surv model further detailed in Section 3.3.3.

3.3.3 The B-surv Model

We aim to estimate true age-cohort specific survival probabilities in the presence of only SSH data. We developed a Bayesian survival model (B-Surv), which accounts for reporting errors. In section 3.3.3.1 we introduce the likelihood in terms of cumulative hazard function with birth year (cohort) b and survival time t since the start of reference period t_0. In Section 3.3.4 we describes the parametrization of the true survival probabilities.

3.3.3.1 Likelihood function

When observing the survival status, time to death, and birth year without errors, and accounting for right censoring (i.e., the sibling is still alive at the end of the period of interest), the likelihood function is given by,

$$p(\boldsymbol{b}, \boldsymbol{t}, \boldsymbol{s}|\boldsymbol{\alpha}) = \prod_{i=1}^{N} \{f(b_i, t_i|\boldsymbol{\alpha})^{1-s_i}[1 - F(b_i, t_i|\boldsymbol{\alpha})]^{s_i}\}, \tag{3.7}$$

$$= \prod_{i=1}^{N} \{h(b_i, t_i|\boldsymbol{\alpha})^{1-s_i} S(b_i, t_i|\boldsymbol{\alpha})\} \tag{3.8}$$

where $f(b, t|\boldsymbol{\alpha})$, $S(b, t|\boldsymbol{\alpha})$, and $h(b, t|\boldsymbol{\alpha})$ refer to the density, survival function, and instantaneous hazard, respectively, at time t for a sibling born in year b, s_i is binary indicator taking value 1 if i-th sibling is alive at end of reference period t^*, b_i refers to the birth year, and t_i refer to the survival time since the start of reference period t_0, i.e. $t = a - (t_0 - b)$, where a refers to the age at death for deceased sibling.

For SSH data, the birth year and survival time t, or equivalently, time since death d, are subject to error and cannot be used directly in the likelihood function. Instead, we assume as per stage I that the reported times are equal to the true time plus error:

$$b_i^{(\text{ssh})} = B_i + \epsilon_i^B, \tag{3.9}$$

$$d_i^{(\text{ssh})} = D_i + \epsilon_i^D, \tag{3.10}$$

where B_i and D_i refer to the true (latent) time of birth and time since death with prior constraints $B_i \sim U(1947, 1997), D_i \sim U(0, 15)$, and the distribution for errors ϵ_i^B and ϵ_i^D is given by the regression model from stage I

$$\begin{pmatrix} \epsilon_i^B \\ \epsilon_i^D \end{pmatrix} \sim N_2 \left(\begin{pmatrix} \hat{\lambda} + \hat{\lambda}_{g[i]} + \sum_k \beta_k x_{ki} \\ \hat{\eta} + \hat{\eta}_{g[i]} + \sum_k \gamma_k x_{ki} \end{pmatrix}, \begin{pmatrix} \hat{\sigma}_b^2 & \hat{\delta}\hat{\sigma}_d\hat{\sigma}_b \\ \hat{\delta}\hat{\sigma}_d\hat{\sigma}_b & \hat{\sigma}_d^2 \end{pmatrix} \right).$$

In this joint distribution, the parameter estimates $(\hat{\lambda}, \hat{\lambda}_{g[i]}, \hat{\beta}_k, \hat{\eta}, \hat{\eta}_{g[i]}, \hat{\gamma}_k, \hat{\sigma}_d^2, \hat{\sigma}_b^2, \hat{\delta})$ are obtained from stage I using Eq. (3.1). Taking account of reporting errors in SSH data, the likelihood function from Eq. (3.8) can be written in terms of the latent birth year B, and time $T = 15 - D$ as follows:

$$p(\boldsymbol{B}, \boldsymbol{T}, \boldsymbol{s}|\boldsymbol{\alpha}) = \prod_{i=1}^{N} \{h(B_i, T_i|\boldsymbol{\alpha})^{1-s_i} S(B_i, T_i|\boldsymbol{\alpha})\}. \tag{3.11}$$

Inference using the likelihood in Eq. (3.8) is challenging due to the assumption of a time-varying hazard, which means there is no closed form expression for the likelihood function. To implement a computationally efficient model, we opt to parametrize the likelihood function in terms of the cumulative hazard $H(b, t)$. The reparametrized likelihood function is as follows:

$$L(F) = \prod_{i=1}^{N} \{f(B_i, T_i|\boldsymbol{\alpha})^{1-s_i^{(\text{ssh})}} [1 - F(B_i, T_i|\boldsymbol{\alpha})]^{s_i^{(\text{ssh})}}\} \tag{3.12}$$

$$= \prod_{i=1}^{N} \{h(B_i, T_i|\boldsymbol{\alpha})^{1-s_i^{(\text{ssh})}} S(B_i, T_i|\boldsymbol{\alpha})\} \tag{3.13}$$

$$= \prod_{i=1}^{N} \left\{ \frac{\partial H(B_i, T_i|\boldsymbol{\alpha})}{\partial t} \}^{1-s_i^{(\text{ssh})}} \{\exp(-H(B_i, T_i|\boldsymbol{\alpha}))\right\} \tag{3.14}$$

$$\log L(F) = \sum_{i=1}^{N} \left\{ (1 - s_i^{(\text{ssh})}) \log \frac{\partial H(B_i, T_i|\boldsymbol{\alpha})}{\partial t} - H(B_i, T_i|\boldsymbol{\alpha}) \right\} \tag{3.15}$$

where $\frac{\partial H(b,t)}{\partial t}$ is the partial derivative of $H(b, t)$ respect to t.

3.3.4 B-Surv process model: parameterization of cumulative hazard function

We parametrize the cumulative hazard using a 2 dimensional cubic spline function, to capture age and cohort specific effects:

$$H(b, t) = \sum_{l=1}^{L} \sum_{p=1}^{P} \tilde{\alpha}_{l,p} G_l(b) G_p(t), \tag{3.16}$$

with $G_l(b)$ and $G_p(t)$ referring to the l-th spline function evaluated at cohort (birth) b and p-th spline function evaluated at survival time t since t_0, and $\tilde{\alpha}_{l,p}$ referring to

the spline coefficients (see Figure 3.4) capturing changes in the cohort versus time direction, respectively.

We use equally spaced cubic B-splines with knots spaced 3 years apart in the reference period $[t_0, t^*]$, and with knots spaced 5 years apart in the cohort (birth year) interval $[1947, 1997]$.

The partial derivative of $H(b,t)$ with respect to t is given by

$$\frac{\partial H(b,t)}{\partial t} = \sum_{l=1}^{L}\sum_{p=1}^{P} \tilde{\alpha}_{l,p} G_l(b) \frac{\partial G_p(t)}{\partial t} \tag{3.17}$$

Definition of B-spline: Given the knots vector $\boldsymbol{t} = (t_1, t_2, ..., t_p, ..., t_P)$. B-spline function of d degree ($d = 3$ for cubic spline) is defined as

$$G_p^d(t) = \frac{t - t_p}{t_{p+d} - t_p} G_p^{d-1}(t) + \frac{t_{p+d+1} - t}{t_{p+d+1} - t_{p+1}} G_{p+1}^{d-1}(t), \tag{3.18}$$

and

$$G_p^0(t) = \begin{cases} 1 & \text{, if } \quad t \in [t_p, t_{p+1}) \\ 0 & \text{, if } \quad t \notin [t_p, t_{p+1}) \end{cases}$$

Derivative of B-spline: The derivative $\frac{\partial G_p(t)}{\partial t}$ is given by

$$\frac{\partial G_p^d(t)}{\partial t} = \frac{d}{t_{p+d} - t_p} G_p^{d-1}(t) - \frac{d}{t_{p+d+1} - t_{p+1}} G_{p+1}^{d-1}(t) \tag{3.19}$$

The coefficients are written as $\tilde{\alpha}_{l,p} = \exp(\alpha_{l,p})$, such that the coefficients and resulting cumulative hazard function $H(b,t)$ is non-negative. To ensure the cumu-

lative hazard is monotone increasing with the increase of time (age), we require the additional constraints:

$$\alpha_{l,p} \mid \alpha_{l,p-1}, \alpha_{l,p+1} \sim U(\alpha_{l,p-1}, \alpha_{l,p+1}). \tag{3.20}$$

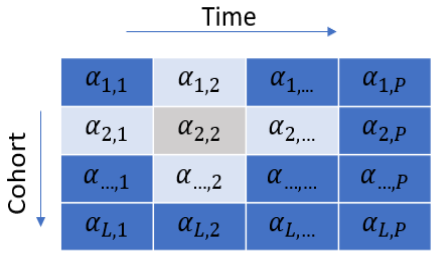


Figure 3.4. Spline coefficients α, capturing changes across cohorts and time.

Conditional Autoregressive (CAR) model We parameterize the $\alpha_{l,p}$ using a conditional autoregressive (CAR) model De Oliveira [2012]:

$$\alpha_{lp} \mid \alpha_{jk} \sim N(\chi \sum_{jk}^{JK} w_{lp,jk}\alpha_{jk}, \phi_{lp}^{-1}) \tag{3.21}$$

where α_{jk}s are the neighbors of α_{lp} such that the expectation of α_{lp} is a function in terms of average of α_{jk}, ϕ_{lp} refers to a spatially varying precision parameter, and $w_{lp,jk}$ is the adjacency indicator ($w_{lp,jk} = 1$ if α_{lp} is a neighbor of α_{jk}. See Figure 3.4), $w_{lp,lp} = 0$.

Conditional autoregressive (CAR) models are popular as prior distributions for spatial random effects with areal spatial data. Historically, MCMC algorithms for CAR models have benefited from efficient Gibbs sampling via full conditional distributions for the spatial random effects. But, these conditional specifications do not work in Stan, where the joint density needs to be specified (up to a multiplicative constant). CAR models can still be implemented in Stan by specifying a multivariate

normal prior on the spatial random effects, parameterized by a mean vector and a precision matrix Morris et al. [2019]. We reparameterize the model as,

$$\alpha_{lp} \sim N(0, [\phi C(I - \chi C^{-1}W)]^{-1}) \tag{3.22}$$

where C is an $LP \times LP$ diagonal matrix with $diag(C)_{lp}$ = the number of neighbors for α_{lp}, χ is a parameter that controls dependence (with $\chi = 0$ implying independence while $\chi = 1$ collapses to an intrinsic conditional autoregressive specification), and W is the adjacency matrix.

Then CAR prior specification simplifies to

$$\alpha_{lp}| \sim N(0, [\phi(C - \chi W)]^{-1}), \tag{3.23}$$

with priors

$$\phi \sim \text{Gamma}(2, 2), \tag{3.24}$$

$$\chi \sim U(0, 1). \tag{3.25}$$

3.3.5 Computation

A HMC algorithm is employed to sample from the posterior distribution of the parameters of both model for reporting errors and B-surv model with the use of Stan Carpenter et al. [2017] and R package Rstan Stan Development Team [2018]. Four parallel chains are run with a total of 1,000 iterations in each chain. The first 500 iterations in each chain are discarded as burn-in so that the resulting chains contain 2,000 samples each. Point estimates are given by medians of the posterior samples. Standard diagnostic checks are used to check convergence and sampling efficiency. These checks are based on trace plots, the improved Rhat diagnostic using rank-normalized draws (Gelman and Rubin [1992], Vehtari et al. [2020b]), and various

calculations of effective sample size (ESS), including the bulk ESS and the tail ESS - giving the minimum of the effective sample sizes of the 5% and 95% quantiles.

3.3.6 Model evaluation and validation

Based on the objective of this study, we focus on assessing whether B-Surv, when fitted to SSH data in stage II, predicts well the HDSS data and associated survival probabilities. We first compare fits of the survival model to

1. SSH data, not accounting for errors;
2. SSH data, accounting for errors;
3. DHSS data.

Model fit 2 corresponds to B-Surv as explained in the previous section. Model fit 1 is added to show the effect of accounting for errors. Model fit 3 is added to check whether survival estimates based on the error-free data are comparable to those based on the SSH data, accounting for error.

To formalize the comparison and assess model performance, we use model 2 to predict birth year and time of death for deceased siblings, and compare it with the true HDSS reported birth year and time of death, and summarize predictive performance in terms of prediction errors and by checking probability integral transform (PIT). The calculation of these measures for birth year and time of death are explained in more detail in the remainder of this section.

