


Simultaneous Estimation and Model Choice for Big Discrete Time-to-Event Data with Additive Predictors

Benjamin Müller 
University of Innsbruck

Nikolaus Umlauf 
University of Innsbruck

Johannes Seiler 
University of Innsbruck

Kenneth Harttgen 
ETH Zurich

Stefan Lang 
University of Innsbruck

Abstract

Discrete-time hazard models are widely used when event times are measured in intervals or are not precisely observed. While these models can be estimated using standard generalized linear model techniques, they rely on extensive data augmentation, making estimation computationally demanding in high-dimensional settings. In this paper, we demonstrate how the recently proposed Batchwise Backfitting algorithm, a general framework for scalable estimation and variable selection in distributional regression, can be effectively extended to discrete hazard models. Using both simulated data and a large-scale application on infant mortality in sub-Saharan Africa, we show that the algorithm delivers accurate estimates, automatically selects relevant predictors, and scales efficiently to large data sets. The findings underscore the algorithm’s practical utility for analysing large-scale, complex survival data with high-dimensional covariates.

Keywords: Discrete time-to-event analysis; Additive predictors; Big data; Batchwise backfitting; Automated variable selection.

1. Introduction

Modelling the time until an event occurs is a central task in many areas of applied statistics. This field, commonly referred to as time-to-event or survival analysis, has a wide range of applications, from modelling survival outcomes of children to predicting the duration of marriages or the failure time of mechanical components (e.g., [Burstein *et al.* 2019](#); [Musick and Micheltore 2015](#); [Christodoulou 2011](#)).

While much of the research in this area focuses on continuous-time models, discrete-time approaches are often more suitable when event times are measured in intervals, as is often the case in social science and health research (e.g., age of the children in months in child survival analyses). Foundational work by [Allison \(1982\)](#) established the basis for discrete-time survival models, and subsequent developments have introduced extensions to enhance their flexibility ([Fahrmeir and Wagenpfeil 1996](#)).

In parallel, the development of generalized additive models (GAMs) has advanced statistical modelling by allowing for nonlinear, smooth effects of covariates ([Hastie and Tibshirani 1986](#)). Building on this, innovations in smoothing techniques ([Eilers and Marx 1996](#)), efficient esti-

mation algorithms (Wood 2003, 2004; Wood *et al.* 2017), and Bayesian extensions (Fahrmeir and Lang 2001; Brezger and Lang 2006) have made GAMs highly adaptable.

These innovations have also influenced time-to-event analysis, leading to growing interest in flexible additive models for survival data (Tutz and Binder 2004; Tutz and Schmid 2016; Berger and Schmid 2018). More recently, machine learning-inspired techniques based on tree methods have grown in popularity in this field (Schmid *et al.* 2016; Puth *et al.* 2020; Spuck *et al.* 2023).

Modern applications of time-to-event analysis increasingly involve large-scale and high-dimensional data sets, which present significant computational and statistical challenges (e.g., complex nonlinear effects and automated variable selection). In such contexts, efficient estimation methods and automatic variable selection are critical. However, to the best of our knowledge, scalable and interpretable approaches for additive discrete time-to-event models remain underdeveloped and insufficiently tested.

This paper addresses this gap by extending the recently proposed Batchwise Backfitting algorithm by Umlauf *et al.* (2024) to additive discrete time-to-event models. Our contribution lies in introducing a scalable estimation strategy that enables simultaneous model fitting and variable selection for big data sets. We evaluate the performance of the Batchwise Backfitting algorithm through a comprehensive simulation study and a real-world application.

The simulation study applies the method across a range of settings, including data sets with up to ten million observations and numerous informative and non-informative covariates. The results demonstrate that the proposed approach achieves comparable or superior performance to established alternatives in terms of estimation accuracy, variable selection, and estimation time. This suggests that the method is well-suited for large-scale, high-dimensional time-to-event data, providing researchers with a practical tool for tackling complex survival analyses that are challenging to address with existing approaches.

To demonstrate the practical utility of the method, we apply Batchwise Backfitting to model infant mortality in ten sub-Saharan African countries. The data set includes approximately 350000 individual children and 26 potential explanatory variables, representing a complex, high-dimensional setting. The results highlight the algorithm’s ability to identify key determinants of child survival while maintaining scalability and interpretability.

The remainder of this paper is structured as follows: Section 2 introduces the model and estimation framework, including the mathematical formulation of the model framework, as well as the additive predictors and details on the estimation approach. Section 3 presents the simulation study, detailing the basic settings and their variations, as well as specifying the performance evaluation metrics. The results of the different methods regarding estimation accuracy and efficiency are discussed at the end of this section. Section 4 applies Batchwise Backfitting to real-world data on infant mortality in sub-Saharan Africa, highlighting key findings, important factors and variable effects. Finally, Section 5 concludes with a discussion of the results, their implications, and potential directions for future research.

2. Flexible discrete hazard models

2.1. Model specification

We consider a discrete time-to-event framework in which time is represented as a sequence of discrete periods, and each individual experiences at most one event. An event is defined as a one-time transition from one state to another (e.g., death, dropout, or system failure). While we use the term *individual* for convenience, the concept can refer to any observational unit, including persons, households, or organisations. The model formulation and notation used in the following build on the frameworks of Tutz and Schmid (2016) and Berger and Schmid (2018).

Although time is generally conceptualised as continuous, in many practical applications it is recorded or analysed in discrete intervals (e.g., days, months, or years) depending on the nature of data collection (Tutz and Schmid 2016). The exact timing of an event is often unknown and is instead observed only within a specific interval. Formally, let $t \in \{1, 2, \dots, k\}$ index the discrete time intervals, which may either reflect naturally discrete time or grouped continuous time, with boundaries $[0, a_1), [a_1, a_2), \dots, [a_{k-1}, \infty)$. For n individuals, let $T_i \in \{1, \dots, k\}$ denote the event time for individual i , where $T_i = t$ implies that the event occurred in the interval $[a_{t-1}, a_t)$. The timing of the event is modelled in relation to p explanatory variables $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$.

Censoring is a common feature in time-to-event data, arising when the event time for an individual is not fully observed. The most frequent case is right censoring, which occurs when the start of observation is known, but the event has not occurred by the end of the observation period. For right-censored data, the observed time is defined as $t_i := \min(T_i, C_i)$, where T_i is the (possibly unobserved) event time and $C_i \in \{1, \dots, k\}$ denotes the censoring time for individual i . Other forms of censoring, such as left or interval censoring, are not considered in this work.

The hazard function is the central quantity in modelling time-to-event data. In a discrete-time setting, it is defined as the conditional probability that an event occurs at time t , given that the individual has “survived” up to that time. For a given set of explanatory variables \mathbf{x}_i , $i = 1, \dots, n$, the discrete-time hazard function is defined as

$$\lambda(t \mid \mathbf{x}_i) = P(T_i = t \mid T_i \geq t, \mathbf{x}_i), \quad t = 1, \dots, k, \quad (1)$$

which represents the probability of transitioning from the initial to the terminal state during interval t , conditional on survival up to t and covariates \mathbf{x}_i .

The corresponding survival function is given by

$$S(t \mid \mathbf{x}_i) = P(T_i > t \mid \mathbf{x}_i) = \prod_{r=1}^t (1 - \lambda(r \mid \mathbf{x}_i)), \quad t = 1, \dots, k, \quad (2)$$

and denotes the probability that the event occurs after time t , or equivalently, that the individual survives the interval $[a_{t-1}, a_t)$.

A general model for the discrete hazard function in Equation (1) can be defined as

$$\lambda(t \mid \mathbf{x}_i) = h(\eta_{it}), \quad (3)$$

where $h(\cdot)$ is a strictly increasing link function, such as the logit function $h(x) = \exp(x)/(1 + \exp(x))$, and η_{it} is the structured additive predictor for individual i at time t . In a flexible specification, η_{it} takes the form (Fahrmeir *et al.* 2022)

$$\eta_{it} = f_0(t; \beta_0) + f_1(\mathbf{x}_i; \beta_1) + \dots + f_p(\mathbf{x}_i; \beta_p), \quad (4)$$

where $f_0(\cdot)$ models the baseline hazard over time, and $f_1(\cdot), \dots, f_p(\cdot)$ represent (possibly nonlinear) covariate effects.

Each function $f_j(\cdot)$ is approximated using basis functions – such as penalized splines (e.g., Eilers and Marx 1996; Wood 2003) – with the (potentially nonlinear) effects expressed as

$$\mathbf{f}_j(\mathbf{X}_j; \beta_j) = \mathbf{X}_j \beta_j, \quad j = 0, \dots, p,$$

where \mathbf{X}_j denotes the design matrix and β_j the associated coefficients. Stacking all components, the predictor of the additive discrete hazard model can be expressed in matrix form as

$$\boldsymbol{\eta} = \sum_{j=0}^p \mathbf{X}_j \beta_j,$$

where $\boldsymbol{\eta}$ is the vector over all individuals and risk intervals. This corresponds to a flexible predictor commonly used in structured additive regression (STAR; Fahrmeir *et al.* 2004; Brezger and Lang 2006; Lang *et al.* 2014).

2.2. Penalized likelihood

In discrete hazard models, the primary focus is on estimating the model specified in equations (3) and (4). If the censoring mechanism is assumed to be independent of the explanatory variables related to the event time (i.e., uninformative censoring), the likelihood function simplifies considerably, as censoring does not affect the distribution of the event time. The individual contribution to the likelihood is given by

$$L_i = c_i P(T_i = t_i)^{\delta_i} P(T_i > t_i)^{1-\delta_i},$$

where the indicator variable $\delta_i = \mathbb{I}(T_i \leq C_i)$ distinguishes between observations with observed events ($\delta_i = 1$) and right-censored observations ($\delta_i = 0$), and $c_i := P(C_i \geq t_i)^{\delta_i} P(C_i = t_i)^{1-\delta_i}$ accounts for the censoring mechanism. Under the assumption of uninformative censoring, c_i can be treated as a constant and is thus omitted from the likelihood used to estimate $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)^\top$.

The likelihood can be expressed as a sum of individual contributions, where each contribution corresponds to the log-likelihood of a Bernoulli-distributed random variable. Specifically, for each individual i , we define a binary indicator y_{is} for each discrete time interval $[a_{s-1}, a_s)$ up to the observed or censored time t_i , with $y_{is} = 1$ if the event occurs in interval s , and $y_{is} = 0$ otherwise. Since each individual can experience the event at most once, the data can be represented as a sequence of independent Bernoulli trials, each with success probability equal to the discrete hazard $\lambda(t \mid \mathbf{x}_i)$. The total log-likelihood of the discrete hazard model

then becomes

$$\begin{aligned}\ell(\boldsymbol{\beta}; \mathbf{y}, \mathbf{X}) &\propto \sum_{i=1}^n \sum_{r=1}^{t_i} [y_{ir} \log \lambda(r \mid \mathbf{x}_i) + (1 - y_{ir}) \log (1 - \lambda(r \mid \mathbf{x}_i))] \\ &= \sum_{i=1}^n \sum_{r=1}^{t_i} [y_{ir} \log h(\eta_{ir}) + (1 - y_{ir}) \log (1 - h(\eta_{ir}))],\end{aligned}\quad (5)$$

where $\lambda(r \mid \mathbf{x}_i) = h(\eta_{ir})$ denotes the discrete hazard at time r for individual i , with $h(\cdot)$ representing a suitable strictly increasing link function, and η_{ir} the structured additive predictor as defined in Equation (4). The vector $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$ contains the covariate values for individual i , and the full design matrix is given by $\mathbf{X} = (\mathbf{X}_0, \dots, \mathbf{X}_p)$, where each \mathbf{X}_j corresponds to the basis functions used to represent the model term $f_j(\cdot)$, and \mathbf{X}_0 represents the design matrix for the smooth baseline effect $f_0(\cdot)$ in Equation (4).

To prevent the model terms $f_j(\cdot)$ from overfitting, regularization is introduced through a penalized log-likelihood of the form

$$\ell(\boldsymbol{\beta}, \boldsymbol{\tau}; \mathbf{y}, \mathbf{X}) \propto \ell(\boldsymbol{\beta}; \mathbf{y}, \mathbf{X}) - \sum_{j=0}^p \boldsymbol{\beta}_j^\top \mathbf{P}_j(\boldsymbol{\tau}_j) \boldsymbol{\beta}_j,$$

where $\boldsymbol{\tau}_j$ is a (possibly vector-valued) smoothing parameter that controls the degree of regularization applied to each model component $f_j(\cdot)$, and $\mathbf{P}_j(\boldsymbol{\tau}_j)$ is a corresponding quadratic penalty matrix determined by the structural assumptions on $f_j(\cdot)$.

For example, in the case of univariate smooth terms modelled via P-splines, $\boldsymbol{\tau}_j$ reduces to a scalar τ_j , and the penalty matrix typically takes the form $\mathbf{P}_j(\tau_j) = \tau_j \mathbf{K}_j$, where \mathbf{K}_j is based on second-order difference penalties on $\boldsymbol{\beta}_j$ (Fahrmeir *et al.* 2022, Section 8.1.2, pp. 449-464). The smoothing parameters $\boldsymbol{\tau}_j$ thus govern the overall functional form of the estimated effects.

2.3. Estimation

Estimation of discrete hazard models reduces to the estimation of a binary regression model but requires prior data augmentation. This involves expanding the original data set by adding pseudo-observations for each individual i and each time point $t = 1, \dots, t_i$ during which the individual is at risk.

This data augmentation step is best illustrated with an example. Imagine we are interested in $t = 1, 2, 3$ and the initial data set in Table 1 on the left.

The individual identification variable is labelled as *id*, t is the time variable, y is the event indicator, where 1 indicates an event and 0 otherwise, and x_1 , x_2 and x_3 are the time-invariant explanatory variables. Note that the individuals 1 and 4 are right-censored, as we do not observe the time to the event, since no event occurred in the observation period ($y = 0$).

Following augmentation, the data set takes the form shown on the right in Table 1. In this example, a total of five rows are added, two for individuals 2 and 4 and one for individual 3. The time variable t needs to cover the individual risk period, the event indicator y is 0 for the added rows, and the individual explanatory variables x_1 , x_2 and x_3 are duplicated.