3.3.6.1 Evaluating predictions of the birth year

To evaluate model performance, we summarize differences between the true HDSS-reported birth year b_i and the posterior predictive distribution obtained from B-Surv in model fits 2. The procedure is as follows:

1. Fit the B-Surv model to SSH data and sample $B_i^{(s)}$ for $s = 1, 2, ..., S$ from its respective posterior distribution.

2. Summarize prediction errors: Calculate the difference between the true (HDSS-reported) year and estimate birth year for sibling i:

$$\text{err}_i^{(birth)} = \hat{B}_i - b_i. \tag{3.26}$$

We report the median and mean square error of the errors.

3. Check PIT/coverage: PIT values are defined of birth year is defined as $P_\kappa = N^{-1} \sum_{i=1}^{N} 1[b_i \leq u_i]$, where u_i refers to the κ percentile of the posterior distribution of B_i from the model fit. We use the same procedure to perform the posterior predictive checks for the survival time T_i since t_0 as well.

3.3.7 Evaluating estimated survival parameters

We perform the posterior predictive checks for the age-cohort specific survival function $S(b, t|\boldsymbol{\alpha})$. The HDSS reported survival time t_i among deceased siblings, given HDSS reported birth year b_i should obtained from the following estimated cumulative density function if it captures the true survival probabilities (leaving out the parameters of spline function $\boldsymbol{\alpha}$ for simplicity of notation):

$$F(b_i, t \mid t \leq t^* - t_0) = \frac{F(b_i, t)}{F(b_i, t^* - t_0)} \tag{3.27}$$

$$= \frac{1 - S(b_i, t)}{1 - S(b_i, t^* - t_0)} \tag{3.28}$$

$$= \frac{1 - \exp(-H(b_i, t))}{1 - \exp(-H(b_i, t^* - t_0))}. \tag{3.29}$$

We can assess the accuracy of the posterior predictive distribution of the survival function using a probability integral transformation. We obtain $u_i^{(s)}$ for each deceased sibling i and posterior sample s as follows:

$$u_i^{(s)} = 1 - \frac{1 - \exp(-H(b_i, t_i|\boldsymbol{\alpha}^{(s)}))}{1 - \exp(-H(b_i, t^* - t_0|\boldsymbol{\alpha}^{(s)}))}, \tag{3.30}$$

where

$$H(b, t|\boldsymbol{\alpha}^{(s)}) = \sum_{l=1}^{L}\sum_{p=1}^{P} \exp(\alpha_{l,p}^{(s)}) G_l(b) G_p(t) \tag{3.31}$$

We assess the distribution of $u_i^{(s)}$, which should follow a $Unif(0,1)$ distribution if the model is well calibrated. If the $u_i^{(s)}$'s have greater density mass located close to 1 than expected under the uniform density, then this means that siblings tend to have a survival time reported in the DHSS that is greater than expected under the model.

3.4 Results

3.4.1 Reporting errors

Table 3.3 lists posterior estimates and 95% credible intervals (CI), for the parameters of the regression model for the SSH reporting errors in birth year and time of death (regression coefficients, standard deviations $(\hat{\sigma}_b, \hat{\sigma}_d)$ and correlation $\hat{\delta}$). There is a negative association between sibling birth year and birth year error ($\hat{\beta}$ = -1.06 with 95% CI given by (-1.27, -0.85), and positive association between birth year and time since death error (0.30 (-0.01, 0.62)). The standard deviation of birth year σ_b and time since death error σ_d and their 95% CI are given by 2.22 (2.16, 2.27) and 2.28 (2.10, 2.52), and the correlation between birth year and time since death error is given by -0.02 (-0.11, 0.06).

Figure C.3 in the appendix shows residual plots, with residuals for the birth year error, and time since death error, plotted against covariates. The plots do not indicate residual trends.

Coefficients/parameters	Birth year error		Time since deat	
	Est.	95% CI	Est.	9
Intercept	0.50	(-2.82,3.70)	1.12	(-1
Birth year	-1.06	(-1.27,-0.85)	0.30	(-0
Time since death	0.06	(-0.06,0.21)	-0.67	(-0
Sibling sex (ref: female)	0.05	(0.08,0.24)	0.06	(-0
Respondent sex (ref: female)	0.01	(-0.18,0.15)	-0.02	(-0
Difference in sex (ref: same sex)	-0.12	(-0.36,0.04)	0.10	(-0
Difference in age (sib vs. resp)	-0.05	(-0.25,0.09)	0.06	(-0
standard deviation (σ_b, σ_d)	2.22	(2.16,2.27)	2.28	(2
correlation δ	-0.02	(-0.11,0.06)	-	

Table 3.3. Posterior estimates of regression parameters: median estimate (Est.) and 95% credible intervals (95% CI).

3.4.2 B Surv model fits

As explained in the methods section, based on the objective of this study, we focus on assessing whether B-Surv, when fitted to SSH data in stage II, predicts well the HDSS data and associated survival probabilities. We compare fits of the survival model to

1. SSH data, not accounting for errors;
2. SSH data, accounting for errors;
3. DHSS data.

Figure 3.5 illustrates the estimated survival functions from model fits 2 and 3. There are differences between survival when using HDSS and SSH data accounting for errors. We see notable differences between model fits in earlier cohorts, i.e., 1951 cohorts. The median of the model fit to SSH data while accounting for the reporting error (blue) is slightly higher than the median of model fit to HDSS data (red) for short time t and is smaller for later time, and it has more uncertainty than model fit to HDSS data.

Figure 3.6 illustrates the estimated survival function from fits 1 and 2. We see slightly difference of median, but more uncertainty for fit 2 (blue) than fit 1 (red).

Figure 3.7 compares the fit to HDSS and SSH data, respectively, ignoring reporting errors. When not accounting for errors, the SSH-based fit is different but differences appear modest, i.e., when comparing medians when taking into account uncertainty in the survival curves.

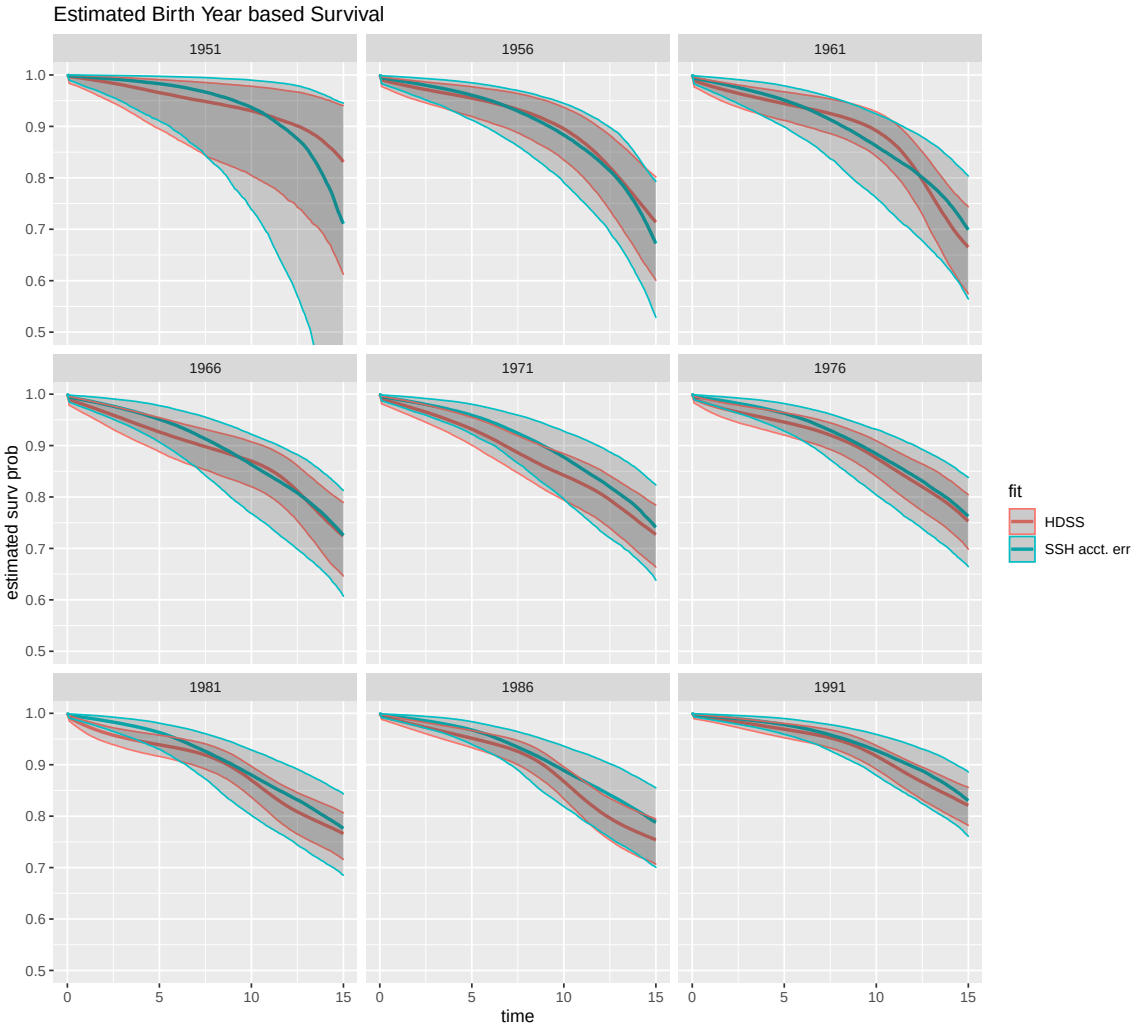


Figure 3.5. Model comparison between fit to HDSS data and fit to SSH data while accounting for errors.

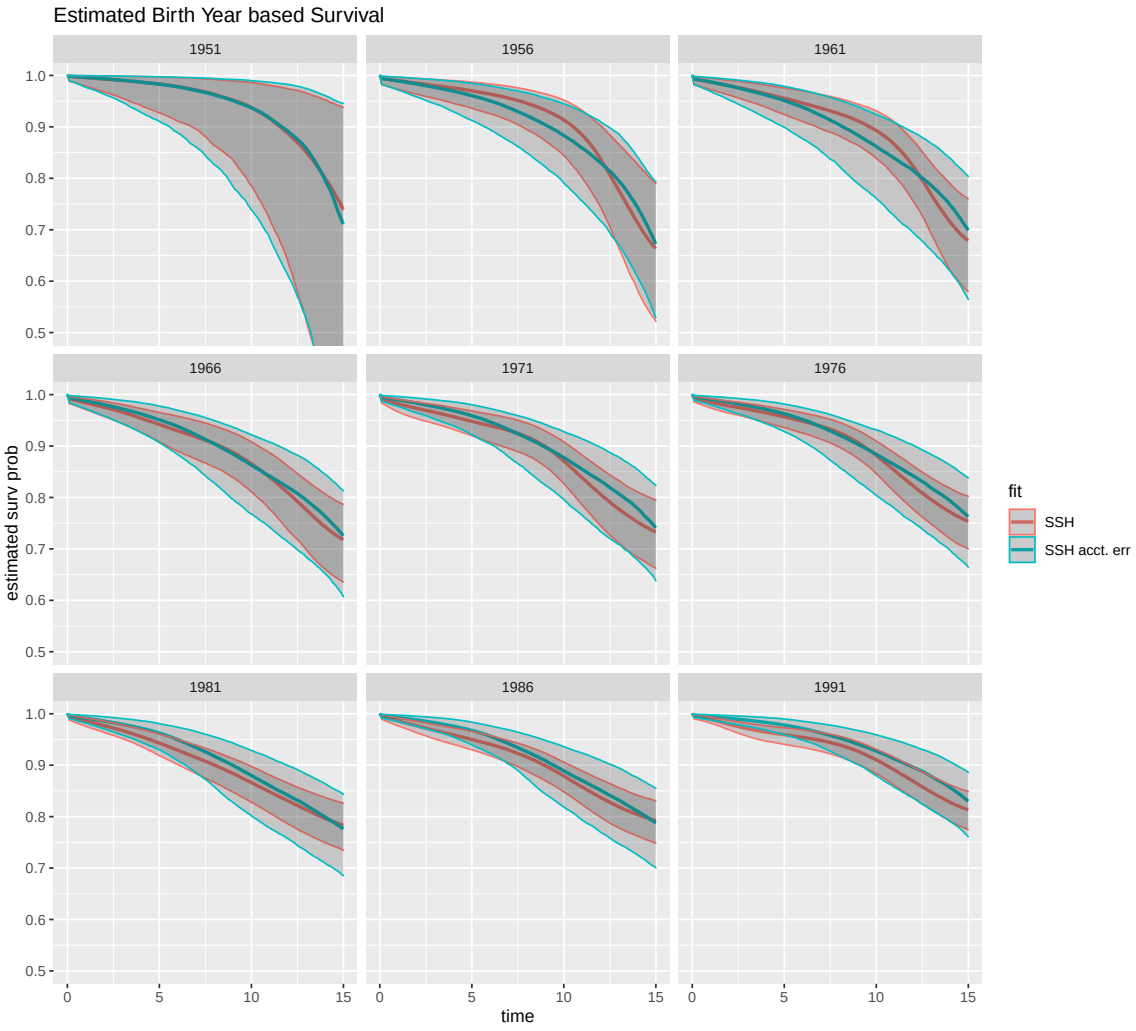


Figure 3.6. Model comparison between fit to SSH data and fit to SSH data without accounting for errors.

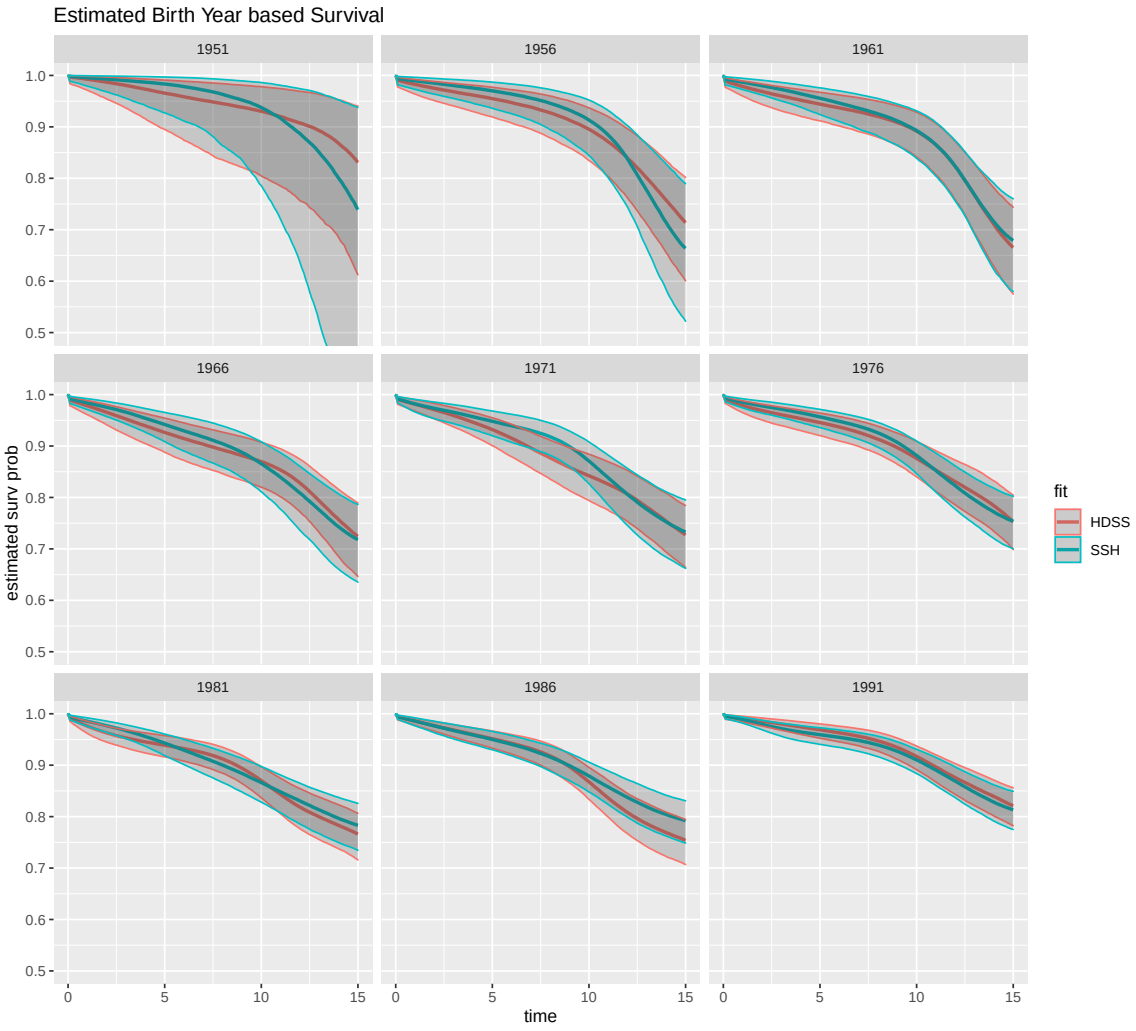


Figure 3.7. Model comparison between fit to HDSS data and fit to SSH data without accounting for errors.

3.4.3 B-surv model validation

Table 3.4 summarizes the residuals defined in Section 3.3.6.1 in terms of mean error and mean square error for model 2 for predicting birth years and time since death. We find that estimated survival time has average positive bias of 0.321, which results in an overestimated sibling survival time compared to the truth. Additionally, the estimated birth year shows a negative bias of -0.125, which yields an underestimation of sibling age.

	mean	mse	PIT(5%, 95%)	$\text{cov}_{95\%}$
$\hat{err}_i^{(birth)}$	-0.125	19.1	(7.2%,94.0%)	91.8%
$\hat{err}_i^{(death)}$	0.321	4.79	(6.0%, 94.0%)	93.0%

Table 3.4. The summary of residual for estimate birth year and survival time. PIT(5%, 95%) represents the PIT with the expected values 5% and 95%. $\text{cov}_{95\%}$ represents the 95% coverage rate.

Figure 3.8 illustrate the residual plot of estimated true birth year and estimated true survival time since t_0. The figure shows the birth year residuals distributed around zero approximately with mean error at -0.125 (see Table 3.4). But we do see the residuals deviate from zero at the extreme ends of the true birth year.

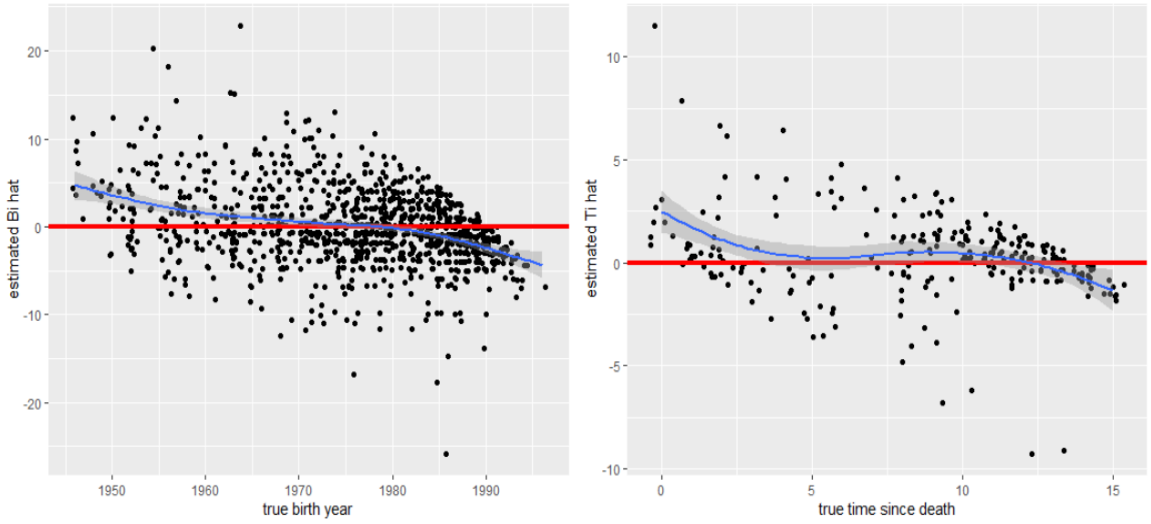


Figure 3.8. Residual check for estimated birth year B_i and estimated survival time T_i since t_0 for deceased sibling.

The predictive distribution of birth year and time of death is also assessed based on coverage and PIT summaries. Coverage of 95% prediction intervals, as summarized in Table 3.4, suggests that intervals, for both birth year and time to death errors, were slightly narrower than expected for nominal coverage. The histogram with PIT values for birth years and time at death in Figure 3.9 do not show obvious trend.

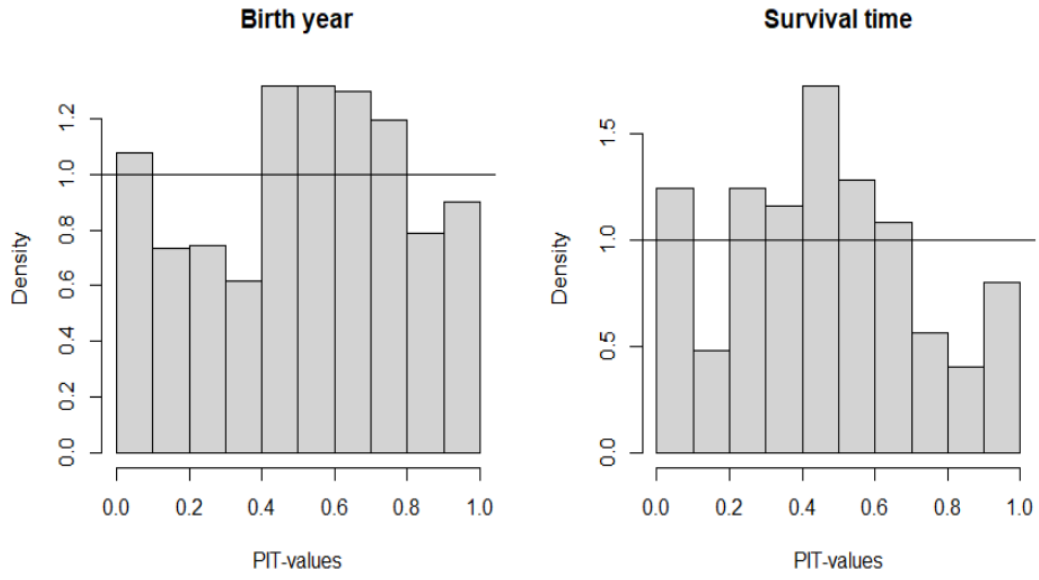


Figure 3.9. An in-sample probability integral transform (PIT) of birth year and survival time in model fit to SSH data while accounting for errors.

The histogram with PIT values for time of death among deceased siblings in Figure 3.10 shows mass being distributed around the center. This suggests that the survival function/parameter is conservative which is consistent with the survival curves.

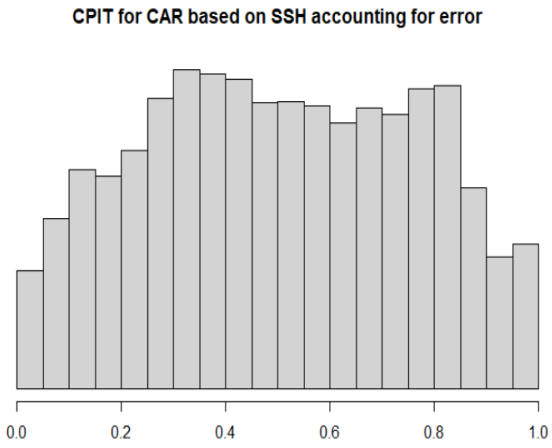


Figure 3.10. Probability integral transform check of survival probability for deceased siblings given HDSS reported birth year b_i and HDSS reported survival time t_i.

3.5 Discussion

We developed a Bayesian model to estimate cohort-age specific survival probabilities using SSH data, while accounting for errors in such data. We used a 2-stage approach to produce estimates. In stage 1, we developed a bivariate hierarchical regression model to predict errors in SSH reporting of the birth year of siblings, and time since death for deceased siblings. In the second stage, we developed a Bayesian B-spline survival model to estimate cohort-age specific survival probabilities in terms of cumulative hazard function.

Current model validation exercises suggest that the model fit to SSH data while accounting for error is reasonable well calibrated. The estimates survival probability capture the true HDSS data. The conservative uncertainty is as our expected given we have extra uncertainty by introducing both bias and variance of the reporting errors.