In many applications, the resulting data expansion can be substantial, often leading to very large data sets that exceed the capacity of standard estimation algorithms. Efficient estimation methods for the GAM framework are therefore essential for handling such high-

Before augmentation:						After augmentation:					
<i>id</i>	<i>t</i>	<i>y</i>	<i>x</i> ₁	<i>x</i> ₂	<i>x</i> ₃	<i>id</i>	<i>t</i>	<i>y</i>	<i>x</i> ₁	<i>x</i> ₂	<i>x</i> ₃
1	1	0	x ₁₁	x ₁₂	x ₁₃	1	1	0	x ₁₁	x ₁₂	x ₁₃
2	3	1	x ₂₁	x ₂₂	x ₂₃	2	1	0	x ₂₁	x ₂₂	x ₂₃
3	2	1	x ₃₁	x ₃₂	x ₃₃	2	2	0	x ₂₁	x ₂₂	x ₂₃
4	5	0	x ₄₁	x ₄₂	x ₄₃	2	3	1	x ₂₁	x ₂₂	x ₂₃
						3	1	0	x ₃₁	x ₃₂	x ₃₃
						3	2	1	x ₃₁	x ₃₂	x ₃₃
						4	1	0	x ₄₁	x ₄₂	x ₄₃
						4	2	0	x ₄₁	x ₄₂	x ₄₃
						4	3	0	x ₄₁	x ₄₂	x ₄₃

Table 1: Toy example for the data augmentation step: Initial data set before augmentation on the left and data set after augmentation on the right.

dimensional settings (see, e.g., [Wood et al. 2017](#); [Lang et al. 2014](#)). Despite this, model term selection for such large models is difficult and not really considered in these implementations.

Recently, [Umlauf et al. \(2024\)](#) proposed the Batchwise Backfitting algorithm – an efficient approach for scalable estimation of distributional regression models with automatic selection of model terms. Since the discrete hazard model as stated in Equation (1) is a special case of a distributional regression model, the resulting Newton–Raphson-type updating equations can be written in an iterative form. These equations are used to maximize the log-likelihood in Equation (5) and estimate the coefficient vectors β_j . For iteration $l + 1$, the update for component j is given by

$$\beta_j^{[l+1]} = (\mathbf{X}_j^\top \mathbf{W} \mathbf{X}_j + \mathbf{P}_j(\tau_j))^{-1} \mathbf{X}_j^\top \mathbf{W}(\mathbf{z} - \boldsymbol{\eta}_{-j}^{[l+1]}), \quad (6)$$

where $\mathbf{z} = \boldsymbol{\eta}^{[l]} + \mathbf{W}^{-1} \mathbf{u}$ is a vector of working observations and $\mathbf{u} = \partial \ell(\beta; \mathbf{y}, \mathbf{X}) / \partial \boldsymbol{\eta}$ is the score vector. $\mathbf{W} = -\text{diag}(\partial^2 \ell(\beta; \mathbf{y}, \mathbf{X}) / \partial \boldsymbol{\eta}^2)$ is a *diagonal weight matrix of size* $N \times N$, where $N = \sum_{i=1}^n t_i$ is the total number of augmented binary observations. Each diagonal entry in \mathbf{W} is evaluated at the current state $\beta^{[l]}$. The expression in Equation (6) corresponds to a backfitting update for the model term $f_j(\cdot)$, with $\boldsymbol{\eta}_{-j}$ denoting the predictor excluding the j -th model component.

The backfitting loop in Equation (6) cycles through model terms $j = 0, \dots, p$, and is iterated until convergence, e.g., when the relative change in the coefficients falls below a prespecified threshold. Smoothing parameters τ_j can be estimated using a stepwise selection procedures (similar to [Belitz and Lang 2008](#)), optimizing each component sequentially within adaptive search intervals and using information criteria (e.g., AIC or BIC). As mentioned above, in many practical cases, τ_j reduces to a scalar. For further algorithmic details, see [Umlauf et al. \(2018\)](#).

Scalable estimation is achieved by using only a randomly selected batch of the data. In discrete hazard models, where each individual contributes t_i observations (one for each time interval during which they are at risk) to the augmented data set, the natural sampling unit is the individual. If an individual is selected for a batch, then all the corresponding observations

are included. Formally, let

$$\mathbf{i} = \left(1, \dots, \sum_{i=1}^n t_i\right)^\top = \left(\mathbf{i}_1^\top, \dots, \mathbf{i}_n^\top\right)^\top$$

denote the vector of stacked row indices in the augmented data set, where

$$\mathbf{i}_i = \left(\sum_{m=1}^{i-1} t_m + 1, \dots, \sum_{m=1}^i t_m\right)^\top$$

gives the indices corresponding to individual i , for $i = 1, \dots, n$. A batch is then defined by selecting a random subset $s \subseteq \{1, \dots, n\}$ of individuals and using all corresponding observations, i.e., $\mathbf{i}_s = \bigcup_{i \in s} \mathbf{i}_i$.

The corresponding response vector $\mathbf{y}_{[\mathbf{i}_s]}$ and covariate matrix $\mathbf{X}_{[\mathbf{i}_s]}$ are used to compute a stochastic updating step of the form

$$\begin{aligned} \beta_j^{[l+1]} &= (1 - \nu) \cdot \beta_j^{[l]} + \\ &\quad \nu \cdot (\mathbf{X}_{[\mathbf{i}_s],j}^\top \mathbf{W}_{[\mathbf{i}_s]} \mathbf{X}_{[\mathbf{i}_s],j} + \mathbf{P}_j(\tau_j))^{-1} \mathbf{X}_{[\mathbf{i}_s],j}^\top \mathbf{W}_{[\mathbf{i}_s]} (\mathbf{z}_{[\mathbf{i}_s]} - \boldsymbol{\eta}_{[\mathbf{i}_s],-j}^{[l+1]}) \\ &= (1 - \nu) \cdot \beta_j^{[l]} + \nu \cdot \beta_{[\mathbf{i}_s],j}^{[l]}, \end{aligned} \tag{7}$$

where ν is the step length control parameter (or *learning rate*) specifying the amount of which $\beta_j^{[l]}$ is updated in the direction of the new estimate $\beta_{[\mathbf{i}_s],j}$ on batch $[\mathbf{i}_s]$. Note that the working weights $\mathbf{W}_{[\mathbf{i}_s]}$ and the score vectors \mathbf{u} used to compute $\mathbf{z}_{[\mathbf{i}_s]}$ are evaluated on the $[l]$ -th estimate $\beta_j^{[l]}$. In each iteration, Equation (7) is evaluated on exactly one batch $[\mathbf{i}_s]$, such that the computational burden can be reduced considerably.

The updating function defined in Equation (7) can be applied in several ways. For instance, if the step length control parameter is set to a small value, such as $\nu = 0.1$, and only the best-fitting model term $f_j(\cdot)$ is updated in each iteration, the algorithm mimics a boosting-type approach. Specifically, the decision to update a model term is based on its log-likelihood contribution, evaluated on another random batch $[\hat{\mathbf{i}}_s]$. This design helps mitigate overfitting, as demonstrated by Umlauf *et al.* (2024). Similarly, smoothing parameters τ_j are selected by minimizing an information criterion (e.g., AIC) on a different independent batch. This simultaneous selection of model terms and smoothing parameters offers a major advantage as it eliminates the need to determine an optimal stopping iteration, which is typically required in classical boosting algorithms (e.g., via computationally intensive cross-validation).

The algorithm is considered to have converged when the “out-of-sample” log-likelihood evaluated on other batches no longer improves. To better account for uncertainty, a common strategy is to refit the selected model with $\nu = 1$, mimicking a resampling step. Estimates are then based on the last iterations, like using Markov chain Monte Carlo (MCMC) simulation. This algorithm has proven to have excellent model term selection performance and can be applied to very large data sets, as it is often the case in discrete hazard models. For a detailed description of the algorithm, please refer to Umlauf *et al.* (2024).

The number of batches and thus the number of iterations L should be chosen sufficiently large such that no (major) further improvements in the out-of-sample log likelihood can be observed. In this work, L is between 200 and 500 in the simulation study and 1000 in the

application. Furthermore, a reasonable batch size M must be defined, which in previous exercises was chosen between 10000 and 50000 (e.g. [Umlauf et al. \(2024\)](#) or [Seiler et al. \(2025a\)](#)). Depending on the complexity and type of data, the batches can be smaller or must be larger to ensure sufficient coverage of the information in the data. For both the simulation and the application, we use a batch size of $M = 20000$.

3. Simulation study

A comprehensive simulation study is conducted to investigate the performance of Batchwise Backfitting for estimating discrete time-to-event models, especially in the context of extensive data sets. We apply Batchwise Backfitting and suitable benchmark methods in different settings and evaluate estimation accuracy, variable selection performance and estimation time. To ensure robustness of the results, each setting is replicated 250 times for each method under study.

3.1. Design and basic setting

Using the framework developed in Section 2, data is simulated for individuals $i = 1, \dots, n$ and time $t = 1, \dots, k$. In the basic setting, the number of individuals n is set to 5000. The number of time points k is kept at 20 in all settings of the simulation study.

The predictor is defined as

$$\eta_{it} = f_0(t) + f_1(x_{i1}) + \dots + f_9(x_{i9}), \quad (8)$$

where $f_0(t)$ represents the baseline hazard function, $f_1(x_{i1}), \dots, f_4(x_{i4})$ are the effects of informative variables x_{i1}, \dots, x_{i4} , and $f_5(x_{i5}) \equiv 0, \dots, f_9(x_{i9}) \equiv 0$ are the effects of non-informative/noise variables x_{i5}, \dots, x_{i9} .

The *baseline hazard* function $f_0(t)$ is specified decreasing logarithmically over time as

$$f_0(t) = a - \frac{1}{2} \log(t).$$

In the basic setting $a = -3$, which corresponds to an event frequency of approximately 10%. This specification is shown in Figure 1 as solid curve.

Equidistant design points are generated from the interval $[0, 1]$ for the variable x_{i4} and from the interval $[-3, 3]$ for the remaining variables. The effects of the informative variables are defined as

$$\begin{aligned} f_1(x_1) &= 0.5 \cdot x_1, & f_2(x_2) &= 1.5 \cdot \sin(x_2), \\ f_3(x_3) &= x_3^{\frac{2}{3}} - 1.5, & f_4(x_4) &= \sin(2 \cdot (4 \cdot x_4 - 2)) + 2 \cdot e^{-(16^2) \cdot (x_4 - 0.5)^2}, \end{aligned}$$

see Figure 2. The unique values of the effects $f_1(x_1), \dots, f_4(x_4)$ are scaled to have a specific standard deviation (SD), which is set to $SD = 1$ in the basic setting. In the following, we refer to $f_1(x_1)$ as *linear*, $f_2(x_2)$ as *sinus*, $f_3(x_3)$ as *squared*, $f_4(x_4)$ as *complex* and $f_5(x_5), \dots, f_9(x_9)$ as *noise 1* to *noise 5*.

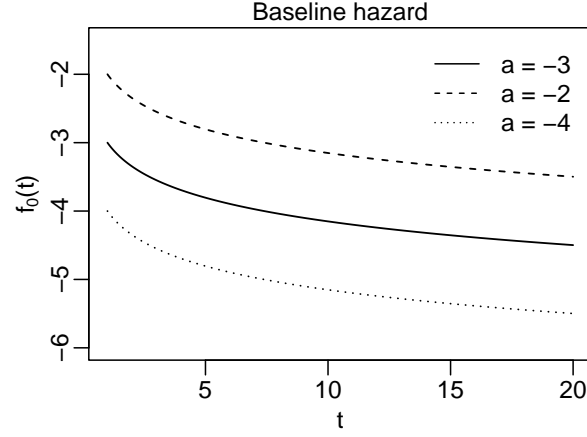


Figure 1: Specifications of the baseline hazard function $f_0(t)$. $a = -3$ is used in the basic setting. $a = -2$ and $a = -4$ are variations to investigate the influence of different event frequencies.

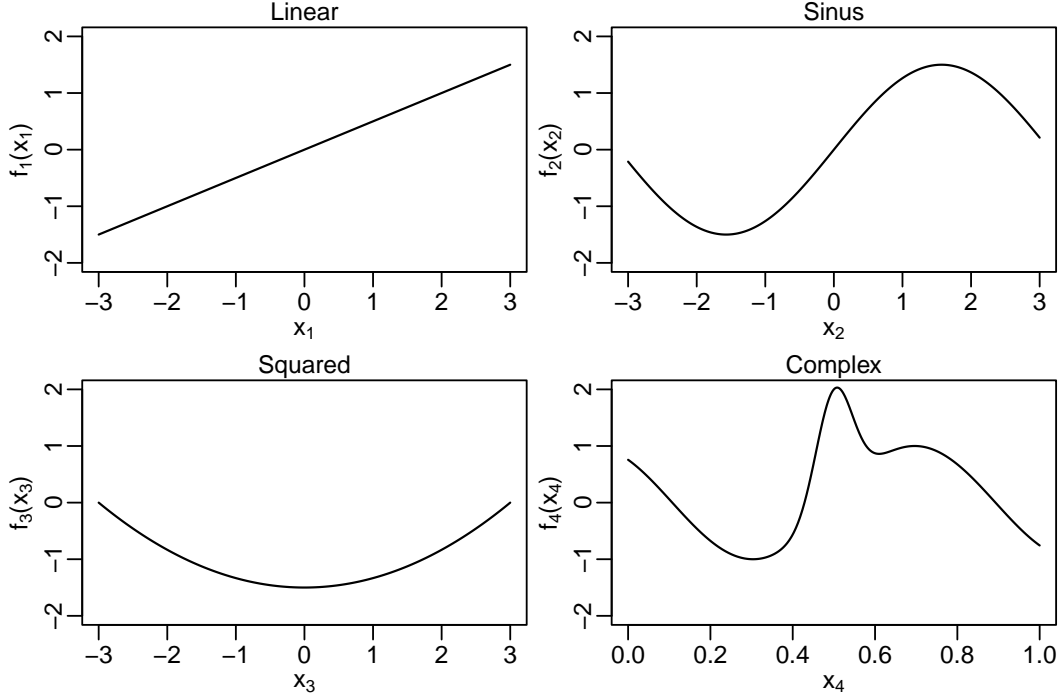


Figure 2: Univariate informative effects on η_{it} .

The event indicator y_{it} is derived by comparing $\lambda(t|\mathbf{x}_i)$ with random draws from the uniform distribution $u_{it} \sim \mathcal{U}(0, 1)$, i.e.,

$$y_{it} = \begin{cases} 1, & \text{if } u_{it} \leq \lambda(t|\mathbf{x}_i), \\ 0, & \text{otherwise.} \end{cases}$$

For each individual, only the first occurrence of an event (i.e., the first time $y_{it} = 1$) is considered as the event time T_i , and a random censoring time C_i is generated, where $C_i \sim \mathcal{U}(\{1, 2, \dots, 20\})$. The individual's observation time is then defined as $t_i = \min(T_i, C_i, 20)$.