While our approach to estimating adult mortality from SSH data is the first to explicitly account for reporting errors, there are limitations to the model due to data availability limitations. Limited data availability restricted the analyses and resulted in simplifying model assumptions. The main limitation is that we were not able to account for errors introduced by unmatched siblings, referring to siblings that were reported on by the respondent but not found in the HDSS data, as well as omitted siblings. Prior studies suggest that deceased siblings are more likely to be omitted by the respondents Masquelier et al. [2021]. Thus, omitted siblings will introduce additional errors and bias estimates. Another limitation relates to the model assumptions, i.e., the CAR setup is symmetric, which means it put the same weight on both cohort direction and time (aging) effect. We would consider to add extra parameter to allow asymmetric CAR setup in the future.

APPENDIX A

SUPPLEMENTARY TABLES AND FIGURES FOR CHAPTER 1

A.1 Tables and figures

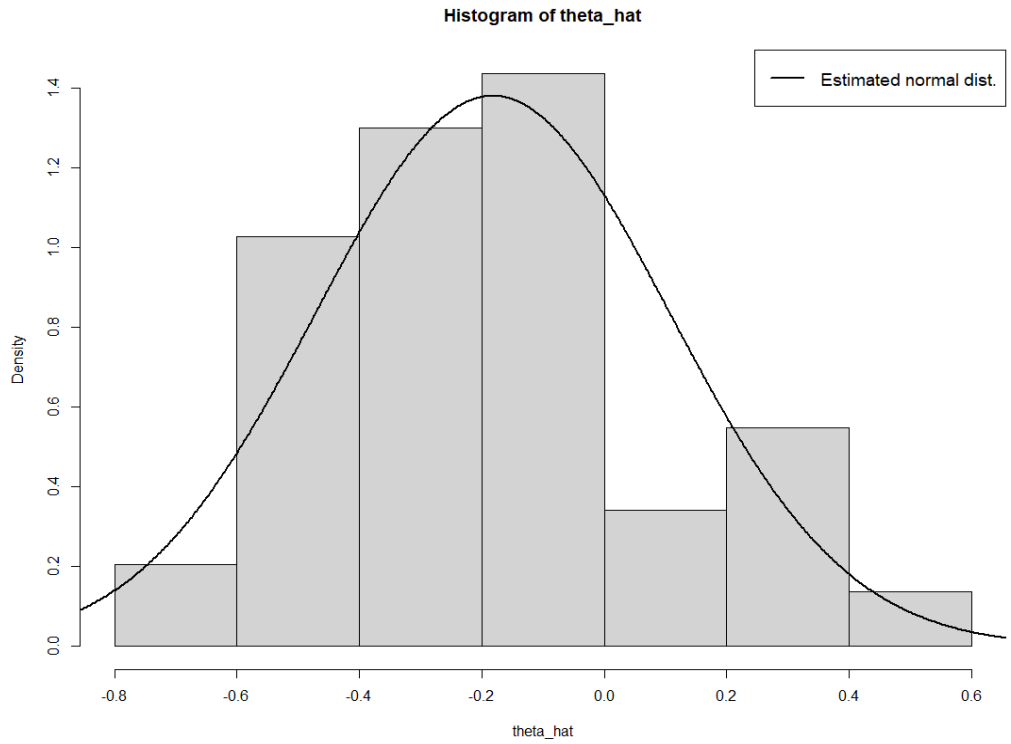


Figure A.1. Histogram of estimates of θ_i. There are 73 observations.

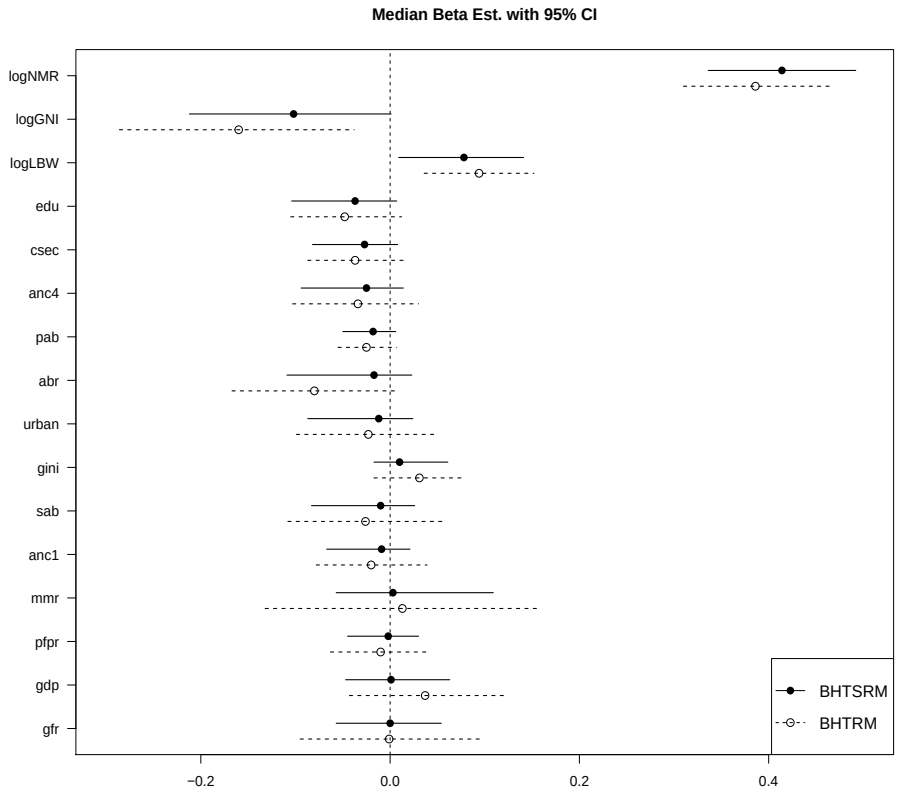


Figure A.2. Estimates of β and 95% CI for reference BHTSRM and BHTRM.

	BHTSRM		BHTRM	
covariates	estimate	sd	estimate	sd
log(nmr)	0.414	0.040	0.430	0.042
log(gni)	-0.102	0.056	-0.165	0.061
log(lbw)	0.078	0.033	0.093	0.030
edu	-0.037	0.031	-0.045	0.030
csec	-0.027	0.025	-0.026	0.025
anc4	-0.025	0.029	-0.043	0.035
pab	-0.018	0.015	-0.022	0.016
abr	-0.017	0.035	-0.031	0.031
urban	-0.012	0.029	-0.019	0.037
gini	0.010	0.020	-0.012	0.025
sab	-0.010	0.027	0.041	0.052
anc1	-0.009	0.022	-0.031	0.031
mmr	0.003	0.038	0.028	0.077
pfpr	-0.002	0.017	-0.037	0.028
gdp	0.001	0.025	0.031	0.034
gfr	0.000	0.025	0.026	0.054

Table A.1. Overview of estimates for regression coefficients under reference BHTSRM and BHTRM. "estimate" and "sd" represent the median and standard deviation of posterior samples.

Validation Desirable*	N.test -	Mean err. 0	Mean abs. err. -	be 5% 5%	be 10% 10%	ab 90% 10%	ab 95% 5%
BHTSRM Recent	112	-0.001	0.091	3.5%	6.2%	2.7%	1.8%
BHTRM Recent	112	-0.001	0.091	2.7%	5.3%	2.7%	1.8%
BHTSRM In-sample	1531	-0.002	0.090	1.8%	3.5%	4.2%	1.6%
BHTRM In-sample	1531	-0.001	0.090	1.8%	3.5%	4.1%	1.8%

Validation Desirable*	ICgroup -	N.test -	Mean err. 0	Mean abs. err. N/A	be 5% 5%	be 10% 10%	ab 90% 10%	ab 95% 5%
BHTSRM In-sample	High	777	-0.002	0.089	2.2%	4.3%	4.1%	1.5%
	Low	754	0.001	0.090	1.3%	2.7%	4.2%	1.7%
BHTRM In-sample	High	777	-0.002	0.090	2.2%	4.4%	4.3%	1.5%
	Low	754	0.001	0.090	1.3%	2.5%	4.0%	2.0%
BHTSRM recent	High	48	0.008	0.082	4.2%	6.3%	2.1%	2.1%
	Low	65	-0.014	0.099	3.1%	6.2%	3.1%	1.5%
BHTRM recent	High	48	0.006	0.084	4.2%	6.3%	2.1%	2.1%
	Low	65	-0.008	0.099	1.5%	4.6%	3.1%	1.5%

Table A.2. Validation result aggregated across all countries (top) and broken down by high and low income countries (bottom). "N.test" represent the number of observations in the test set. Validation exercises "Recent" and "In-sample" represent leaving out recent observations, and approximate leave-one-out validation. The outcome measures are as follows: mean of error, mean absolute error, and % of left-out observations below (be) and above (ab) their respective 90% and 80% prediction intervals. "BHTRM" validation are results for models with vague prior on β. Desirable* refers to outcomes for models that are unbiased and well calibrated.

A.2 List of covariates

Table A.3: Candidate covariates with its sources and methodology

Var	Source	Definition	Methodology Notes
abr	United Nations Department of Economic and Social Affairs (DESA), Population Division United Nations Population Fund (UNFPA). Data are based on DHS, MICS and other national household surveys	Adolescent Birth Rate (number of live births to adolescent women per 1,000 adolescent women)	Extrapolated to 2019 assuming a flat trend, linear interpolation applied when data between 2000-2019 unavailable, imputation using regional year data for countries without any available data, and smoothing applied.
anc1	UNICEF/WHO. Data are based on DHS, MICS and other national household surveys	Antenatal care 1+ visit - Percentage of women (age 15–49) attended at least once during pregnancy by skilled health personnel.	Extrapolated to 2019 assuming a flat trend, linear interpolation applied when data between 2000-2019 unavailable, imputation using regional year data for countries without any available data, and smoothing applied.
Continued on next page			

Table A.3 – continued from previous page

Var	Source	Definition	Methodology Notes
anc4	UNICEF/WHO. Data are based on DHS, MICS and other national household surveys	Antenatal care 4+ visits - Percentage of women (age 15–49) attended at least four times during pregnancy by any provider.	Extrapolated to 2019 assuming a flat trend, linear interpolation applied when data between 2000-2019 unavailable, imputation using regional year data for countries without any available data, and smoothing applied.
csec	UNICEF. Data are based on DHS, MICS and other national household surveys	C-section rate - Percentage of deliveries by Caesarian section.	Extrapolated to 2019 assuming a flat trend, linear interpolation applied when data between 2000-2019 unavailable, imputation using regional year data for countries without any available data, and smoothing applied.

Continued on next page

Table A.3 – continued from previous page

Var	Source	Definition	Methodology Notes
gfr	United Nations Department of Economic and Social Affairs (DESA), Population Division, World Population Prospects 2019 Edition	General fertility rate. Number of live births divided by the female population age 15-49 years.	No additional processing applied.
gdp	World Bank	World Bank Gross domestic product per capita	No additional processing applied.
gini	World Bank, Development Research Group. Data are based on primary household survey data obtained from government statistical agencies and World Bank country departments.	Gini index measures the extent to which the distribution of income (or, in some cases, consumption expenditure) among individuals or households within an economy deviates from a perfectly equal distribution.	Extrapolated to 2019 assuming a flat trend, linear interpolation applied when data between 2000-2019 unavailable, imputation using regional year data for countries without any available data, and smoothing applied.

Continued on next page

Table A.3 – continued from previous page

Var	Source	Definition	Methodology Notes
gni	World Bank, International Comparison Program	Gross national income per capita	Extrapolated to 2019 assuming a flat trend, linear interpolation applied when data between 2000-2019 unavailable, and imputation using regional year data for countries without any available data.
lbw	UNICEF/WHO estimates, 2019 Edition. Data based on vital registration data and national household surveys	Percentage of live births that weighted less than 2500 grams.	Extrapolated to 2019 assuming a flat trend, linear interpolation applied when data between 2000-2019 unavailable, imputation using regional year data for countries without any available data, and smoothing applied.
edu	United Nations Development Programme. Data are based Barro and Lee (2013), UNESCO Institute for Statistics (2013).	Average number of years of education received by people ages 25 and older, converted from educational attainment levels using official duration of each level.	Extrapolated to 2019 assuming a flat trend, linear interpolation applied when data between 2000-2019 unavailable, imputation using regional year data for countries without any available data, and smoothing applied.