This results in approximately $(n \cdot k)/2$ rows in the augmented data set used for estimation. In the basic setting (where $n = 5000$), the augmented data set consists of around 50000 rows.

3.2. Further settings

The further settings are variations of the basic setting to evaluate the performance of the estimation methods under different circumstances. In each setting, one feature is varied, while all other components remain as in the basic setting. The number of individuals, the baseline hazard specification and the scaling of the effects are varied and a predictor with a spatial effect is simulated:

- **Number of individuals n :** To investigate the performance for both small and large data sets, simulations are performed with number of individuals $n = 1000; 10000; 50000; 100000; 500000$ and 1000000 . This results in augmented data sets that range from approximately 10000 to 10000000 rows.
- **Baseline hazard specification $f_0(t)$:** Alternative baseline hazard specifications are considered to analyse the influence of different event frequencies. Simulations are performed with $a = -2$ and $a = -4$, which corresponds to an event frequency of approximately 20% and 5%, respectively. The two variations are also visualised in Figure 1.
- **Effect scaling SD :** The scaling of the unique values of the effects is varied to $SD = 0.5$ and $SD = 2$ to investigate the influence of different effect sizes. However, since the scaling of the effects does not reveal any remarkable pattern, we refrain from presenting and discussing these settings in the remainder.
- **Spatial effect $f_{spa}(lon, lat)$:** In addition to the univariate effects, a bivariate spatial effect $f_{spa}(lon, lat)$ is added to the model. It is defined as

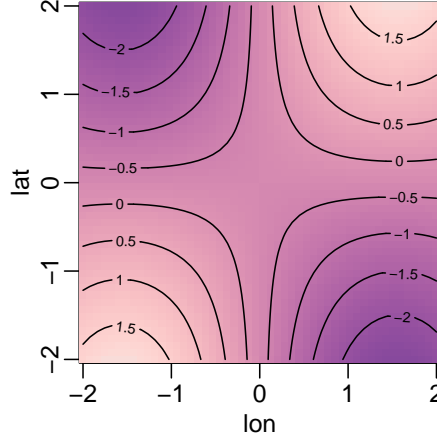
$$f_{spa}(lon, lat) = 2.5 \cdot \sin(lon) \cdot \sin(0.5 \cdot lat) - 0.3,$$

and the updated predictor is $\theta_{it} = \eta_{it} + f_{spa}(lon, lat)$, where η_{it} is defined in Equation (8). For lon and lat equidistant design points are sampled from the interval $[-3, 3]$. The spatial effect is visualised in Figure 3.

3.3. Methods and implementation

In the following, we refer to Batchwise Backfitting as BBFIT. The method is implemented in the R package **bamlss** (Umlauf *et al.* 2018, 2021) and the Batchwise Backfitting algorithm (Umlauf *et al.* 2024) can be used by setting `optimizer = opt_bbfit` in the main function `bamlss()`. The method is applied as outlined in Section 2, see also Umlauf *et al.* (2024) or Seiler *et al.* (2025a) for a less technical description.

Model comparison is based on the out-of-sample AIC. We use a combination of the boosting and resampling variants of the algorithm in a two-step procedure. In the first step, the boosting variant is executed for $L = 200$ iterations including all possible covariates with the default step length $\nu = 0.1$. To perform a variable selection, the argument `select = TRUE` is specified. This step serves as a covariate preselection for the second step, the resampling variant of the algorithm. The second step is also executed for $L = 200$ iterations, with the

Figure 3: Bivariate spatial effect on θ_{it} .

first 100 iterations discarded as a burn-in phase. In this step, the model contains only the selected covariates. The standard step length $\nu = 1$ is applied, and slice sampling (Neal 2003) is used for the smoothing parameters.

The batches are generated by sampling individuals at random and including all observations of a sampled individual in a batch to account for the augmented data structure (see Section 2.3). We define a (maximum) batch size $M = 20000$. This implies that for the setting with 1000 individuals, corresponding to an augmented data set with around 10000 rows, each batch consists of the entire model data set including all individuals. For settings with more individuals, only a portion of the data set is included in each batch until M is reached.

The model specification follows the approach introduced by Wood (2003), using thin-plate regression splines $\mathbf{s}(\cdot)$ for the baseline hazard function and all univariate effects of the explanatory variables, with the default basis dimension ($\mathbf{k} = 10$). For the bivariate spatial effect, a full tensor product smooth is used in its default form via `te()`, as described in Wood (2006).

The performance of BBFIT is compared with two benchmark methods:

1. As a basic benchmark, we use an estimation method for generalized linear models with stepwise model selection based on the Akaike information criterion (AIC). Specifically, the function `stepAIC()` from the R package **MASS** (Venables and Ripley 2002) with the specification `direction = 'both'` is applied. The baseline hazard function and all effects of the explanatory variables are modelled with third-degree polynomials. In the following, we refer to this approach as GLM.
2. For more flexible specifications, an estimation method for generalised additive models optimised for (very) large data sets is employed. Here we use the `bam()` function (Wood *et al.* 2015, 2017) from the R package **mgcv** (Wood 2003, 2004, 2011; Wood *et al.* 2016, 2017). The smoothing parameter is estimated with a fast version of a restricted maximum likelihood approach. This is the default approach of `bam()` specified by `method = "fREML"`. For variable selection, we set `select = TRUE`. In the following, this approach is referred to as BAM.

3.4. Performance measures

The measures aim to evaluate the performance of the estimation methods in terms of effect estimation, variable selection and estimation time:

- **Effect estimation:** To evaluate the performance of the methods in terms of effect estimation, all estimated effects are plotted against the true effect and the means and quantiles are displayed to detect systematic biases. Furthermore, the MSE of the estimated effects is calculated, defined as the mean of the squared differences between the true effects and the estimated effects, i.e. $MSE_f = \frac{1}{n} \sum_{i=1}^n (\hat{f}(x_i) - f(x_i))^2$, where $f(\cdot)$ is the true, $\hat{f}(\cdot)$ is the estimated effect and x_1, \dots, x_n are design points spanning the entire range of x .
- **Variable selection:** For the selection performance, we examine selection frequencies, with particular attention to the noise/uninformative variables. This frequency is computed by dividing the number of times each variable is selected by the total number of replications. For the methods GLM and BBFIT, selection is determined by whether a variable is included or excluded from the model. For BAM, however, there is no such clear rule, since effects of uninformative variables are estimated to be close to zero, but never exactly zero. To establish a selection rule for BAM, we evaluate several thresholds based on *equivalent degrees of freedom (EDF)* and *p-values*, and identify *p-value* < 0.01 as the best criterion, see Section 3.6 below for more details.
- **Estimation time:** Estimation time refers to the time required to estimate the model, which is exclusively the time that the functions need for the estimation. The estimation time is based on replications that are run on a “standard” PC (see Section 3.5) to ensure practical relevance.

3.5. Computational details

The simulation was run on the high-performance cluster (HPC) infrastructure LEO4 of the University of Innsbruck. This HPC infrastructure runs on a Linux system (CentOS 7), and 50 computing nodes with Intel Xeon (Broadwell/Skylake) processors with up to 3000 gigabyte (GB) available memory. To analyse the estimation time for the methods under study, some replications were re-run on a “standard” PC, as this is a typical setup for users. This PC runs with the Ubuntu 24.04.1 LTS operating system. The model is a Lenovo ThinkPad T14s Gen 3 with an AMD Ryzen 7 PRO 6850U processor, AMD Radeon Graphics (x 16) and 32 GB of available memory.

3.6. Results

Basic setting

Figure 4 shows the estimated effects of all informative variables and the first noise variable, while Figure 5 shows the corresponding distribution of the MSE. Since the other noise variables show very similar patterns, their results are omitted here, but can be found in Figures 21 and 22 in the Appendix A.1.

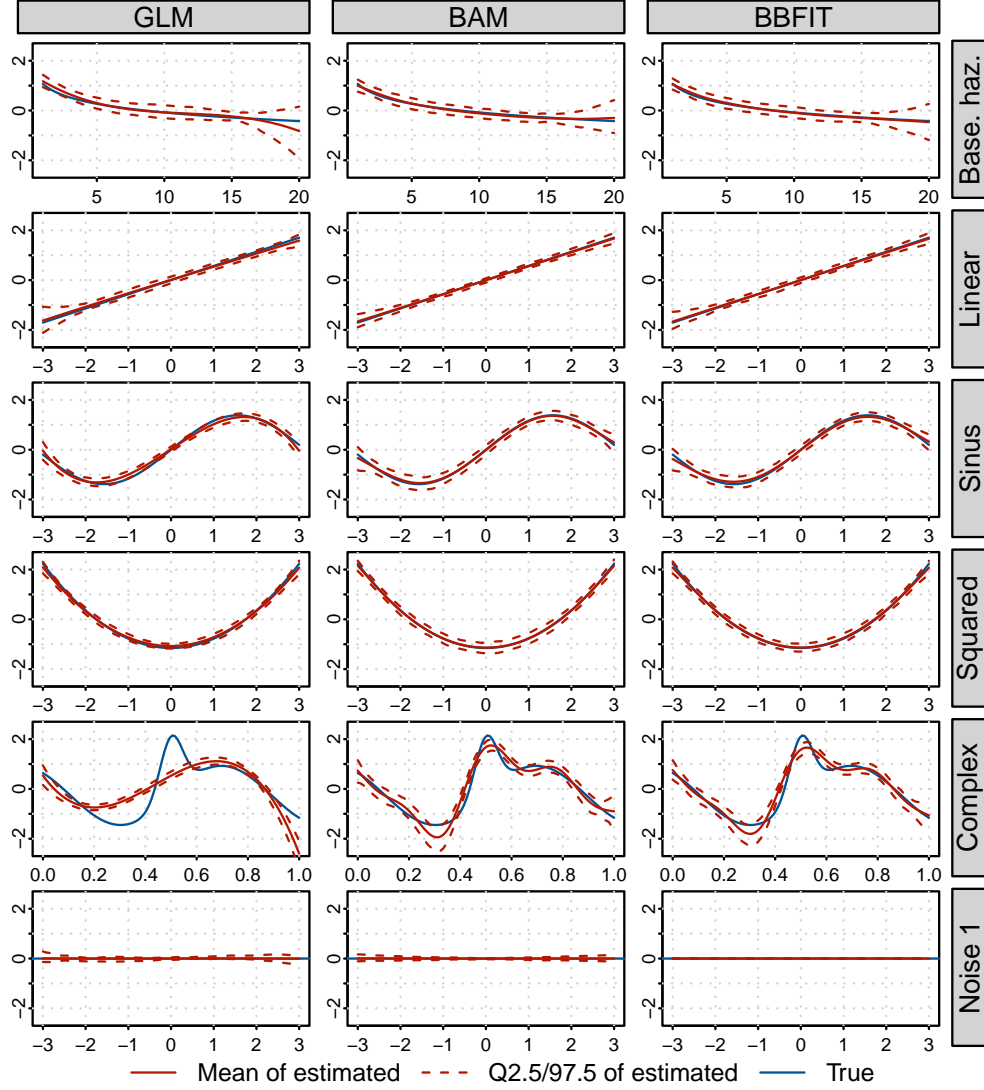


Figure 4: Estimated effects of all informative variables and the first noise variable in the basic setting (see Section 3.1).

To analyse the estimated effects in more detail, we created a web app that allows replications and effects to be visualised separately or as an overlay in a dynamic plot. The app also offers the possibility to explore the further settings described below. It can be accessed via bmueller5000.github.io/bb4sa-shinylive/ and the corresponding data can be found in fileshare.uibk.ac.at/d/d78d077c7d184beab2ba/.

Further results regarding the selection frequencies are discussed in Figure 6. Since the informative variables are always selected, we restrict the presentation to the noise variables. As mentioned above, the selection approach of BAM does not provide a clear selection rule. Therefore, different thresholds for the *EDF* and *p-values* are tested as selection criteria and visualised in Figure 7.

From Figures 4 to 7 and the web app, we draw the following conclusions:

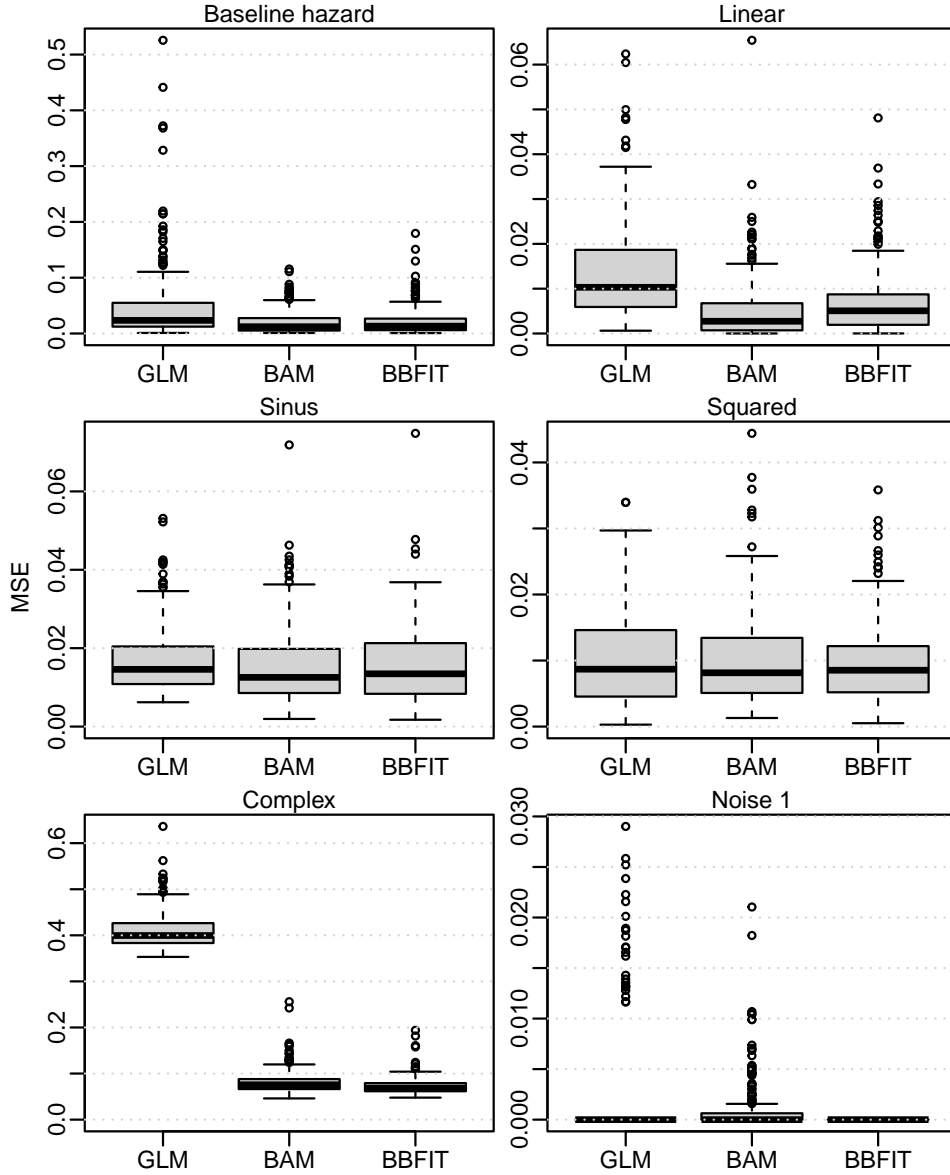


Figure 5: MSE of the baseline hazard, the complex effect and the first noise effect in the basic setting (see Section 3.1). Note: The MSEs of BBFIT for Noise 1 (bottom right) are always zero as the noise variables are never selected.

- *Bias*: All methods provide largely unbiased estimates for the linear, sinus, and the squared effect f_1 , f_2 , f_3 and substantial bias for the complex effect f_4 . This bias is, as expected, more pronounced for the GLM method, since polynomials of degree three are not flexible enough to capture such complex effects. The GLM method also shows a clear bias for the baseline hazard f_0 . BAM shows a small bias in the baseline hazard f_0 at higher t values, which is almost invisible in Figure 4. Surprisingly, this bias increases with the number of individuals, see the more detailed discussion below.
- *MSE*: The main findings from the visible inspection of the bias are reflected in the MSEs. For the linear, sinus, and the squared effect f_1 , f_2 , f_3 the MSEs are almost zero

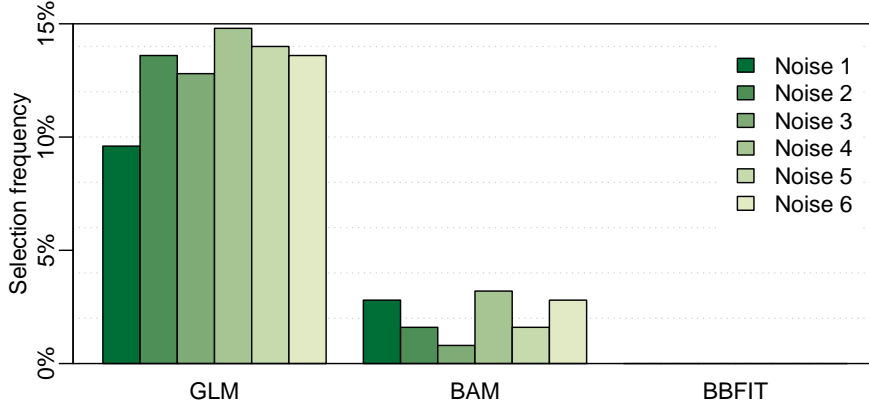


Figure 6: Selection frequencies of the noise variables in the basic setting (see Section 3.1).

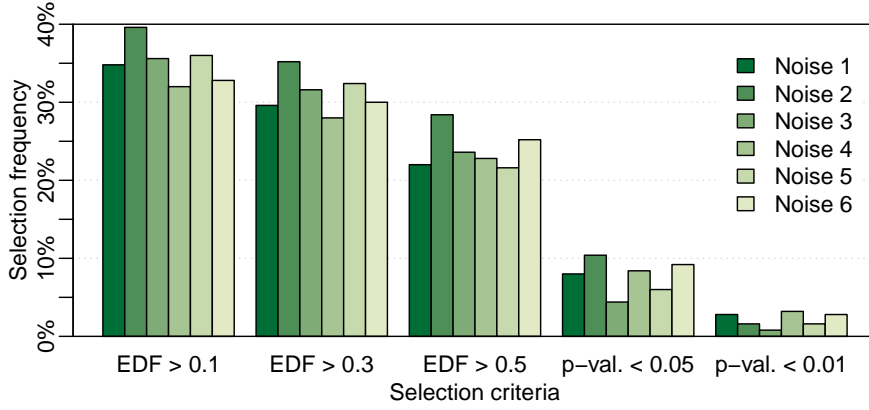


Figure 7: Selection frequencies of the noise variables for different selection criteria of BAM in the basic setting (see Section 3.1).

for all methods. For the baseline hazard f_0 , BAM and BBFIT yield lower MSEs than GLM and for the complex effect f_4 , GLM is clearly outperformed. Regarding the noise variables, BBFIT achieves MSEs of exactly zero in all replications. BAM shows low but non-zero MSE values due to occasional false positive selections, while GLM produces considerably higher MSEs for Noise 1.

- *Selection frequencies:* Compared to the p -values, selection via the EDF consistently leads to higher selection frequencies for the noise variables in BAM (Figure 7). The best results are obtained with the rule $p\text{-value} < 0.01$ which is therefore used for BAM in the remainder. Figure 6 shows exceptional selection performance of BBFIT as the noise variables are never selected. The selection frequency of BAM (based on the rule $p\text{-value} < 0.01$) varies between 0.8% and 3.2% and GLM is not competitive with selection frequencies between 10% and 15%.

Further settings: Number of individuals

Figure 8 shows the average MSE across all 250 replications for varying numbers of individuals, focusing on the baseline hazard f_0 , the complex effect f_4 , and the first noise effect f_5 . These

three effects show the most notable patterns. A plot including all effects is provided in Appendix A.2, Figure 23. Estimation using the GLM method was not feasible for data sets with more than 100000 individuals and is therefore excluded from the analysis of large data sets.

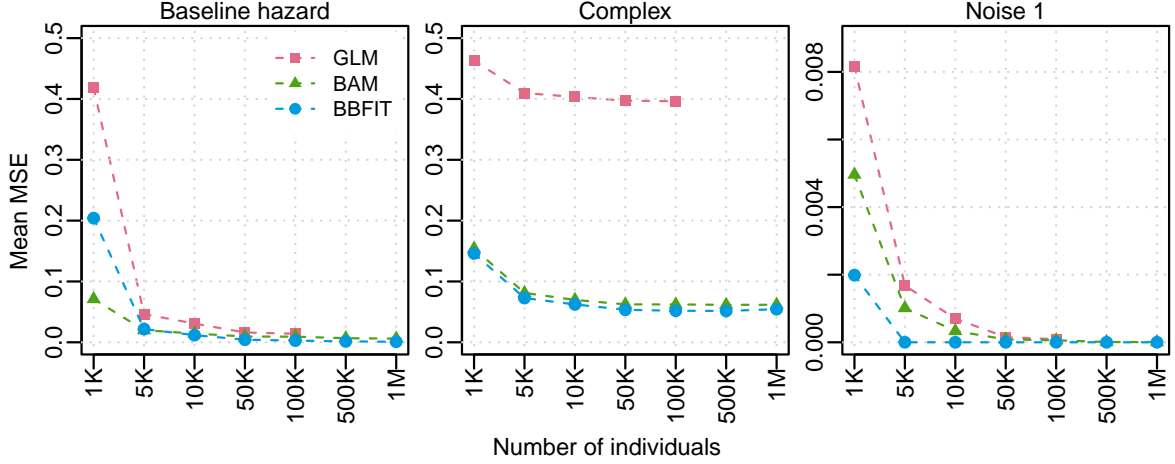


Figure 8: Mean MSE for different number of individuals of the baseline hazard, the complex effect and the first noise effect.

With increasing numbers of individuals, the estimation accuracy of the methods increases, which is reflected in a decreasing average MSE. However, the average MSE is almost constant for all methods and effects from 5000 individuals. As already discussed in the basic setting, GLM performs worst in the estimation of the complex effect f_4 due to model specification limitations. BAM and BBFIT perform more or less equally well for 50000 to 1000000 individuals, and for fewer individuals no method is superior to the other.

Figure 9 displays the estimated baseline hazards based on 500000 individuals, comparing the BAM method with variable selection (`select=TRUE`), which is used as the standard throughout this work, to the `bam()` method without variable selection (`select=FALSE`).

As mentioned above and clearly present with 500000 individuals, BAM shows a bias in the baseline hazard at higher t values. To investigate this bias further, we tested different methods for estimating the smoothing parameters, as well as varied the number of nodes and the smoothing basis for the splines. These variations did not resolve the issue. However, we found that the bias is not present when using the methods without variable selection, indicating that the variable selection approach might cause this issue. Further research is needed to test for systematic biases with the selection method of `bam()` in a discrete time-to-event data structure and in general, less special, data structures.

Figure 10 shows the mean selection frequency of the six noise variables for different number of individuals. The observed patterns are consistent for the individual noise variables, implying that the mean selection frequency effectively summarises the qualitative implications.

Independent of the number of individuals, the method GLM performs worst with a mean selection frequency of around 10% to 12%. BAM performs significantly better with a selection rate around 1% to 4%. The best selection performance is achieved using BBFIT with perfect selection for 5000 or more individuals. Interestingly, the selection performance of all methods

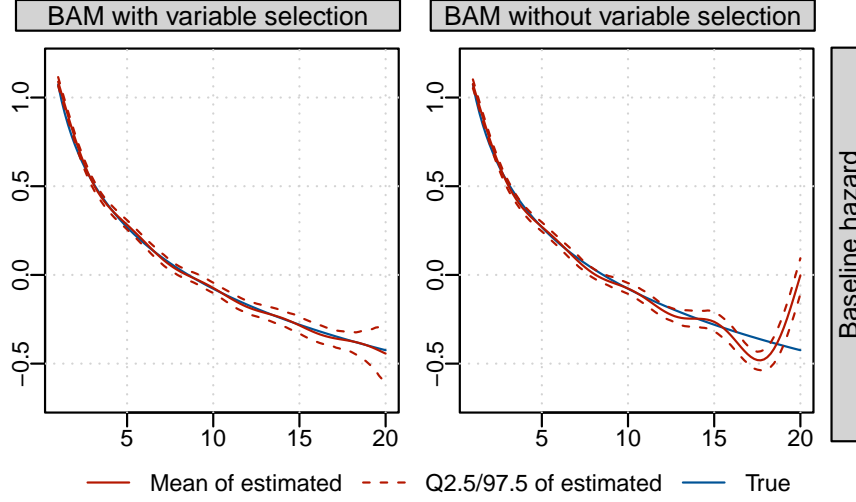


Figure 9: Estimated effects of the baseline hazard with 500000 individuals.

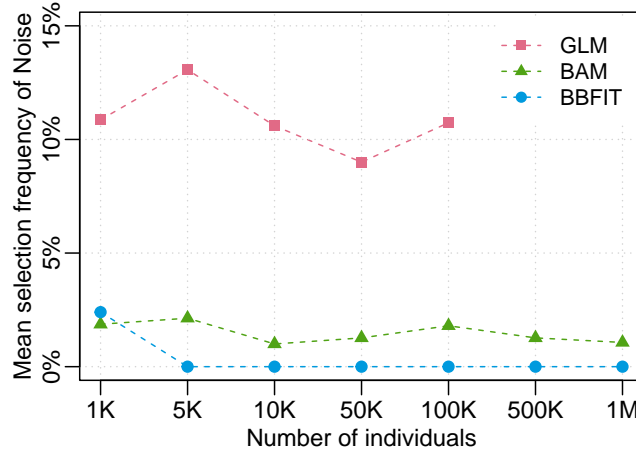


Figure 10: Mean selection frequency of the six noise variables for different number of individuals.

is relatively unaffected by the number of individuals.

Figure 11 shows the median estimation time for different numbers of individuals based on 25 replications on a standard PC (see Section 3.5). The median estimation time for individuals up to 100000 ranges from a few seconds to around 3 minutes and is comparable for the three methods. For very large sample sizes of 500000 or even 1000000 individuals, BBFIT clearly outperforms BAM with median estimation time of BBFIT being about half as long as with BAM.

Figure 12 further shows the distribution of estimation times for 100000, 500000 and 1000000 individuals as box plots including the mean. While the estimation times for GLM and BBFIT are relatively stable, BAM shows some large outliers. For 500000 individuals, for example, 4 of 25 replications have estimation times of about 250 minutes, while the rest have about 13 minutes, resulting in a large discrepancy between the mean and the median. These large outliers in estimation times are not only observed on our PC, but also occur when running

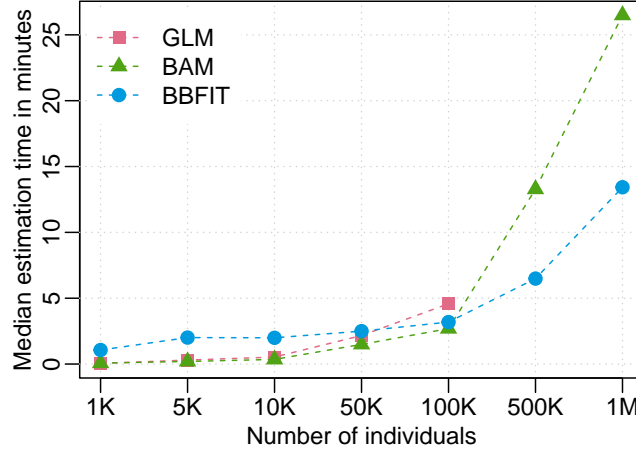


Figure 11: Median estimation time in minutes for different number of individuals based on 25 replications run on a “standard” PC (see Section 3.5).

the models on the HPC. For these exceptionally long estimation times, a warning is issued that the algorithm did not converge. However, when looking at the estimated effects of these replications, we found no qualitative differences compared to the replications with normal estimation times and no warning.

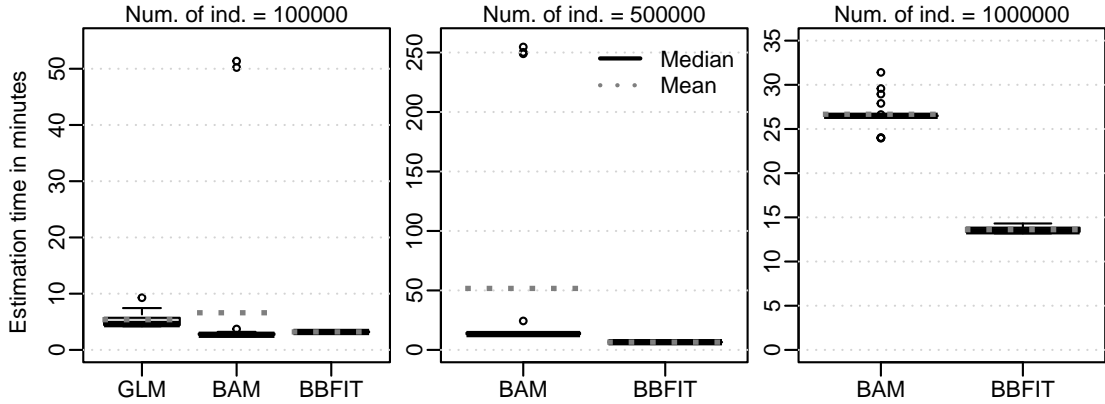


Figure 12: Boxplots of estimation time in minutes for 100000, 500000 and 1000000 individuals based on 25 replications run on a “standard” PC (see Section 3.5).