Continued on next page

Table A.3 – continued from previous page

Var	Source	Definition	Methodology Notes
mmr	UN MMEIG estimates, 2019 edition. Data are based on DHS, MICS and other national household surveys	The number of maternal deaths during a given time period per 100,000 live births during the same time period.	Extrapolated to 2019 assuming a flat trend, and imputation using regional year data for countries without any available data.
nmr	UN IGME, 2019 Edition. Data are based on from vital registration, household survey and population census.	Probability of dying in the first 28 days of life, expressed per 1,000 live births.	No additional processing applied.
pab	UNICEF/WHO. Data based on administrative reporting and TT coverage surveys.	Percentage of pregnant women protected by tetanus toxoid containing vaccines (TTCV) who would give birth to a child protected against tetanus as a result of maternal transfer of antibodies through the placenta.	Extrapolated to 2019 assuming a flat trend, linear interpolation applied when data between 2000-2019 unavailable, imputation using regional year data for countries without any available data, and smoothing applied.

Continued on next page

Table A.3 – continued from previous page

Var	Source	Definition	Methodology Notes
pfpr	Malaria Atlas Project, estimates, 2019 edition. Data based on national household surveys, routine surveillance systems, and geographic and climate data	Plasmodium falciparum parasite rate.	Extrapolated to 2019 assuming a flat trend, linear interpolation applied when data between 2000-2019 unavailable, and smoothing applied.
sab	UNICEF/WHO estimates, 2019 edition. Data based on admin records, DHS, MICS and other national household surveys	The proportion of births attended by skilled health personnel.	Extrapolated to 2019 assuming a flat trend, linear interpolation applied when data between 2000-2019 unavailable, and imputation using regional year data for countries without any available data.
urban	United Nations Department of Economic and Social Affairs, Population Division, World Urbanization Prospects 2018	Percentage of population living in urban areas.	No additional processing applied.

APPENDIX B

SUPPLEMENTARY INFORMATION FOR CHAPTER 2

Numerous methods for Bayesian model selection and assessment have been proposed. Various approaches and their theoretical properties have been reviewed in prior work, for example by Bernardo and Smith [1994], Vehtari and Ojanen [2012], and Piironen and Vehtari [2017a]. Drawing from these prior studies, we will summarize some widely used Bayesian variable selection methods in section B.1. In section B.2, We will discuss the feasibility of using these methods for variable selection in a BHTRM model as given in Eq. (2.2), where the outcome of interest is the sum of a linear regression model, group-specific intercepts α_c, and temporal smoother terms $\delta_{c,t}$.

B.1 Literature review of Bayesian variable selection

Piironen and Vehtari [2017a] has categorized the model selection methods into $\mathcal{M}$-close, $\mathcal{M}$-completed, and $\mathcal{M}$-open views. There are 2^K candidate models if outcome y has K candidate covariates. $\mathcal{M}$-close view assumes one of the 2^K candidate models is true data generating model. Based on the assumption, one can set prior probabilities for each candidate model and build the Bayesian model average (BMA) solution (Barbieri and Berger [2004]). However, the $\mathcal{M}$-completed view yields the idea of true data generating model. But it assumes that there is still a reference model, M_R, which is the best available model to predict the unobserved observations based on some criteria. One can identify the relevant covariates according to the posterior distribution of the reference model based on some approaches. The approaches are

known as $\mathcal{M}$-completed approaches. The $\mathcal{M}$-open views does not depends on the assumption of true data generating model or reference model. The candidate models are compared based on criteria like WAIC, DIC, ect.

B.1.1 $\mathcal{M}$-open methods

Information criteria offers a computationally appealing way of estimating the performance of the model. Candidate models can be compared using its expected predictive accuracy on new data. There are several widely-used methods to estimate expected log posterior predictive density (ELPD). The within-sample log-posterior density is subject to biased, some information criteria like Akaike information criterion (AIC), and deviance information criterion (DIC), are proposed to correct the bias with a penalty by introducing the number of parameters.

B.1.1.1 deviance information criterion (DIC)

DIC estimates the generalization performance of the model with parameters set as $\hat{\beta}$ which maximizes the observed data likelihood $p(y|x, \beta)$.

$$\text{DIC} = \frac{1}{N} \sum \log p(y_i \mid x_i, \hat{\beta}) - \frac{P_{\text{eff}}}{N} \tag{B.1}$$

where P_{eff} is the effective number of parameters which can be estimated as

$$P_{\text{eff}} = 2 \sum_{i=1}^{N} (\log p(y_i \mid x_i, \hat{\beta}) - E[\log p(y_i \mid x_i, \beta)]) \tag{B.2}$$

The expectation is calculated based on the posterior. Thus, DIC is not theoretically justified because it measures the fit when the parameters are fixed to a point estimate, which is questionable because of Bayesian perspective.

B.1.1.2 widely applicable information criterion (WAIC) and Pareto Smoothed Importance Sampling LOO (PSIS-LOO)

Another fully Bayesian criterion is WAIC (Watanabe [2010]). WAIC can be calculated as

$$\text{WAIC} = \frac{1}{N}\sum_{i=1}^{N} \log p(y_i \mid x_i, \beta) - \frac{V}{n} \tag{B.3}$$

where V is the functional variance given by

$$\text{V} = \sum_{i=1}^{N} \{\text{E}[(\log p(y_i \mid x_i, \beta))^2] - \text{E}[\log p(y_i \mid x_i, \beta)]^2\} \tag{B.4}$$

Both of the expectation account for the uncertainty in the parameters. Watanabe [2010] proved that WAIC is asymptotically equal to the Bayesian LOO-CV. Vehtari et al. [2017] proposed the Pareto Smoothed Importance Sampling LOO implemented in the loo package. It is computationally efficient as it does not require completely re-fitting the model, unlike cross-validation. It is more robust than WAIC in the finite case with weak priors or influential observations.

B.1.2 $\mathcal{M}$-complete methods

$\mathcal{M}$-complete view is based on the reference model which is considered as the best model to predict the unobserved data. Many Bayesian variable selection approaches are designed based on the posterior distribution of the reference model. To identify the relevant covariates based on the credible interval is one of the most convenient approach. Another widely applied approach is the Spike and slab prior (Mitchell and Beauchamp [1988]) which is always considered as the gold standard for Bayesian variable selection. Piironen et al. [2020] proposed a projection method based on the reference model with horseshoe prior (Carvalho et al. [2009a]).

B.1.2.1 Credible interval approach

The simplest $\mathcal{M}$-complete methods is to select variable based on the credible intervals given by the reference model. van der Pas et al. [2017] show that the marginal credible intervals given by horseshoe prior are narrow enough to be informative for variable selection in the normal means problem:

$$y_i = \beta_i + \epsilon_i, i = 1, ..., n, \tag{B.5}$$

with ϵ i.i.d. $N(0, \sigma^2)$, and β_i is the normal mean. Bhattacharya et al. [2016] compare various aspects of the horseshoe prior to frequentist procedures in the linear regression problem, and obtain highly promising results for large signals.

B.1.2.2 Spike and slab prior

The spike and slab (Mitchell and Beauchamp [1988],George and McCulloch [1993]) is a popular shrinkage prior for sparse Bayesian estimation. The prior is often written as a two-component mixture of normal distributions

$$\beta_k \mid \lambda_k, c, \sigma \sim \lambda_k \mathrm{N}(0, c^2) + (1 - \lambda_k)\mathrm{N}(0, \sigma^2), \tag{B.6}$$

$$\lambda_k \sim \mathrm{Bern}(\pi) \tag{B.7}$$

where σ is much smaller than c and often set as 0 so that $N(0, \sigma^2)$ is the 'spike' and $N(0, c^2)$ is the 'slab'. And $\lambda_k \in \{0, 1\}$ is the indicator variable denotes whether the coefficient β_k is zero when $\lambda_k = 1$ or nonzero when $\lambda_k = 0$. The final decision of covariate x_k can be informed by the posterior distribution $P(\lambda_k \mid y, x)$. If $P(\lambda_k = 1 \mid y, x) > 0.5$, we drop kth covariate.

Ročková and George [2018] introduced the spike and slab LASSO (SSL). It is an approach based on a prior which provides a continuum between the penalized

likelihood LASSO and the Bayesian spike and slab prior. The spike and slab LASSO prior is usually specified as

$$\beta_k \mid \lambda_k, \tau_0, \tau_1 \sim \lambda_k \phi(\beta_k \mid \tau_0) + (1 - \lambda_k)\phi(\beta_k \mid \tau_1), \tag{B.8}$$

$$\lambda_k \sim \text{Bern}(\pi) \tag{B.9}$$

where $\phi(\beta \mid \tau) = (\tau/2)e^{-\tau|\beta|}$ denotes the Laplace density with scale parameter τ. It is mixture of two Laplace priors $\phi(\beta \mid \tau)$ with different scale parameter τ_0 and τ_1. When $\tau_1 >> \tau_0$, the SSL has the similar structure as spike and slab prior, because the density is very peaked around zero when τ is large. Since it is a mixture of two Laplace distributions, the SSL prior is a two-group refinement of LASSO with different penalties on the coefficients. Thus, the posterior distribution $p(\beta_k \mid x, y)$ can be used to perform variable selection, which is the same as basic spike and slab prior.

B.1.2.3 Projection approach

Another $\mathcal{M}$-complete method is projection method which is proposed by Goutis and Robert [1998], and further discussed by Dupuis and Robert [2003b]. The key characteristic of this approach is to find an excellent trade-off between sparsity and predictive accuracy. It is to simplify the full model M^* by projecting the information in the posterior onto the candidate submodel so that the predictive distribution is as close to the reference model as possible. Given the parameters of the full model β^*, the projected parameters β^p in the parameter space of submodel M^p are defined as

$$\beta^p = \underset{\beta}{\text{argmin}} \frac{1}{N} \sum_{i=1}^{N} KL(p(\tilde{y} \mid x_i, \beta^*, M^*) \mid\mid p(\tilde{y} \mid x_i, \beta, M^p)) \tag{B.10}$$

where KL refers to the Kullback-Leibler Divergence, and $\tilde{y}$ refers to the unobserved outcome.The model choice can be based on the strict minimization of the discrepancy measure.

B.1.3 $\mathcal{M}$-close methods

Bayesian Model Average (BMA) (Martini and Spezzaferri [1984]) is one of the $\mathcal{M}$-close view variable selection methods. Suppose there is an exhaustive list of candidate models, $\{M_k\}_{k=1}^K$ the distribution over the model space is,

$$p(M|D) \propto p(D|M)p(M) \tag{B.11}$$

The predictions from Bayesian Model Averaging(BMA) are

$$p(\tilde{y}|D) = \sum_{k=1}^{K} p(\tilde{y}|D, M_k)p(M_k|D) \tag{B.12}$$

In BMA each model is weighted by its marginal likelihood,

$$p(M_k|y) = \frac{p(y|M_k)p(M_k)}{\sum_{k=1}^{K} p(y|M_k)p(M_k)} \tag{B.13}$$

where

$$p(y|M) = \int p(y|\Gamma_k, M_k)p(\Gamma_k|M_k)d\Gamma_k, \tag{B.14}$$

and $\Gamma_k = \{\beta_k\}$ in linear regression setting. In the $\mathcal{M}$-close view, BMA will asymptotically select the correct model. Since the BMA weights by marginal likelihood, these weights extremely sensitive to the choices of the priors $p(\Gamma_k)$ for each model. The sensitivity to prior distributions make the BMA weights suspect. The difficulty of computing marginal likelihood generally make the BMA hard to generalize (Piironen and Vehtari [2017a]).

B.2 Application of variable selection approach in BHTRM

BHTRM used for SBR data combines a linear regression and temporal smooth term, which can be considered as a semi-parametric model. Due to the particularity

of BHTRM, most of variable selection methods are not available to impose variable selection for BHTRM. In this section, we will discuss the availability of above approaches.

B.2.1 $\mathcal{M}$-open methods

The $\mathcal{M}$-open view abandons the idea of reference model and true data generating model. The variable selection is usually conducted by model comparison between a few number of candidate models based on information criteria. Thus, this type of approach can be used for any model including BHTRM. However, if we do not have any preference or knowledge for the covariates based on prior information, we have to combine the approach with stepwise model searching strategies. It is a robust and convenient approach for small models, but it may not be a good choice for BHTRM, which is computational expensive.

We use an iterative forward method to test the performance of PSIS-LOO in the BHTRM setting:

1. Start from the empty model and compute ELPD loo.
2. Add the covariate which best improves ELPD loo.
3. Repeat step 2 until no improvement of ELPD is observed.