Further settings: Baseline hazard specification

The alternative baseline hazard specifications $a = -2$ and $a = -4$ (see Figure 1) have no considerable influence on the estimation performance in terms of estimation accuracy of the methods under study. However, we find an effect on the selection frequencies. Figure 13 shows the mean selection frequencies of the six noise variables for the three baseline hazard specifications. A notable result occurs with the specification $a = -4$ (event frequency of around 5%). Here, the selection performance of BBFIT rises from 0 to about 5% and is outperformed by BAM.

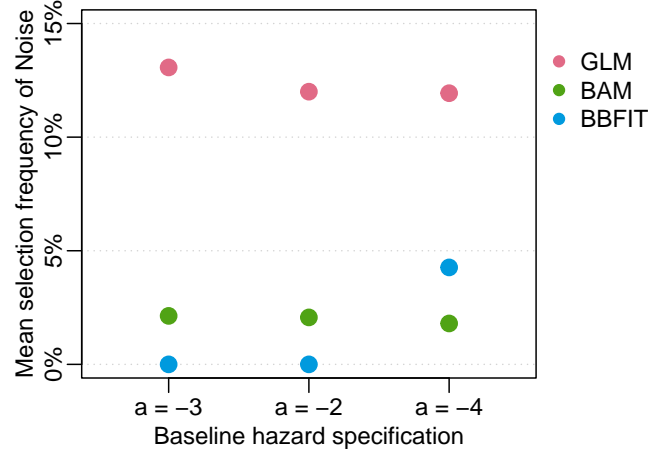


Figure 13: Mean selection frequency of the six noise variables for different baseline hazard specifications.

Further settings: Spatial effect

The bivariate spatial effect can only be added with BAM and BBFIT. The GLM method does not support any multivariate functional forms. Figure 14 shows the distributions of the MSE based on 250 replications.

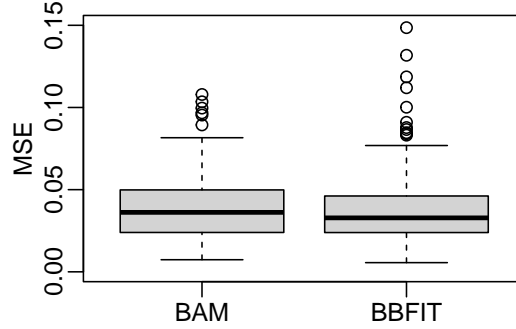


Figure 14: MSE of bivariate spatial effect.

Both methods estimate the spatial effect equally well. BAM shows a marginally larger bias in some areas, but overall the MSE is similar for the two methods.

4. Application

Despite considerable improvements in the past, infant mortality rates (i.e., the number of deaths of children under one year of age per 1000 live births in a given year) remain high in some parts of the world. This is especially true in sub-Saharan Africa, where the most recently reported infant mortality rate was 4.4% (i.e., 44 per 1000 live births) in 2023 (United Nations Children's Fund 2025). While this reflects a substantial reduction from 10.2% in 1990, the rate remains high compared to other regions in the world. For instance, the reported infant mortality rate in South Asia – the second most affected region – was 3%, and in North America, it was about ten times lower than in sub-Saharan Africa in the same year. In

absolute terms, around 1.8 million children under one year of age died in sub-Saharan Africa, corresponding to around 51% of all infant deaths globally in 2023.

Child mortality (including neonatal, infant, and under-five mortality) is a complex issue influenced by a range of factors, such as inadequate access to health care, poor nutrition, and environmental conditions. Therefore, child mortality rates serve as crucial indicators of overall human development. This is prominently recognized in the third Sustainable Development Goal (*SDG 3*) – *Good health and well-being* (United Nations 2015). Promoting child health is a high priority in *SDG 3*, with one of the targets being to reduce neonatal and under-five mortality rates in all countries by 2030.

The analysis of infant mortality involves modelling the probability of death between birth and the first birthday (Rutstein and Rojas 2006). In our application, we employ a time-to-event model with an additive predictor to model and estimate infant mortality in ten countries in eastern sub-Saharan Africa.

4.1. Data

For modelling infant mortality in eastern sub-Saharan Africa, Demographic and Health Surveys (DHS, ICF (2004-2017)) provide representative and standardised survey data commonly used to monitor health related outcomes in low- and middle-income countries. This data is merged with remotely sensed data on climate, demography, environmental information to create a unique data source. The geo-spatial data was initially compiled for Seiler *et al.* (2025a) to estimate the prevalence of anemia in children. They provide a detailed overview of the data and details of the preprocessing steps in the Section *Data* and especially in the *Supplementary Information* (Seiler *et al.* 2025b).

In this application, we use a subset of that data on individual children from the eastern sub-Saharan Africa countries (based on the UN geoscheme classification, United Nations Statistics Division 2025) Burundi, Ethiopia, Kenya, Malawi, Mozambique, Rwanda, United Republic of Tanzania, Uganda, Zambia, and Zimbabwe. Further data cleaning involved the exclusion of surveys older than the year 2000, observations with missing covariates and observations born more than five years before the survey date. The resulting data set contains 351705 individual children, of which about 4.5% died in the first year of life. As discussed in Section 2, the data must be augmented before fitting the models. This augmentation step substantially increases the data set up to around 3.7 million rows, which are ultimately used for estimation.

All basic information for modelling mortality is recorded for these children. This includes an indicator whether the child is still alive at the time of observation (*dead*), as well as information on the age at the time of observation or the age at death on a monthly basis (*age*). In addition, the data set contains numerous potential explanatory variables, such as the age of the mother at birth (*magebirth*), information on the material wealth of the household (*ai*) and the average temperature anomalies at the household’s location (*t2m2*). We can draw on a total of 25 potential explanatory variables, which are listed in Table 2.

These variables can be broadly categorised into child-related, maternal, socio-economic, community-related and climatic factors, all of which are considered potentially important explanatory factors associated with infant mortality. The selection of variables is informed by the *UNICEF Conceptual Framework on the Determinants of Maternal and Child Nutrition* (UNICEF 2020). While the framework focuses on undernutrition, it is particularly relevant for studying child survival outcomes, as nearly half of all child deaths can be attributed to

Table 2: Covariates included in the full model as used in the selection step.

Variable	Unit	Type	Description	Source/Reference
<i>Response</i>				
dead	1 if "dead"; 0 if "alive"	Binary	Death/living status at the day of the interview	DHS; ICF (2004-2017)
<i>Continuous or quasi-continuous covariates</i>				
age	Months	Metric	Age of child	DHS; ICF (2004-2017)
ai	Index	Continuous	Asset index of the household	DHS; ICF (2004-2017)
altitude	Meters	Continuous	Altitude in m above sea level	NOAA ETOPO; Amante and Eakins (2009); NOAA National Geophysical Data Center (2009)
bord	Count	Metric	Birth order within household	DHS; ICF (2004-2017)
l_distance	Kilometers	Continuous	Log of distance to the closest body of water	GRG Washington; NASA JPL (2013)
hhs	Count	Metric	Household size	DHS; ICF (2004-2017)
hghedyear	Years	Metric	Highest completed year of schooling	DHS; ICF (2004-2017)
magebirth	Years	Metric	Age of mother at birth	DHS; ICF (2004-2017)
minc12	Prevalence	Continuous	Malaria incidence	Malaria Atlas Project; Weiss <i>et al.</i> (2019); Battle <i>et al.</i> (2019)
ndvi12	Index	Continuous	Normalised difference vegetation index	GLIMS; MODIS; Tucker <i>et al.</i> (2005); Pinzon and Tucker (2014); Didan and Barreto (2016)
pre12	Meters	Continuous	Precipitation	ERA5; Hersbach <i>et al.</i> (2020)
rgdp	US\$	Continuous	Real GDP of the country	WDI; World Bank (2022)
t2m12	Kelvin	Continuous	2 m surface temperature	ERA5; Hersbach <i>et al.</i> (2020)
l_ttcity	Hours	Continuous	Log of travel time to city	Malaria Atlas Project; Weiss <i>et al.</i> (2018)
l_ttmotor	Hours	Continuous	Log of travel time to healthcare facility by motorised vehicle	Malaria Atlas Project; Weiss <i>et al.</i> (2020)
l_ttwalk	Hours	Continuous	Log of travel time to healthcare facility by foot	Malaria Atlas Project; Weiss <i>et al.</i> (2020)
lon; lat	Degree	Continuous	Longitude and latitude coordinates	DHS; ICF (2004-2017)
<i>Discrete covariates</i>				
gender	"female"; "male"	Binary	Sex of the child	DHS; ICF (2004-2017)
d_conf25;	"yes" if $x \geq 5$;	Binary	Indicator of reported conflicts in the past within a buffer of 25/50/100 km	UCDP; Davies <i>et al.</i> (2022); Sundberg and Melander (2013)
d_conf50;	"no" if $x < 5$			
d_conf100				
d_n120_12	"yes" if $x \geq 5$;	Binary	Indicator of observed night-time light digital number values	Night-time light; Li <i>et al.</i> (2020)
	"no" if $x < 5$			
s_1c12	Classification	Categorical	Land-cover classification	MODIS Land Cover; Friedl <i>et al.</i> (2010); Friedl and Sulla-Menashe (2015)
s_soil	Classification	Categorical	Soil type classification	Soilgrids; Poggio <i>et al.</i> (2020)

some form of malnutrition (Black *et al.* 2008).

Previous research has identified several socio-demographic key factors that influence infant and child mortality. These include the sex of the child (see e.g., Harttgen *et al.* 2020; Kiross *et al.* 2021; Shobiye *et al.* 2022), which has been associated with survival in various contexts, and birth order (see e.g., Hobcraft *et al.* 1985), which potentially affects health outcomes through parental investment and sibling competition. Maternal age at birth is also critical, with both younger and older mothers facing higher risks due to biological and socioeconomic factors (see e.g., Harttgen *et al.* 2020; Shobiye *et al.* 2022). Parental education, especially of the mother, can be considered to serve as a protective factor by improving health knowledge, child care practices, and health care utilization (see e.g., Hanmer *et al.* 2003; Ekholuenetale *et al.* 2020; Shobiye *et al.* 2022). In addition, urban or rural residence was found to influence access to health care and living conditions, while household wealth likely influences nutrition, access to health care, and general living standards, all of which are closely linked to child mortality risk (see e.g., Harttgen *et al.* 2020; Ekholuenetale *et al.* 2020).

In addition to socio-demographic factors, the presence of armed conflict, along with environmental and climatic conditions, can provide valuable insights into explaining child mortality. For instance, children growing up near areas affected by armed conflict face a substantially higher risk of dying before their first and fifth birthdays (Wagner *et al.* 2018; Boerma *et al.* 2019). Furthermore, malaria, one of the leading infectious diseases in tropical regions, is strongly associated with both infant and child mortality (see e.g., Battle *et al.* 2019; Weiss *et al.* 2019). Malaria disproportionately affects young children, particularly in regions with high transmission rates, leading to an increased risk of death from severe complications of the disease. Moreover, improved modelling techniques and the growing availability of socio-economic, environmental, and climatic geo-spatial data have proven invaluable in explaining infant and child mortality. For example, data on factors such as accessibility, land cover, precipitation, surface temperature, and vegetation are particularly effective when high-resolution geo-spatial modelling is applied (see e.g., Golding *et al.* 2017; Burstein *et al.* 2019). These factors, including climatic and environmental conditions, can act as proxies for infectious diseases or indicators of development, making them essential for assessing child and infant survival.

Before presenting the full modelling framework and the results, we examine the covariate *age of the mother at birth* (`magebirth`) in more detail, which is found to be one of the most important predictors in our analysis (see Figure 16). This variable is used to illustrate how individual covariates can be explored and interpreted within our modelling approach. Specifically, we use it to discuss key analytical steps, including the structure of the survival data and the presentation of time-to-event outcomes. Starting with basic descriptive patterns in the data, we show in Section 4.3 how the estimated effects can be visualised and interpreted. This illustration is intended to make the full results easier to follow and to highlight the value of variable-specific analyses.

Figure 15 shows the distribution of `magebirth` on the left, where age groups with fewer than 500 observations are highlighted in red.

Based on our sample, most mothers give birth between the ages of 18 and 28, with a median age of 25. The number of observations declines steadily after age 22, with only 356 mothers aged 46 or older at the time of birth. On the right, the relative frequency of observed events (deaths) is displayed, which shows a contrasting pattern: relatively low mortality rates below

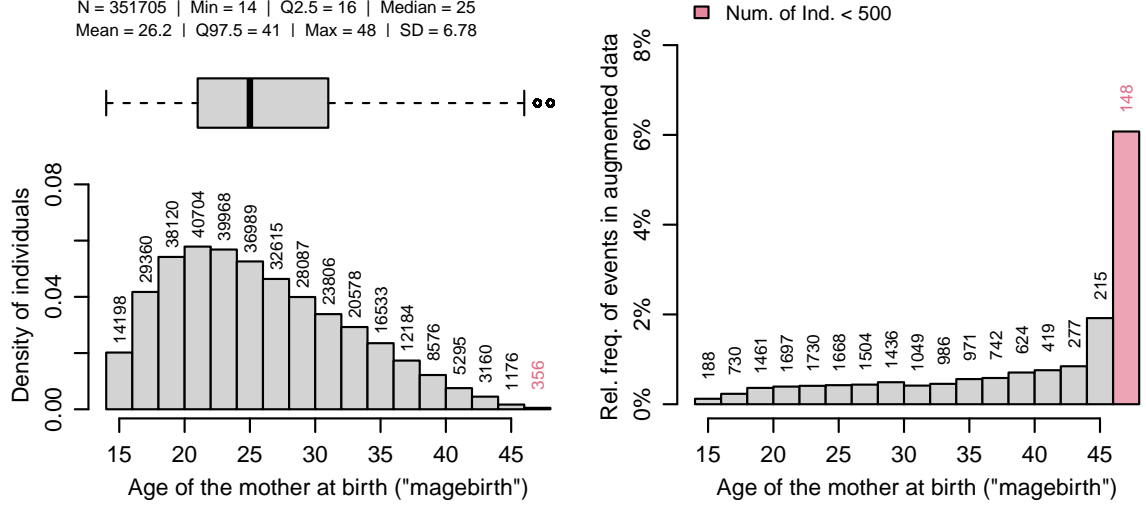


Figure 15: Distribution of the covariate *age of the mother at birth* (*magebirth*) on the left and relative frequency of events (deaths) in the augmented data on the right.

age 44, followed by a pronounced increase above that age, and a particularly high rate in the oldest age group. These insights are useful, when interpreting the estimated effects and other survival-specific outcomes in Section 4.3 below.