B.2.2 $\mathcal{M}$-complete methods

$\mathcal{M}$-complete methods assume that there is a reference model which can predict the unobserved data well. Variable selection can be informed by the reference model. But how to construct the reliable reference model is an open question. Because the prior of SSL is a mixture of two Laplace distribution, the posterior distribution $p(\beta \mid x, y)$ is exactly sparse and can be used to perform parameter estimation. Bai et al. [2020] shows that the model with SSL has good predictive performance. As a continuous version of spike and slab prior, Piironen et al. [2020] suggest that model

with horseshoe prior also has good performance in prediction. Thus, the models with SSL prior and horseshoe prior can be considered as the reliable reference model to impose variable selection.

B.2.2.1 Credible interval

Bhattacharya et al. [2016] compare the variable selection accuracy between the use of credible interval based on horseshoe prior with frequentist procedures in the linear regression problem, and obtain highly promising results for large signals. However, almost none of the small and medium signals are detected. Given that all signals are relatively small compare to the outcome variance in SBR data. It could be the first challenge to use credible interval to selection variable in SBR data. Using credible interval requires a threshold value α. The choice of α are always arbitrary. It could be another issue of credible interval approach. But it is still an available and most straightforward approach for BHTRM.

B.2.2.2 Spike and slab and SSL

Although spike and slab type approach is one of the most popular approach and it performs well in generalized linear regression setting, it cannot make use of the advantage of STAN (Stan Development Team [2018],Carpenter et al. [2017]) which does not allow the discrete parameter: the local binary indicator λ_k in Eq. (B.6). But we will use the continuous version spike and slab - the horseshoe prior - instead.

B.2.2.3 Projection method

Although projection approach has been demonstrated for generalized linear multilevel models and generalized additive multilevel models, the proposed approach is still not working for BHTRM setting. The full model M^* not only contains linear regression $\sum_k x_{ik}\beta_k$, but also a country-specific intercept α_c, a smoothing term $\delta_{c,t}$, and non-homogeneous variance σ_d.

To apply for the projection approach, we develop a projection approach for BHTRM version. We propose a two-step projection approach. First, we project the full model M^* onto a submodel M^b by constraining the smoothing term to 0, i.e., $\delta^b_{c,t} = 0$. Then the projected parameters β^p in the parameter space of sub-submodel M^p can be defined as

$$\beta^p = \underset{\beta}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} KL(p(\tilde{y}^b \mid x_i, \beta^*, \alpha, \sigma, \delta^b, M^b) \mid\mid p(\tilde{y}^b \mid x_i, \beta, \alpha, \sigma, \delta^b, M^p)) \quad \text{(B.15)}$$

where $\tilde{y}^c$ refers to the data without smoothing term.

The discrepancy between the first projection submodel, M^b, and the submodel of the projection submodel, M^p, is then defined to be the expectation of this divergence over the posterior of the first order projection model. Dupuis and Robert [2003b] introduce the notation of explanatory power to measure the distance between models. In BHTSRM setting, we use the mean square error to obtain the bast sub-submodel.

We calculate the discrepancy by samples $\{\beta^*, \alpha, \sigma\}_{s=1}^{S}$ from the posterior of the reference model, calculating the projected parameters $\{\beta^p\}_{s=1}^{S}$, and then calculate the discrepancy as

$$\Delta(M^b||M^p) = \frac{1}{n} \sum_{i=1}^{n} (\tilde{y}_i^b - \tilde{y}_i^p)^2 \quad \text{(B.16)}$$

where $\tilde{y}_i^b = \frac{1}{S} \sum \tilde{y}_i^{b(s)}$, and $\tilde{y}_i^p = \frac{1}{S} \sum \tilde{y}_i^{p(s)}$,

$$\tilde{y}_i^{b(s)} \sim N(\alpha_{c[i]}^{(s)} + \sum_k x_{ik} \beta_k^{*(s)}, \sigma_{d[i]}^{2(s)}) \quad \text{(B.17)}$$

$$\tilde{y}_i^{p(s)} \sim N(\alpha_{c[i]}^{(s)} + \sum_k x_{ik} \beta_k^{p(s)}, \sigma_{d[i]}^{2(s)}) \quad \text{(B.18)}$$

After fitting the model with all the variables using regularized horseshoe prior, we use the projection predictive variable selection strategy above. As a search heuristic, we use forward searching, that is, starting from the empty model, we add variables one at a time, each time choosing the variable that decreasing the mean square error the most.

B.2.3 $\mathcal{M}$-close approach: BMA

BMA has been successfully implemented in generalized linear regression, but the BHTSRM makes the BMA approach more complicated, unstable, and time consuming because of the difficulty of computation of the likelihood and sampling issues. BMA is a undesired approach for BHTRM, we do not test the performance in this paper.

APPENDIX C

SUPPLEMENTARY TABLE AND FIGURES FOR CHAPTER 3

C.1 Notation and Definitions

Notation	Description
t^*	End year of the reference (study) period 2013.
$t_0 = t^* - 15$	Start year of the reference (study) period 2013.
$b_i^{(\text{ssh})}$	SSH reported birth year for sibling i.
$d_i^{(\text{ssh})}$	SSH reported time since death for deceased sibling i.
$s_i^{(\text{ssh})}$	SSH reported survival status at end of study period for sibling i.
b_i	HDSS reported birth year for sibling i.
d_i	HDSS reported time since death for deceased sibling i.
s_i	HDSS reported survival status at end of study period for sibling i.
$e_i^{(\text{birth})}$	Birth year error for sibling i given by $b_i^{(\text{ssh})} - b_i$.
$e_i^{(\text{death})}$	Time since death error for sibling i given by $d_i^{(\text{ssh})} - d_i$.
(λ_g, η_g)	Interviewer random effects.
(β_k, γ_k)	Covariate effects.
$(\sigma_b^2, \sigma_d^2, \delta)$	(Co-) variances related to errors.
(τ, φ)	Global shrinkage parameters in horseshoe priors.
B_i	The true (unknown) birth year for sibling i.
D_i	The true (unknown) time since death for sibling i.
T_i	The true (unknown) time to death since t_0 for sibling i.
b	Continuous cohort(birth year) b.
$S(b,t)$	Survival function for cohort b at time t.
$f(b,t)$	Time to death density function for cohort b at time t.
$H(b,t)$	Cumulative hazard function for cohort b at time t.
$\alpha_{l,p}$	two dimensional spline coefficient.
(χ,ϕ)	Parameters of CAR.

Table C.1. Descriptions of mathematical notations

C.1.1 Graphical analysis of covariates

Figure C.2 illustrates the relationship between observed reporting errors and the covariates.

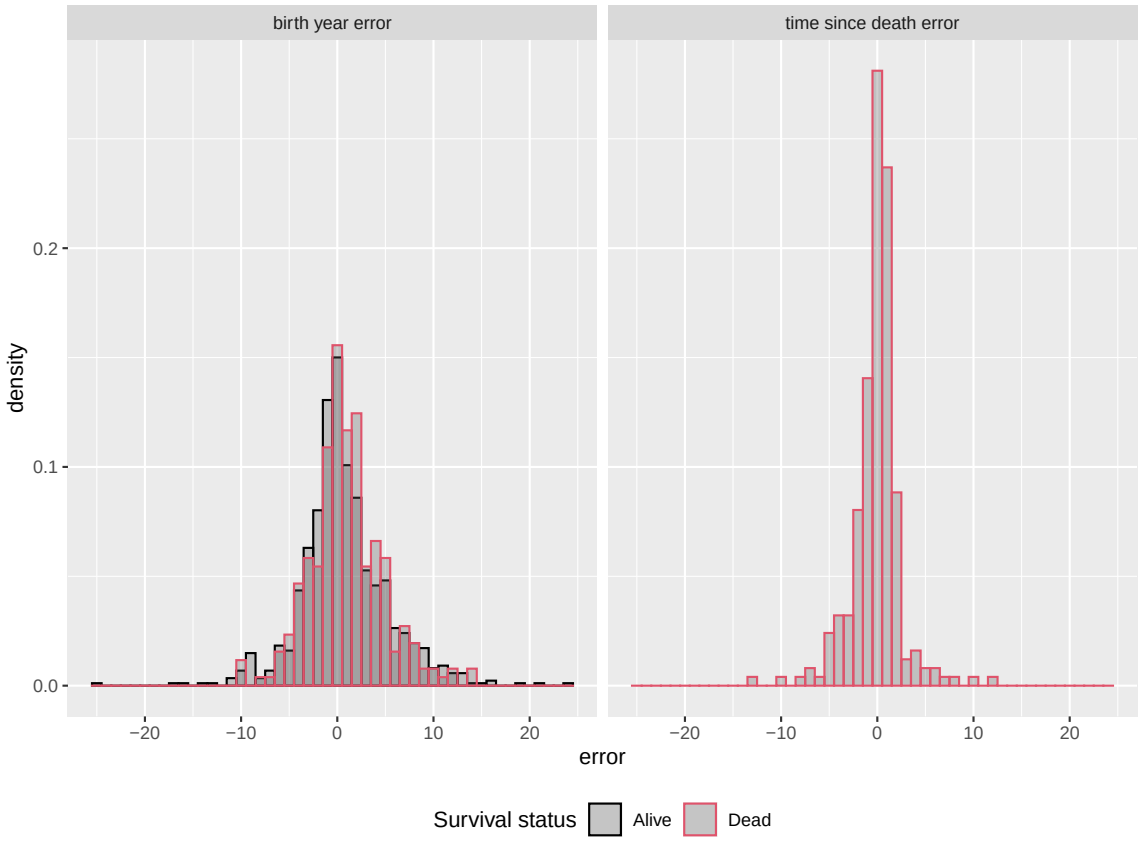


Figure C.1. Exploratory data analysis. Distribution of birth year error and time since death error.

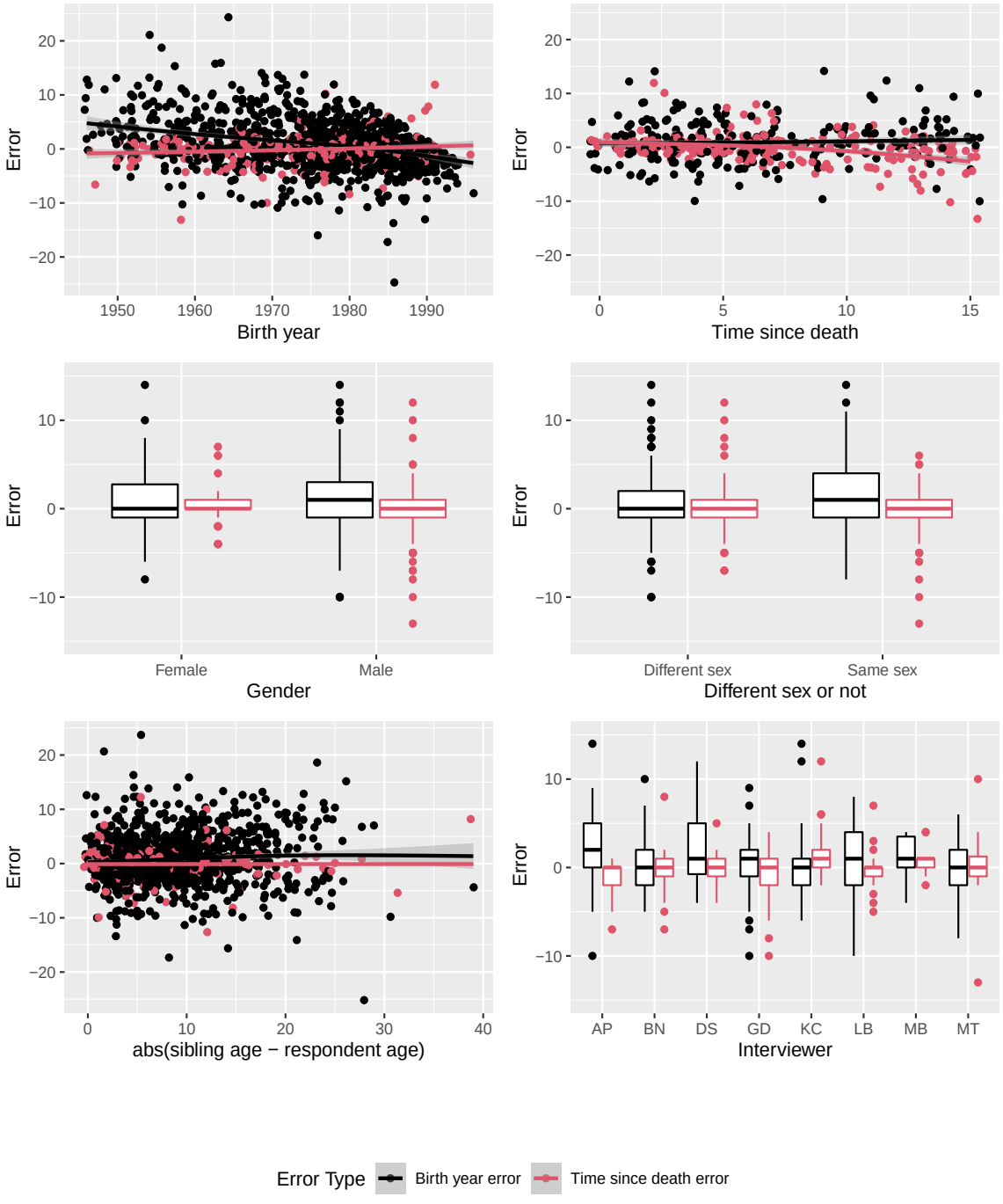


Figure C.2. Exploratory data analysis. Relationship between errors and covariates.