4.2. Model

The probability of death in month $t = 1, \dots, 11$ is modelled with a discrete time-to-event model. The additive predictor is defined as

$$\begin{aligned} \eta = & f_0(\text{age}) + f_1(\text{ai}) + f_2(\text{altitude}) + f_3(\text{bord}) + f_4(\text{l_distance}) + f_5(\text{iyyear}) + \\ & f_6(\text{hhs}) + f_7(\text{higheduyear}) + f_8(\text{magebirth}) + f_9(\text{minc12}) + f_{10}(\text{ndvi12}) + \\ & f_{11}(\text{pre12}) + f_{12}(\text{l_rgdp}) + f_{13}(\text{t2m12}) + f_{14}(\text{l_ttcity}) + f_{15}(\text{l_ttmotor}) + \\ & f_{16}(\text{l_ttwalk}) + f_{17}(\text{lon, lat}) + f_{18}(\text{gender}) + f_{19}(\text{d_conf25}) + f_{20}(\text{d_conf50}) + \\ & f_{21}(\text{d_conf100}) + f_{22}(\text{d_n120_12}) + f_{23}(\text{s_lc12}) + f_{24}(\text{s_soil}), \end{aligned}$$

where f_0, \dots, f_{24} can be a spline, a spatial or a random effect. The variables are described in more detail in the previous section and in Table 2.

The model is estimated using BBFIT, following the same specification approach as in the simulation study – which includes both boosting and resampling steps – but with an increased number of iterations. Specifically, $L = 1000$ iterations are used in the boosting step to select the most important covariates, and $L = 400$ iterations are used in the resampling step, with the first 200 discarded as burn-in.

4.3. Results

Figure 16 displays the relative updating frequencies of the boosting step on the left and the log-likelihood contribution on the right. It is the frequency with which a model term yields the best improvement of the out-of-sample log-likelihood by the total number of iterations.

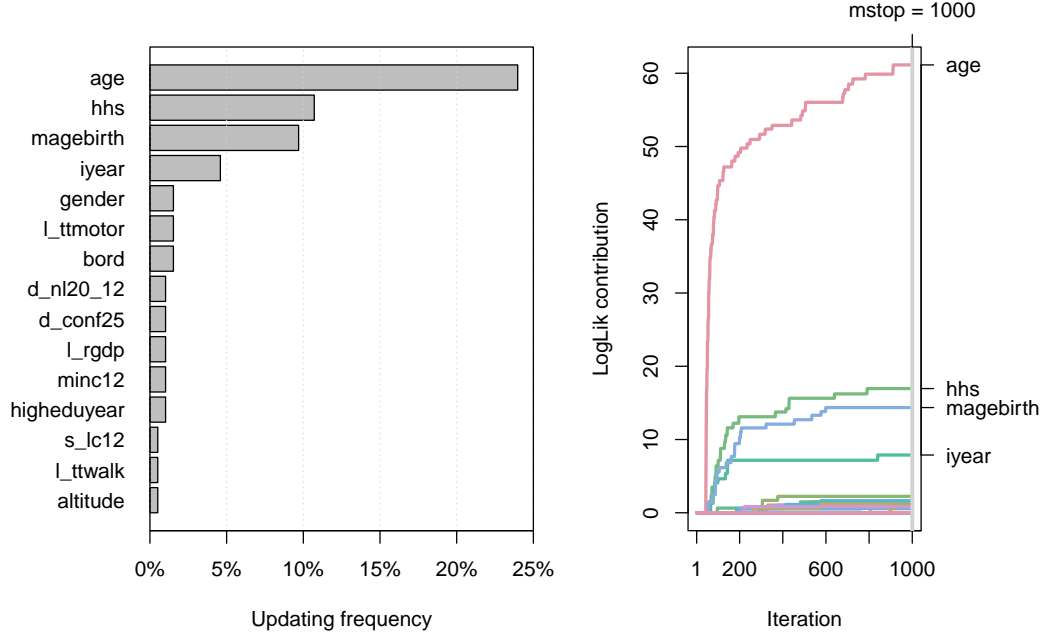


Figure 16: Updating frequencies based on boosting step of BBFIT (left) and log-likelihood contribution plot (right). Note that variables that have never been updated are omitted. See Table 2 for variable description.

We find that the age of the child (*age*), the household size (*hhs*), the age of the mother at birth (*magebirth*) and the interview year (*iyear*) are the most important explanatory variables in terms of updating frequency. The age of the child appears to have the by far the greatest importance based on these metrics. While these four variables stand out in terms of importance, the others contribute far less and are not examined further.

As discussed in Section 4.1, we focus here first on the covariate *age of the mother at birth* (*magebirth*) to illustrate the presentation of time-to-event outcomes. Figure 17 displays the estimated effect centred at zero.

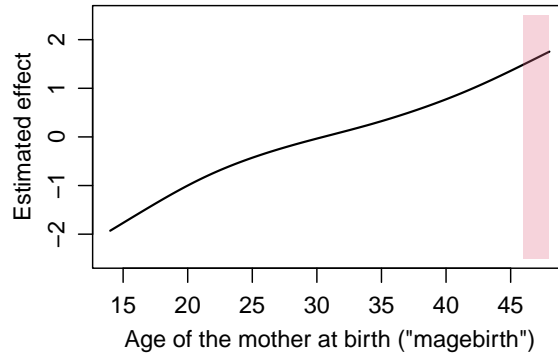


Figure 17: Estimated effects of the covariate *age of the mother at birth* (*magebirth*) centred at zero. The age groups with fewer than 500 observations are highlighted in red (see Figure 15).

We find a near-linear increase in mortality risk between ages 18 and 44. At the lower and

upper ends of the distribution, a slightly steeper increase in risk is visible. In general, this indicates that children with older mothers are more likely to die than children with younger mothers.

While this trend appears plausible, interpretation on the predictor scale is limited, as it does not directly translate to outcome probabilities and the estimated effects are centred at zero. Of primary interest is how covariates influence the survival probability, as defined in Equation (2). Figure 18 visualises this relationship for `magebirth`. All survival probabilities shown are marginal, i.e., all variables, except the one under consideration, are held fixed at their mean (continuous variables) or at the mode (discrete variables) level.

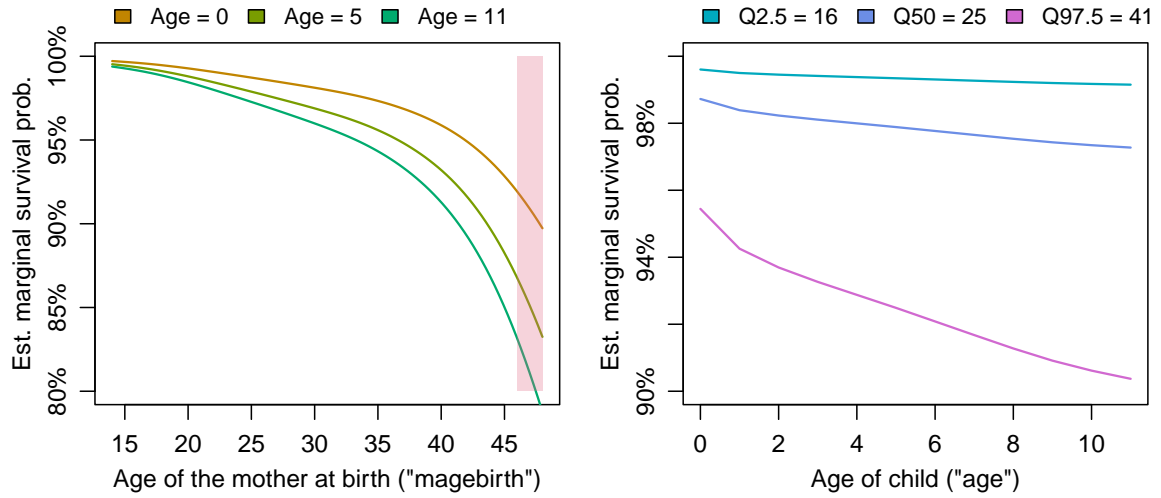


Figure 18: Estimated marginal survival curves for children aged 0, 5 and 11 months, across the age of the mother at birth on the left, and estimated marginal survival probabilities over child age for 16, 25 and 41 (2.5th, 50th, and 97.5th percentiles respectively) years of maternal age at birth. The age groups with fewer than 500 observations are highlighted in red (see Figure 15).

The left panel of Figure 18 shows estimated marginal survival curves for children up to an age of one month (Age = 0 corresponding to age interval $[0, 1)$), half a year (Age = 5), and one year (Age = 11) corresponding to infant mortality, as a function of the mother's age at birth. Survival probabilities are close to 100% for younger maternal ages but decline notably with increasing maternal age, particularly after age 35. The downward trend is more pronounced for older children, indicating that the impact of maternal age accumulates over time.

The right panel shows estimated marginal survival probabilities over child age for three selected maternal ages: 16 (Q2.5), 25 (Q50 or median) and 41 (Q97.5) years. Children born to younger mothers consistently show higher marginal survival probabilities across all ages, while children born to older mothers experience lower survival probabilities, with the gap widening as children age.

These findings underscore the value of variable-specific analyses in understanding complex patterns in survival outcomes. They highlight how flexible modelling approaches benefit from a careful combination of descriptive data exploration (as illustrated in Section 4.1) and meaningful interpretation of estimated effects. By examining `magebirth` in detail, we

demonstrate how individual covariates can be linked to outcome dynamics and how model results can be interpreted in a substantively relevant way. Equivalent analyses are provided for all covariates in Appendix B.1 Figures 24 to 33. Here we continue with a summary of the estimated effects and the estimated marginal probability of survival for the most important variables identified above – excluding further discussion of `magebirth`.

Figure 19 displays the estimated effects centred at zero. The estimated effect of the age of the child (`age`) decreases logarithmically over time. This indicates that the probability of dying in the first months of life is greater than later when the first months have been survived. The estimated effect of household size (`hhs`) follows a U-shape. The probability of dying is lower in households with 5 to 15 members and higher in households with fewer or more members. However, the effect is more pronounced in the range 2 to 4 household size. One interpretation could be that smaller households go through a learning phase or have limited childcare options due to single parents, while larger households may not have enough attention and resources available for individuals. The estimated effect of the year of interview (`iyear`) decreases more or less linearly, which is promising as it indicates a decreasing trend in the probability of death over time.

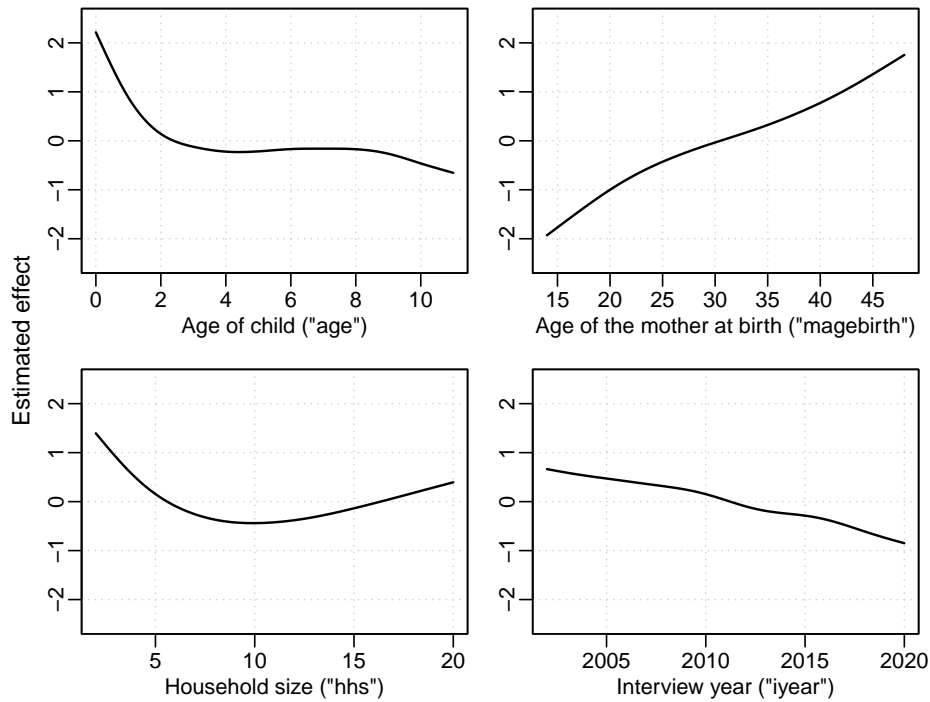


Figure 19: Estimated effects of the most important variables (see Figure 16) centred at zero.

As discussed with the covariate `magebirth` above, the centred effects are difficult to interpret meaningfully on the predictor scale. Of primary interest is how these factors influence the probability of survival. The estimated marginal probability of survival for the first year of life ($\text{Age} = 11$) is visualised in Figure 20.

The three effects (excluding `age` which is obviously used to construct the survival curves) – `magebirth`, `hhs`, and `iyear` – have a substantial influence on the marginal probability of survival. For example, children in small households with less than 5 members have

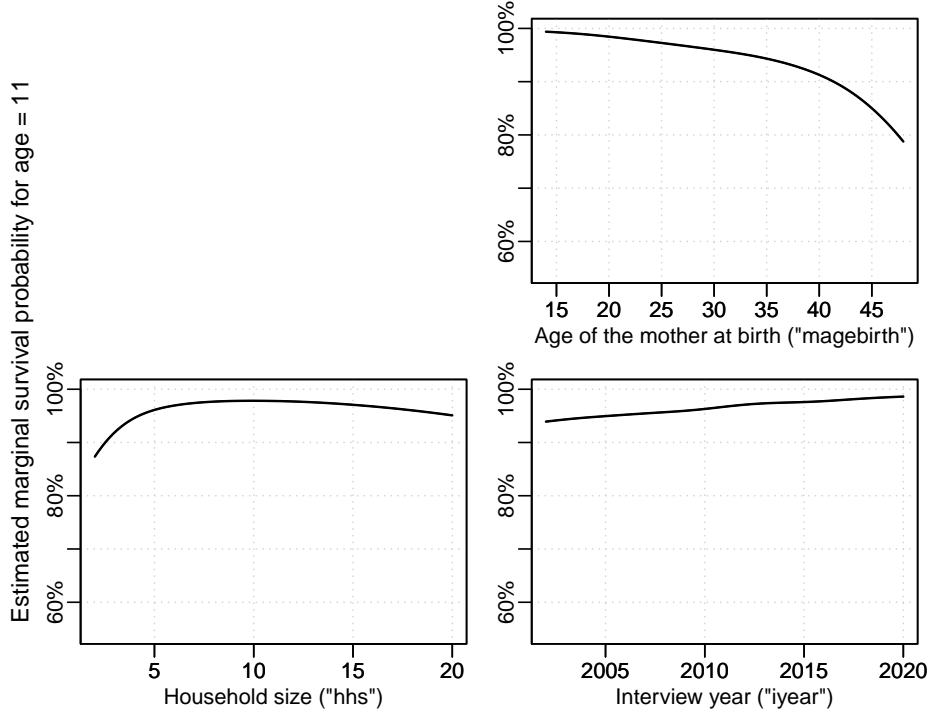


Figure 20: Estimated marginal survival probability for the first year of life (Age = 11) of the most important variables (see Figure 16).

a much lower marginal probability of surviving the first year of 90% to 95% compared to those in households with 8 to 10 members with around 97%. For households with more than 15 members, the marginal probability of surviving the first year decreases again to around 95%. Similar interpretations are possible for the other explanatory variables. The estimated effects and the marginal survival curves for all selected effects are shown in Appendix B.2 in Figure 34 and 35.

5. Conclusion

This paper introduces an efficient estimation framework for discrete time-to-event models with additive predictors by extending the recently proposed Batchwise Backfitting algorithm (BBFIT). Our contribution lies in combining scalable estimation with simultaneous variable selection in high-dimensional survival settings – addressing key limitations of existing approaches.

Through a comprehensive simulation study, we demonstrate that the proposed method (BBFIT) achieves high estimation accuracy, excellent variable selection performance, and significantly reduced computation times, even with data sets comprising up to ten million rows. Compared to benchmark methods such as GLM and BAM, BBFIT shows superior robustness and efficiency, especially in high-dimensional or large-scale contexts. In particular, BBFIT consistently avoids the selection of non-informative covariates, while BAM and GLM have a non-negligible false selection rate.

The application to modelling infant mortality in ten countries in eastern sub-Saharan Africa

further illustrates the practical utility of our approach. Despite the complexity of the data – both in terms of sample size and number of covariates – BBFIT identifies a small subset of influential factors (e.g., child’s age, household size, and survey year) and provides interpretable smooth effect estimates. This real-world example highlights how BBFIT can enhance predictive performance and yield meaningful insights in public health and development research, while also demonstrating its applicability to a broad range of real-world applications.

While the proposed framework is a powerful tool for large-scale discrete time-to-event modelling, several directions for future work remain. For instance, exploring alternative model structures (e.g., piecewise exponential models or Box-Cox-transformed hazard models) could expand its applicability beyond the current link and distributional assumptions. Another potential extension is the development of principled importance measures to quantify the relevance of selected predictors beyond selection frequencies or updating rates.

On the computational side, future work may focus on extending the algorithm towards a Bayesian framework to account for the model uncertainty. Furthermore, exploiting the specific structure of the likelihood – where all entries are zero except possibly the last – could lead to additional algorithmic efficiencies. These enhancements would enable more informative inference and further improve the scalability of the method for application to more complex settings in modern time-to-event analysis.

Disclosure statement

No potential conflict of interest was reported by the authors.

Acknowledgements

The computational results presented have been achieved (in part) using the HPC infrastructure LEO of the University of Innsbruck. We thank ICF International, Inc. and USAID for conducting the DHS (ICF 2004-2017), and providing public access to the data. This research was partially funded by the Austrian Science Fund (FWF) grant. doi: [10.55776/P33941](https://doi.org/10.55776/P33941).

Software

The computational results presented have been derived using the statistical software R (R Core Team 2025) using the following attached R packages: **bamlss** (Umlauf *et al.* 2018, 2021, 2024), **mgcv** (Wood 2003, 2011; Wood *et al.* 2016; Wood 2017), **MASS** (Venables and Ripley 2002), **discSurv** (Welchowski *et al.* 2022), **sf** (Pebesma 2018; Pebesma and Bivand 2023), **colorspace** (Stauffer *et al.* 2009; Zeileis *et al.* 2009, 2020), **lubridate** (Grolemund and Wickham 2011).

References

- Allison PD (1982). “Discrete-Time Methods for the Analysis of Event Histories.” *Sociological Methodology*, **13**, 61. ISSN 00811750. doi:10.2307/270718. URL <https://www.jstor.org/stable/270718?origin=crossref>.
- Amante C, Eakins B (2009). “ETOPO1 1 Arc-Minute Global Relief Model: procedures, Data Sources and Analysis.” Retrieved January 27, 2021, from NOAA Technical Memorandum NESDIS NGDC-24 doi: 10.7289/V5C8276M.
- Battle K, Lucas T, Nguyen M, Howes R, Nandi A, Twohig K, Pfeffer D, Cameron E, Rao P, Casey D, Gibson H, Rozier J, Dalrymple U, Keddie S, Collins E, Harris J, Guerra C, Thorn M, Bisanzio D, Fullman N, Huynh C, Kulikoff X, Kutz M, Lopez A, Mokdad A, Naghavi M, Nguyen G, Shackelford K, Vos T, Wang H, Lim S, Murray C, Price R, Baird J, Smith D, Bhatt S, Weiss D, Hay S, Gething P (2019). “Mapping the Global Endemicity and Clinical Burden of *Plasmodium vivax*, 2000-17: a Spatial and Temporal Modelling Study.” *Lancet*, **394**(10195), 332–343. doi: 10.1016/S0140-6736(19)31096-7.
- Belitz C, Lang S (2008). “Simultaneous selection of variables and smoothing parameters in structured additive regression models.” *Computational Statistics & Data Analysis*, **53**(1), 61–81. ISSN 01679473. doi:10.1016/j.csda.2008.05.032. URL <https://linkinghub.elsevier.com/retrieve/pii/S0167947308003009>.
- Berger M, Schmid M (2018). “Semiparametric regression for discrete time-to-event data.” *Statistical Modelling*, **18**(3-4), 322–345. ISSN 1471-082X, 1477-0342. doi:10.1177/1471082X17748084. URL <http://journals.sagepub.com/doi/10.1177/1471082X17748084>.
- Black RE, Allen LH, Bhutta ZA, Caulfield LE, Onis Md, Ezzati M, Mathers C, Rivera J (2008). “Maternal and child undernutrition: global and regional exposures and health consequences.” *The Lancet*, **371**(9608), 243–260. ISSN 0140-6736. doi:[https://doi.org/10.1016/S0140-6736\(07\)61690-0](https://doi.org/10.1016/S0140-6736(07)61690-0). URL <https://www.sciencedirect.com/science/article/pii/S0140673607616900>.
- Boerma T, Tappis H, Saad-Haddad G, Das J, Melesse DY, DeJong J, Spiegel P, Black R, Victora C, Bhutta ZA, Barros AJD (2019). “Armed conflicts and national trends in reproductive, maternal, newborn and child health in sub-Saharan Africa: what can national health surveys tell us?” *BMJ Global Health*, **4**(Suppl 4). doi: 110.1136/bmjgh-2018-001300.
- Brezger A, Lang S (2006). “Generalized structured additive regression based on Bayesian P-splines.” *Computational Statistics & Data Analysis*, **50**(4), 967–991. doi:10.1016/j.csda.2004.10.011. MAG ID: 2134080536.
- Burstein R, Henry NJ, Collison ML, Marczak LB, Sligar A, Watson S, Marquez N, Abbasalizad-Farhangi M, Abbasi M, Abd-Allah F, Abdoli A, Abdollahi M, Abdollahpour I, Abdulkader RS, Abrigo MRM, Acharya D, Adebayo OM, Adekanmbi V, Adham D, Afshari M, Aghaali M, Ahmadi K, Ahmadi M, Ahmadpour E, Ahmed R, Akal CG, Akinyemi JO, Alahdab F, Alam N, Alamene GM, Alene KA, Alijanzadeh M, Alinia C, Alipour V, Aljunid SM, Almalki MJ, Al-Mekhlafi HM, Altirkawi K, Alvis-Guzman N, Amegah AK, Amini S,

Amit AML, Anbari Z, Androudi S, Anjomshoa M, Ansari F, Antonio CAT, Arabloo J, Arefi Z, Aremu O, Armoon B, Arora A, Artaman A, Asadi A, Asadi-Aliabadi M, Ashraf-Ganjouei A, Assadi R, Ataeinia B, Atre SR, Quintanilla BPA, Ayanore MA, Azari S, Babae E, Babazadeh A, Badawi A, Bagheri S, Bagherzadeh M, Baheiraei N, Balouchi A, Barac A, Bassat Q, Baune BT, Bayati M, Bedi N, Beghi E, Behzadifar M, Behzadifar M, Belay YB, Bell B, Bell ML, Berbada DA, Bernstein RS, Bhattacharjee NV, Bhattarai S, Bhutta ZA, Bijani A, Bohlouli S, Breitborde NJK, Britton G, Browne AJ, Nagaraja SB, Busse R, Butt ZA, Car J, Cárdenas R, Castañeda-Orjuela CA, Cerin E, Chanie WF, Chatterjee P, Chu DT, Cooper C, Costa VM, Dalal K, Dandona L, Dandona R, Daoud F, Daryani A, Das Gupta R, Davis I, Davis Weaver N, Davitoiu DV, De Neve JW, Demeke FM, Demoz GT, Deribe K, Desai R, Deshpande A, Desyibelew HD, Dey S, Dharmaratne SD, Dhimal M, Diaz D, Doshmangir L, Duraes AR, Dwyer-Lindgren L, Earl L, Ebrahimi R, Ebrahimpour S, Effiong A, Eftekhari A, Ehsani-Chimeh E, El Sayed I, El Sayed Zaki M, El Tantawi M, El-Khatib Z, Emamian MH, Enany S, Eskandarieh S, Eyawo O, Ezalarab M, Faramarzi M, Fareed M, Faridnia R, Faro A, Fazaeli AA, Fazlzadeh M, Fentahun N, Fereshtehnejad SM, Fernandes JC, Filip I, Fischer F, Foigt NA, Foroutan M, Francis JM, Fukumoto T, Fullman N, Gallus S, Gebre DG, Gebrehiwot TT, Gebremeskel GG, Gessner BD, Geta B, Gething PW, Ghadimi R, Ghadiri K, Ghajarzadeh M, Ghashghae A, Gill PS, Gill TK, Golding N, Gomes NGM, Gona PN, Gopalani SV, Gorini G, Goulart BNG, Graetz N, Greaves F, Green MS, Guo Y, Haj-Mirzaian A, Haj-Mirzaian A, Hall BJ, Hamidi S, Haririan H, Haro JM, Hasankhani M, Hasanpoor E, Hasanzadeh A, Hassankhani H, Hassen HY, Hegazy MI, Hendrie D, Heydarpour F, Hird TR, Hoang CL, Hollerich G, Rad EH, Hoseini-Ghahfarokhi M, Hossain N, Hosseini M, Hosseinzadeh M, Hostiuc M, Hostiuc S, Househ M, Hsairi M, Ilesanmi OS, Imani-Nasab MH, Iqbal U, Irvani SSN, Islam N, Islam SMS, Jürisson M, Balalami NJ, Jalali A, Javidnia J, Jayatilleke AU, Jenabi E, Ji JS, Jobanputra YB, Johnson K, Jonas JB, Shushtari ZJ, Jozwiak JJ, Kabir A, Kahsay A, Kalani H, Kalhor R, Karami M, Karki S, Kasaeian A, Kassebaum NJ, Keiyoro PN, Kemp GR, Khabiri R, Khader YS, Khafaie MA, Khan EA, Khan J, Khan MS, Khang YH, Khatab K, Khater A, Khater MM, Khatony A, Khazaei M, Khazaei S, Khazaei-Pool M, Khubchandani J, Kianipour N, Kim YJ, Kimokoti RW, Kinyoki DK, Kisa A, Kisa S, Kolola T, Kosen S, Koul PA, Koyanagi A, Kraemer MUG, Krishan K, Krohn KJ, Kugbey N, Kumar GA, Kumar M, Kumar P, Kuupiel D, Lacey B, Lad SD, Lami FH, Larsson AO, Lee PH, Leili M, Levine AJ, Li S, Lim LL, Listl S, Longbottom J, Lopez JCF, Lorkowski S, Magdeldin S, Abd El Razek HM, Abd El Razek MM, Majeed A, Maleki A, Malekzadeh R, Malta DC, Mamun AA, Manafi N, Manda AL, Mansourian M, Martins-Melo FR, Masaka A, Massenburg BB, Maulik PK, Mayala BK, Mazidi M, McKee M, Mehrotra R, Mehta KM, Meles GG, Mendoza W, Menezes RG, Meretoja A, Meretoja TJ, Mestrovic T, Miller TR, Miller-Petrie MK, Mills EJ, Milne GJ, Mini GK, Mir SM, Mirjalali H, Mirrakhimov EM, Mohamadi E, Mohammad DK, Darwesh AM, Mezerji NMG, Mohammed AS, Mohammed S, Mokdad AH, Molokhia M, Monasta L, Moodley Y, Moosazadeh M, Moradi G, Moradi M, Moradi Y, Moradi-Lakeh M, Moradinazar M, Moraga P, Morawska L, Mosapour A, Mousavi SM, Mueller UO, Muluneh AG, Mustafa G, Nabavizadeh B, Naderi M, Nagarajan AJ, Nahvijou A, Najafi F, Nangia V, Ndwandwe DE, Neamati N, Negoi I, Negoi RI, Ngunjiri JW, Thi Nguyen HL, Nguyen LH, Nguyen SH, Nielsen KR, Ningrum DNA, Nirayo YL, Nixon MR, Nnaji CA, Nojomi M, Noroozi M, Nosratnejad S, Noubiap JJ, Motlagh SN, Ofori-Asenso R, Ogbo FA, Oladimeji KE, Olagunju AT, Olfatifar M, Olum S, Olusanya BO, Oluwasanu MM, Onwujekwe OE, Oren E, Ortega-Altamirano DDV, Ortiz A, Osarenotor O, Osei FB, Osgood-Zimmerman