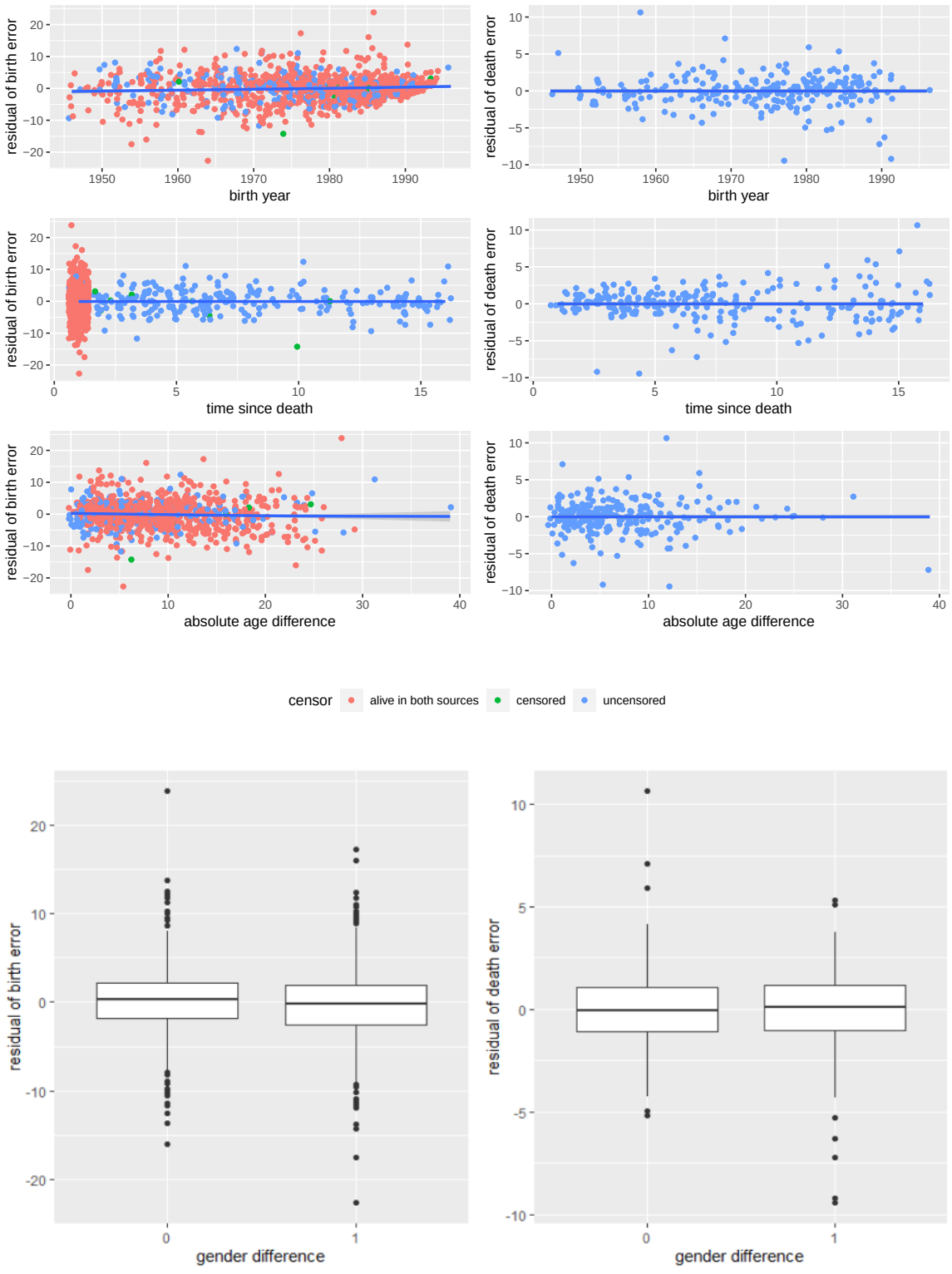


Figure C.3. Residual plot from SSH error model.

BIBLIOGRAPHY

Imran Ahmed, Said Mohammed Ali, Seeba Amenga-Etego, Shabina Ariff, Rajiv Bahl, Abdullah H Baqui, Nazma Begum, Nita Bhandari, Kiran Bhatia, Zulfiqar A Bhutta, Godfrey Biemba, Saikat Deb, Usha Dhingra, Brinda Dube, Arup Dutta, Karen Edmond, Fabian Esamai, Wafaie Fawzi, Amit Kumar Ghosh, Peter Gisore, Caroline Grogan, Davidson H Hamer, Julie Herlihy, Lisa Hurt, Muhammad Ilyas, Fyezah Jehan, Michel Kalonji, Jasmine Kaur, Rasheda Khanam, Betty Kirkwood, Aarti Kumar, Alok Kumar, Vishwajeet Kumar, Alexander Manu, Irene Marete, Honorati Masanja, Sarmila Mazumder, Usma Mehmood, Shambhavi Mishra, Dipak K Mitra, Erick Mlay, Sanjana Brahmawar Mohan, Mamun Ibne Moin, Karim Muhammad, Alfa Muhihi, Samuel Newton, Serge Ngaima, Andre Nguwo, Imran Nisar, Maureen O'Leary, John Otomba, Pawankumar Patil, Mohammad Abdul Quaiyum, Mohammed Hefzur Rahman, Sunil Sazawal, Katherine EA Semrau, Caitlin Shannon, Emily R Smith, Sajid Soofi, Seyi Soremekun, Venantius Sunday, Sunita Taneja, Antoinette Tshefu, Yaqub Wasan, Kojo Yeboah-Antwi, Sachiyo Yoshida, and Anita Zaidi. Population-based rates, timing, and causes of maternal deaths, stillbirths, and neonatal deaths in south Asia and sub-Saharan Africa: a multi-country prospective cohort study. *The Lancet Global Health*, 6(12):e1297–e1308, December 2018. ISSN 2214109X. doi: 10.1016/S2214-109X(18)30385-1. URL `https://linkinghub.elsevier.com/retrieve/pii/S2214109X18303851`.

Leontine Alkema, Sanqian Zhang, Doris Chou, Alison Gemmill, Ann-Beth Moller, Doris Ma Fat, Lale Say, Colin Mathers, and Daniel Hogan. A Bayesian approach to the global estimation of maternal mortality. *The Annals of Applied Statistics*, 11(3):1245–1274, September 2017. ISSN 1932-6157. doi: 10.1214/16-AOAS1014. URL `https://projecteuclid.org/euclid.aoas/1507168829`.

Ray Bai, Veronika Rockova, and Edward I George. Spike-and-slab meets lasso: a review of the spike-and-slab lasso. *arXiv preprint arXiv:2010.06451*, 2020.

Maria Maddalena Barbieri and James O. Berger. Optimal predictive model selection. *The Annals of Statistics*, 32(3), June 2004. ISSN 0090-5364. doi: 10.1214/009053604000000238. URL `http://arxiv.org/abs/math/0406464`. arXiv: math/0406464.

Jos M. Bernardo and Adrian F. M. Smith, editors. *Bayesian Theory*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, USA, May 1994. ISBN 978-0-470-31687-0 978-0-471-92416-6. doi: 10.1002/9780470316870. URL `http://doi.wiley.com/10.1002/9780470316870`.

Anirban Bhattacharya, Antik Chakraborty, and Bani K. Mallick. Fast sampling with Gaussian scale mixture priors in high-dimensional regression. *Biometrika*, 103(4):985–991, December 2016. ISSN 0006-3444, 1464-3510. doi: 10.1093/biomet/asw042. URL https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/asw042.

Hannah Blencowe, Simon Cousens, Fiorella Bianchi Jassir, Lale Say, Doris Chou, Colin Mathers, Dan Hogan, Suhail Shiekh, Zeshan U Qureshi, Danzhen You, and Joy E Lawn. National, regional, and worldwide estimates of stillbirth rates in 2015, with trends from 2000: a systematic analysis. *The Lancet Global Health*, 4 (2):e98–e108, 2016. ISSN 2214109X. doi: 10.1016/S2214-109X(15)00275-2. URL https://linkinghub.elsevier.com/retrieve/pii/S2214109X15002752.

Carl L Bose, Melissa Bauserman, Robert L Goldenberg, Shivaprasad S Goudar, Elizabeth M McClure, Omrana Pasha, Waldemar A Carlo, Ana Garces, Janet L Moore, Menachem Miodovnik, and Marion Koso-Thomas. The Global Network Maternal Newborn Health Registry: a multi-national, community-based registry of pregnancy outcomes. *Reproductive Health*, 12(S2):S1, December 2015. ISSN 1742-4755. doi: 10.1186/1742-4755-12-S2-S1. URL https://reproductive-health-journal.biomedcentral.com/articles/10.1186/1742-4755

Sarah E.K. Bradley, William Winfrey, and Trevor N. Croft. Contraceptive use and perinatal mortality in the dhs: An assessment of the quality and consistency of calendars and histories. 2015. DHS Methodological Reports No. 17. Rockville, Maryland, USA: ICF International.

Bob Carpenter, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. *Stan* : A Probabilistic Programming Language. *Journal of Statistical Software*, 76(1), 2017. ISSN 1548-7660. doi: 10.18637/jss.v076.i01. URL http://www.jstatsoft.org/v76/i01/.

Carlos M Carvalho, Nicholas G Polson, and James G Scott. Handling sparsity via the horseshoe. In *Artificial Intelligence and Statistics*, pages 73–80. PMLR, 2009a.

Carlos M Carvalho, Nicholas G Polson, and James G Scott. Handling Sparsity via the Horseshoe. *Artificial Intelligence and Statistics*, page 8, 2009b.

Carlos M. Carvalho, Nicholas G. Polson, and James G. Scott. Handling sparsity via the horseshoe. 5:73–80, 16–18 Apr 2009c. URL https://proceedings.mlr.press/v5/carvalho09a.html.

I D Currie and M Durban. Flexible smoothing with P-splines: a unified approach. *Statistical Modelling: An International Journal*, 2(4):333–349, December 2002. ISSN 1471-082X, 1477-0342. doi: 10.1191/1471082x02st039ob. URL http://journals.sagepub.com/doi/10.1191/1471082x02st039ob.

Victor De Oliveira. Bayesian analysis of conditional autoregressive models. *Annals of the Institute of Statistical Mathematics*, 64(1):107–133, 2012.

Valerie Delaunay, Laetitia Douillot, Aldiouma Diallo, Djibril Dione, Jean-François Trape, Oleg Medianikov, Didier Raoult, and Cheikh Sokhna. Profile: the niakhar health and demographic surveillance system. *International journal of epidemiology*, 42(4):1002–1011, 2013.

Paul Demeny, Geoffrey McNicoll, and Dennis Hodgson. *Encyclopedia of population*. Hodgson, Dennis (2003). Contemporary Population Thought, 2003.

Jérome A. Dupuis and Christian P. Robert. Variable selection in qualitative models via an entropic explanatory power. *Journal of Statistical Planning and Inference*, 111(1-2):77–94, February 2003a. ISSN 03783758. doi: 10.1016/S0378-3758(02)00286-0. URL https://linkinghub.elsevier.com/retrieve/pii/S0378375802002860.

Jérome A. Dupuis and Christian P. Robert. Variable selection in qualitative models via an entropic explanatory power. *Journal of Statistical Planning and Inference*, 111(1-2):77–94, February 2003b. ISSN 03783758. doi: 10.1016/S0378-3758(02)00286-0. URL https://linkinghub.elsevier.com/retrieve/pii/S0378375802002860.

Paul H. C. Eilers and Brian D. Marx. Flexible smoothing with B -splines and penalties. *Statistical Science*, 11(2):89–121, May 1996. ISSN 0883-4237. doi: 10.1214/ss/1038425655. URL http://projecteuclid.org/euclid.ss/1038425655.

Andrew Gelman. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1(3): 515–534, September 2006. ISSN 1936-0975. doi: 10.1214/06-BA117A. URL http://projecteuclid.org/euclid.ba/1340371048.

Andrew Gelman and B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–72, 1992.

Edward I George and Robert E McCulloch. Variable selection via gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889, 1993.

Constantinos. Goutis and Christian P. Robert. Model choice in generalised linear models: a Bayesian approach via Kullback-Leibler projections. *Biometrika*, 85(1): 29–37, March 1998. ISSN 0006-3444, 1464-3510. doi: 10.1093/biomet/85.1.29. URL https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/85.1.29.

Wendy Graham, William Brass, and Robert W. Snow. Estimating Maternal Mortality: The Sisterhood Method. *Studies in Family Planning*, 20(3):125–135, 1989. ISSN 0039-3665. doi: 10.2307/1966567. URL https://www.jstor.org/stable/1966567. Publisher: [Population Council, Wiley].

Stéphane Helleringer, Gilles Pison, Almamy M. Kanté, Géraldine Duthé, and Armelle Andro. Reporting Errors in Siblings' Survival Histories and Their Impact on Adult Mortality Estimates: Results From a Record Linkage Study in Senegal. *Demography*, 51(2):387–411, April 2014a. ISSN 0070-3370. doi: 10.1007/s13524-013-0268-3. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4487521/.

Stéphane Helleringer, Gilles Pison, Bruno Masquelier, Almamy Malick Kanté, Laetitia Douillot, Géraldine Duthé, Cheikh Sokhna, and Valérie Delaunay. Improving the Quality of Adult Mortality Data Collected in Demographic Surveys: Validation Study of a New Siblings' Survival Questionnaire in Niakhar, Senegal. *PLOS Medicine*, 11(5):e1001652, May 2014b. ISSN 1549-1676. doi: 10.1371/journal.pmed.1001652. URL https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001652. Publisher: Public Library of Science.

Lucia Hug, Danzhen You, Hannah Blencowe, Anu Mishra, Zhengfan Wang, Miranda J Fix, Jon Wakefield, Allisyn C Moran, Victor Gaigbe-Togbe, Emi Suzuki, et al. Global, regional, and national estimates and trends in stillbirths from 2000 to 2019: a systematic assessment. *The Lancet*, 398(10302):772–785, 2021.

Shyama Kuruvilla, Flavia Bustreo, Taona Kuo, Ck Mishra, Katie Taylor, Helga Fogstad, Geeta Rao Gupta, Kate Gilmore, Marleen Temmerman, Joe Thomas, Kumanan Rasanathan, Ted Chaiban, Anshu Mohan, Anna Gruending, Julian Schweitzer, Hannah Sarah Dini, John Borrazzo, Hareya Fassil, Lars Gronseth, Rajat Khosla, Richard Cheeseman, Robin Gorna, Lori McDougall, Kadidiatou Toure, Kate Rogers, Kate Dodson, Anita Sharma, Marta Seoane, and Anthony Costello. The *Global strategy for women's, children's and adolescents' health (2016–2030)* : a roadmap based on evidence and country experience. *Bulletin of the World Health Organization*, 94(5): 398–400, May 2016. ISSN 0042-9686. doi: 10.2471/BLT.16.170431. URL http://www.who.int/entity/bulletin/volumes/94/5/16-170431.pdf.

Li Liu, Henry D. Kalter, Yue Chu, Narjis Kazmi, Alain K. Koffi, Agbessi Amouzou, Olga Joos, Melinda Munos, and Robert E. Black. Understanding Misclassification between Neonatal Deaths and Stillbirths: Empirical Evidence from Malawi. *PLOS ONE*, 11(12):e0168743, December 2016. ISSN 1932-6203. doi: 10.1371/journal.pone.0168743. URL https://dx.plos.org/10.1371/journal.pone.0168743.

A San Martini and Fulvio Spezzaferri. A predictive model selection criterion. *Journal of the Royal Statistical Society: Series B (Methodological)*, 46(2):296–303, 1984.

Bruno Masquelier. Adult Mortality From Sibling Survival Data: A Reappraisal of Selection Biases. *Demography*, 50(1):207–228, October 2012. ISSN 0070-3370. doi: 10.1007/s13524-012-0149-1. URL https://doi.org/10.1007/s13524-012-0149-1.

Bruno Masquelier, Mufaro Kanyangarara, Gilles Pison, Almamy Malick Kanté, Cheikh Tidiane Ndiaye, Laetitia Douillot, Géraldine Duthé, Cheikh Sokhna, Valérie Delaunay, and Stéphane Helleringer. Errors in reported ages and dates in surveys of adult mortality: A record linkage study in Niakhar (Senegal). *Population Studies*, 75(2):269–287, May 2021. ISSN 0032-4728. doi: 10.1080/00324728.2020.1854332. URL https://doi.org/10.1080/00324728.2020.1854332. Publisher: Routledge _eprint: https://doi.org/10.1080/00324728.2020.1854332.

Lene Mikkelsen, David E Phillips, Carla AbouZahr, Philip W Setel, Don de Savigny, Rafael Lozano, and Alan D Lopez. A global assessment of civil registration and vital statistics systems: monitoring data quality and progress. *The Lancet*, 386(10001): 1395–1406, October 2015. ISSN 01406736. doi: 10.1016/S0140-6736(15)60171-4. URL https://linkinghub.elsevier.com/retrieve/pii/S0140673615601714.

Toby J Mitchell and John J Beauchamp. Bayesian variable selection in linear regression. *Journal of the american statistical association*, 83(404):1023–1032, 1988.

Mitzi Morris, Katherine Wheeler-Martin, Dan Simpson, Stephen J Mooney, Andrew Gelman, and Charles DiMaggio. Bayesian hierarchical spatial models: Implementing the besag york mollié model in stan. *Spatial and spatio-temporal epidemiology*, 31:100301, 2019.

Ziad Obermeyer, Julie Knoll Rajaratnam, Chang H. Park, Emmanuela Gakidou, Margaret C. Hogan, Alan D. Lopez, and Christopher J. L. Murray. Measuring Adult Mortality Using Sibling Survival: A New Analytical Method and New Results for 44 Countries, 1974–2006. *PLOS Medicine*, 7(4):e1000260, April 2010. ISSN 1549-1676. doi: 10.1371/journal.pmed.1000260. URL https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1000260. Publisher: Public Library of Science.

Jon Pedersen and Jing Liu. Child Mortality Estimation: Appropriate Time Periods for Child Mortality Estimates from Full Birth Histories. *PLOS Medicine*, 9(8): e1001289, August 2012. ISSN 1549-1676. doi: 10.1371/journal.pmed.1001289. URL https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001289. Publisher: Public Library of Science.

Juho Piironen and Aki Vehtari. Comparison of Bayesian predictive methods for model selection. *Statistics and Computing*, 27(3):711–735, May 2017a. ISSN 0960-3174, 1573-1375. doi: 10.1007/s11222-016-9649-y. URL http://arxiv.org/abs/1503.08650. arXiv: 1503.08650.

Juho Piironen and Aki Vehtari. Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics*, 11 (2), January 2017b. ISSN 1935-7524. doi: 10.1214/17-EJS1337SI. URL http://arxiv.org/abs/1707.01694. arXiv: 1707.01694.

Juho Piironen, Markus Paasiniemi, and Aki Vehtari. Projective Inference in High-dimensional Problems: Prediction and Feature Selection. *Electronic Journal of Statistics*, 14(1), January 2020. ISSN 1935-7524. doi: 10.1214/20-EJS1711. URL http://arxiv.org/abs/1810.02406. arXiv: 1810.02406.

Gilles Pison, Laetitia Douillot, Almamy M Kante, Ousmane Ndiaye, Papa N Diouf, Paul Senghor, Cheikh Sokhna, and Valérie Delaunay. Health & Demographic Surveillance System Profile: Bandafassi Health and Demographic Surveillance System (Bandafassi HDSS), Senegal. *International Journal of Epidemiology*, 43(3):739–748, June 2014. ISSN 0300-5771. doi: 10.1093/ije/dyu086. URL https://doi.org/10.1093/ije/dyu086.

Veronika Ročková and Edward I George. The spike-and-slab lasso. *Journal of the American Statistical Association*, 113(521):431–444, 2018.

Stan Development Team. RStan: the R interface to Stan, 2018. URL http://mc-stan.org/. R package version 2.18.2.

Cynthia Stanton, Joy E Lawn, Hafiz Rahman, Katarzyna Wilczynska-Ketende, and Kenneth Hill. Stillbirth rates: delivering estimates in 190 countries. *The Lancet*, 367 (9521):1487–1494, May 2006. ISSN 01406736. doi: 10.1016/S0140-6736(06)68586-3. URL https://linkinghub.elsevier.com/retrieve/pii/S0140673606685863.

UN Inter-agency Group for Child Mortality Estimation. *A neglected tragedy: the global burden of stillbirths. Report of the UN Inter-agency Group for Child Mortality Estimation, 2020.* New York. United Nations Children's Fund, 2020, 2020. ISBN 978-92-806-5141-6. URL https://data.unicef.org/wp-content/uploads/2020/10/UN-IGME-2020-Stillbirth-Repor

Stéphanie van der Pas, Botond Szabó, and Aad van der Vaart. Uncertainty quantification for the horseshoe (with discussion). *Bayesian Analysis*, 12(4):1221–1274, 2017.

Aki Vehtari and Janne Ojanen. A survey of Bayesian predictive methods for model assessment, selection and comparison. *Statistics Surveys*, 6(none), January 2012. ISSN 1935-7516. doi: 10.1214/12-SS102.

Aki Vehtari, Andrew Gelman, and Jonah Gabry. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5): 1413–1432, September 2017. ISSN 0960-3174, 1573-1375. doi: 10.1007/s11222-016-9696-4. URL http://arxiv.org/abs/1507.04544. arXiv: 1507.04544.

Aki Vehtari, Jonah Gabry, Mans Magnusson, Yuling Yao, and Andrew Gelman. loo: Efficient leave-one-out cross-validation and waic for bayesian models, 2019. URL https://mc-stan.org/loo. R package version 2.2.0.

Aki Vehtari, Jonah Gabry, Mans Magnusson, Yuling Yao, Paul-Christian Bürkner, Topi Paananen, and Andrew Gelman. loo: Efficient leave-one-out cross-validation and waic for bayesian models, 2020a. URL https://mc-stan.org/loo/. R package version 2.4.1.

Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. Rank-Normalization, Folding, and Localization: An Improved R hat for Assessing Convergence of MCMC. *Bayesian Analysis*, July 2020b. ISSN 1936-0975. doi: 10.1214/20-BA1221. URL https://projecteuclid.org/euclid.ba/1593828229.

Zhengfan Wang, Miranda J. Fix, Lucia Hug, Anu Mishra, Danzhen You, Hannah Blencowe, Jon Wakefield, and Leontine Alkema. Estimating the Stillbirth Rate for 195 Countries Using A Bayesian Sparse Regression Model with Temporal Smoothing. *arXiv:2010.03551 [stat]*, October 2020. URL http://arxiv.org/abs/2010.03551. arXiv: 2010.03551.

Sumio Watanabe. Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. page 24, 2010.

WHO Department of Information, Evidence, and Research. *WHO methods and data sources for life tables 1990–2016*. 2018. URL https://www.who.int/healthinfo/statistics/LT$_m$*ethod.pdf*.

Robert Woods. Long-term tends in fetal mortality: implications for developing countries. *Bulletin of the World Health Organization*, 2008(6):460–466, June 2008. ISSN 00429686. doi: 10.2471/BLT.07.043471. URL http://www.who.int/bulletin/volumes/86/6/07-043471.pdf.

World Health Organization. *Every newborn: an action plan to end preventable deaths*. World Health Organization, Geneva, 2014. ISBN 978-92-4-150744-8. OCLC: 918973731.

World Health Organization. *International classification of Disease (ICD), revised definition for ICD-11*. 2019.