- AE, Otstavnov SS, Owolabi MO, P A M, Pagheh AS, Pakhale S, Panda-Jonas S, Pandey A, Park EK, Parsian H, Pashaei T, Patel SK, Pepito VCF, Pereira A, Perkins S, Pickering BV, Pilgrim T, Pirestani M, Piroozi B, Pirsahab M, Plana-Ripoll O, Pourjafar H, Puri P, Qorbani M, Quintana H, Rabiee M, Rabiee N, Radfar A, Rafiei A, Rahim F, Rahimi Z, Rahimi-Movaghar V, Rahimzadeh S, Rajati F, Raju SB, Ramezankhani A, Ranabhat CL, Rasella D, Rashedi V, Rawal L, Reiner Jr RC, Renzaho AMN, Rezaei S, Rezapour A, Riahi SM, Ribeiro AI, Roever L, Roro EM, Roser M, Roshandel G, Roshani D, Rostami A, Rubagotti E, Rubino S, Sabour S, Sadat N, Sadeghi E, Saeedi R, Safari Y, Safari-Faramani R, Safdarian M, Sahebkar A, Salahshoor MR, Salam N, Salamati P, Salehi F, Zahabi SS, Salimi Y, Salimzadeh H, Salomon JA, Sambala EZ, Samy AM, Santric Milicevic MM, Jose BPS, Saraswathy SYI, Sarmiento-Suárez R, Sartorius B, Sathian B, Saxena S, Sbarra AN, Schaeffer LE, Schwebel DC, Sepanlou SG, Seyedmousavi S, Shaahmadi F, Shaikh MA, Shams-Beyranvand M, Shamshirian A, Shamsizadeh M, Sharafi K, Sharif M, Sharif-Alhoseini M, Sharifi H, Sharma J, Sharma R, Sheikh A, Shields C, Shigematsu M, Shiri R, Shiue I, Shuval K, Siddiqi TJ, Silva JP, Singh JA, Sinha DN, Sisay MM, Sisay S, Sliwa K, Smith DL, Somayaji R, Soofi M, Soriano JB, Sreeramareddy CT, Sudaryanto A, Sufiyan MB, Sykes BL, Sylaja PN, Tabarés-Seisdedos R, Tabb KM, Tabuchi T, Taveira N, Temsah MH, Terkawi AS, Tessema ZT, Thankappan KR, Thirunavukkarasu S, To QG, Tovani-Palone MR, Tran BX, Tran KB, Ullah I, Usman MS, Uthman OA, Vahedian-Azimi A, Valdez PR, Van Boven JFM, Vasankari TJ, Vasseghian Y, Veisani Y, Venketasubramanian N, Violante FS, Vladimirov SK, Vlassov V, Vos T, Vu GT, Vujcic IS, Waheed Y, Wakefield J, Wang H, Wang Y, Wang YP, Ward JL, Weintraub RG, Weldegewergs KG, Weldesamuel GT, Westerman R, Wiysonge CS, Wondafrash DZ, Woyczynski L, Wu AM, Xu G, Yadegar A, Yamada T, Yazdi-Feyzabadi V, Yilgwan CS, Yip P, Yonemoto N, Lebni JY, Younis MZ, Yousefifard M, Yousof HASA, Yu C, Yusefzadeh H, Zabeh E, Moghadam TZ, Bin Zaman S, Zamani M, Zandian H, Zangeneh A, Zerfu TA, Zhang Y, Ziapour A, Zodepey S, Murray CJL, Hay SI (2019). “Mapping 123 million neonatal, infant and child deaths between 2000 and 2017.” *Nature*, **574**(7778), 353–358. doi:10.1038/s41586-019-1545-0.
- Christodoulou SE (2011). “Water Network Assessment and Reliability Analysis by Use of Survival Analysis.” *Water Resources Management*, **25**(4), 1229–1238. ISSN 0920-4741, 1573-1650. doi:10.1007/s11269-010-9679-8. URL <http://link.springer.com/10.1007/s11269-010-9679-8>.
- Davies S, Pettersson T, Öberg M (2022). “Organized Violence 1989–2021 and Drone Warfare.” *Journal of Peace Research*, **59**(4), 593–610. doi: 10.1177/00223433221108428.
- Didan K, Barreto A (2016). “NASA MEaSUREs Vegetation Index and Phenology (VIP) Phenology EVI2 Yearly Global 0.05Deg CMG [Dataset].” doi: 10.5067/MEaSUREs/VIP/VIPPHEN_EVI2.004.
- Eilers PHC, Marx BD (1996). “Flexible smoothing with B-splines and penalties.” *Statistical Science*, **11**(2). ISSN 0883-4237. doi:10.1214/ss/1038425655. URL <https://projecteuclid.org/journals/statistical-science/volume-11/issue-2/Flexible-smoothing-with-B-splines-and-penalties/10.1214/ss/1038425655.full>.
- Ekholuenetale M, Wegbom AI, Tudeme G, Onikan A (2020). “Household factors associated with infant and under-five mortality in sub-Saharan Africa countries.” **14**(1), 10. doi: 10.1186/s40723-020-00075-1.

- Fahrmeir L, Kneib T, Lang S (2004). “Penalized Structured Additive Regression for Space-Time Data: A Bayesian Perspective.” *Statistica Sinica*, **14**(3), 731–761. ISSN 1017-0405. Publisher: Institute of Statistical Science, Academia Sinica, URL <https://www.jstor.org/stable/24307414>.
- Fahrmeir L, Kneib T, Lang S, Marx BD (2022). *Regression: Models, Methods and Applications*. Springer Berlin Heidelberg, Berlin, Heidelberg. ISBN 978-3-662-63881-1 978-3-662-63882-8. doi:10.1007/978-3-662-63882-8. URL <https://link.springer.com/10.1007/978-3-662-63882-8>.
- Fahrmeir L, Lang S (2001). “Bayesian Inference for Generalized Additive Mixed Models Based on Markov Random Field Priors.” *Journal of the Royal Statistical Society Series C: Applied Statistics*, **50**(2), 201–220. ISSN 0035-9254, 1467-9876. doi:10.1111/1467-9876.00229. URL <https://academic.oup.com/jrssc/article/50/2/201/6990749>.
- Fahrmeir L, Wagenpfeil S (1996). “Smoothing Hazard Functions and Time-Varying Effects in Discrete Duration and Competing Risks Models.” *Journal of the American Statistical Association*, **91**(436), 1584–1594. ISSN 0162-1459. doi:10.1080/01621459.1996.10476726. Publisher: ASA Website _eprint: <https://www.tandfonline.com/doi/pdf/10.1080/01621459.1996.10476726>, URL <https://www.tandfonline.com/doi/abs/10.1080/01621459.1996.10476726>.
- Friedl M, Sulla-Menashe D (2015). “MCD12C1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 0.05Deg CMG V006 [Dataset].” Retrieved July 07, 2022, from doi:10.5067/MODIS/MCD12C1.006.
- Friedl MA, Sulla-Menashe D, Tan B, Schneider A, Ramankutty N, Sibley A, Huang X (2010). “MODIS Collection 5 global land cover: algorithm refinements and characterization of new datasets.” *Remote Sensing of Environment*, **114**(1), 168–182. doi:10.1016/j.rse.2009.08.016.
- Golding N, Burstein R, Longbottom J, Browne AJ, Fullman N, Osgood-Zimmerman A, Earl L, Bhatt S, Cameron E, Casey DC, Dwyer-Lindgren L, Farag TH, Flaxman AD, Fraser MS, Gething PW, Gibson HS, Graetz N, Krause LK, Kulikoff XR, Lim SS, Mappin B, Morozoff C, Reiner Robert C J, Sligar A, Smith DL, Wang H, Weiss DJ, Murray CJL, Moyes CL, Hay SI (2017). “Mapping Under-5 and Neonatal Mortality in Africa, 2000-15: a Baseline Analysis for the Sustainable Development Goals.” *Lancet*, **390**(10108), 2171–2182. Publisher: Elsevier.
- Grolemund G, Wickham H (2011). “Dates and Times Made Easy with lubridate.” *Journal of Statistical Software*, **40**(3), 1–25. URL <https://www.jstatsoft.org/v40/i03/>.
- Hanmer L, Lensink R, White H (2003). “Infant and child mortality in developing countries: Analysing the data for Robust determinants.” *The Journal of Development Studies*, **40**(1), 101–118. doi:10.1080/00220380412331293687.
- Harttgen K, Lang S, Santer J, et al. (2020). “Modelling Under-five Mortality through Multilevel Structured Additive Regression with Varying Coefficients for Asia and sub-Saharan Africa.” *The Journal of Development Studies*, **56**(2), 401–430. doi:10.1080/00220388.2018.1563681.

- Hastie T, Tibshirani R (1986). “Generalized Additive Models.” *Statistical Science*, **1**(3). doi:10.1214/ss/1177013604.
- Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, Nicolas J, Peubey C, Radu R, Schepers D, Simmons A, Soci C, Abdalla S, Abellan X, Balsamo G, Bechtold P, Biavati G, Bidlot J, Bonavita M, De Chiara G, Dahlgren P, Dee D, Diamantakis M, Dragani R, Flemming J, Forbes R, Fuentes M, Geer A, Haimberger L, Healy S, Hogan RJ, Hólm E, Janisková M, Keeley S, Laloyaux P, Lopez P, Lupu C, Radnoti G, de Rosnay P, Rozum I, Vamborg F, Villaume S, Thépaut JN (2020). “The ERA5 Global Reanalysis.” *Quarterly Journal of the Royal Meteorological Society*, **146**(730), 1999–2049. doi: 10.1002/qj.3803.
- Hobcraft J, McDonald J, Rutstein S (1985). “Demographic Determinants of Infant and Early Child Mortality: A Comparative Analysis.” *Population Studies*, **39**(3), 363–385. doi: 10.1080/0032472031000141576.
- ICF (2004-2017). *Demographic and Health Surveys (various) [Datasets]*. Rockville, Maryland, USA. Funded by USAID.
- Kiross GT, Chojenta C, Barker D, Loxton D (2021). “Individual-, household- and community-level determinants of infant mortality in Ethiopia.” *PLOS ONE*, **16**(3), 1–19. doi: 10.1371/journal.pone.0248501.
- Lang S, Umlauf N, Wechselberger P, Harttgen K, Kneib T (2014). “Multilevel structured additive regression.” *Statistics and Computing*, **24**(2), 223–238. ISSN 0960-3174, 1573-1375. doi:10.1007/s11222-012-9366-0. URL <http://link.springer.com/10.1007/s11222-012-9366-0>.
- Li X, Zhou Y, Zhao M, Zhao X (2020). “A Harmonized Global Nighttime Light Dataset 1992-2018.” *Scientific Data*, **7**(1), 168. doi: 10.1038/s41597-020-0510-y.
- Musick K, Micheltore K (2015). “Change in the Stability of Marital and Cohabiting Unions Following the Birth of a Child.” *Demography*, **52**(5), 1463–1485. ISSN 0070-3370, 1533-7790. doi:10.1007/s13524-015-0425-y. URL <https://read.dukeupress.edu/demography/article/52/5/1463/169462/Change-in-the-Stability-of-Marital-and-Cohabiting>.
- NASA JPL (2013). “NASA Shuttle Radar Topography Mission Water Body Data Shapefiles & Raster Files [Dataset].” Retrieved May 25, 2021, from doi: 10.5067/MEASUREs/SRTM/SRTMSWBD.003.
- Neal RM (2003). “Slice sampling.” *The Annals of Statistics*, **31**(3). ISSN 0090-5364. doi:10.1214/aos/1056562461. URL <https://projecteuclid.org/journals/annals-of-statistics/volume-31/issue-3/Slice-sampling/10.1214/aos/1056562461.full>.
- NOAA National Geophysical Data Center (2009). “ETOPO1 1 Arc-Minute Global Relief Model [Dataset].” Retrieved January 27, 2021, from <https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ngdc.mgg.dem:316>.
- Pebesma E (2018). “Simple Features for R: Standardized Support for Spatial Vector Data.” *The R Journal*, **10**(1), 439–446. doi:10.32614/RJ-2018-009. URL <https://doi.org/10.32614/RJ-2018-009>.

- Pebesma E, Bivand R (2023). *Spatial Data Science: With applications in R*. Chapman and Hall/CRC. doi:10.1201/9780429459016. URL <https://r-spatial.org/book/>.
- Pinzon JE, Tucker CJ (2014). “A Non-Stationary 1981–2012 AVHRR NDVI_{3g} Time Series.” *Remote Sensing*, **6**(8), 6929–6960. doi: 10.3390/rs6086929.
- Poggio L, de Sousa LM, Batjes NH, Heuvelink GBM, Kempen B, Ribeiro E, Rossiter D (2020). “SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty [Dataset].” *SOIL*, **7**(1), 217–240. doi: 10.5194/soil-7-217-2021.
- R Core Team (2025). “R: A language and environment for statistical computing.” <https://www.R-project.org/>.
- Puth MT, Tutz G, Heim N, Münster E, Schmid M, Berger M (2020). “Tree-based modeling of time-varying coefficients in discrete time-to-event models.” *Lifetime Data Analysis*, **26**(3), 545–572. ISSN 1380-7870, 1572-9249. doi:10.1007/s10985-019-09489-7. URL <http://link.springer.com/10.1007/s10985-019-09489-7>.
- Rutstein SO, Rojas G (2006). “Guide to DHS Statistics: Demographic and Health Surveys Methodology.” *Technical report*, ORC Macro: Calverton, Maryland. url: .
- Schmid M, Küchenhoff H, Hoerauf A, Tutz G (2016). “A survival tree method for the analysis of discrete event times in clinical and epidemiological studies.” *Statistics in Medicine*, **35**(5), 734–751. ISSN 0277-6715, 1097-0258. doi:10.1002/sim.6729. URL <https://onlinelibrary.wiley.com/doi/10.1002/sim.6729>.
- Seiler J, Wetscher M, Harttgen K, Utzinger J, Umlauf N (2025a). “High-resolution spatial prediction of anemia risk among children aged 6 to 59 months in low- and middle-income countries.” *Communications Medicine*, **5**(1), 57. ISSN 2730-664X. doi:10.1038/s43856-025-00765-2. URL <https://www.nature.com/articles/s43856-025-00765-2>.
- Seiler J, Wetscher M, Harttgen K, Utzinger J, Umlauf N (2025b). “Supplementary information for "High-resolution spatial prediction of anemia risk among children aged 6 to 59 months in low- and middle-income countries"." *Communications Medicine*. URL https://static-content.springer.com/esm/art%3A10.1038/s43856-025-00765-2/MediaObjects/43856_2025_765_MOESM2_ESM.pdf.
- Shobiye DM, Omotola A, Zhao Y, Zhang J, Ekawati FM, Shobiye HO (2022). “Infant mortality and risk factors in Nigeria in 2013–2017: A population-level study.” *eClinicalMedicine*, **51**. doi: 10.1016/j.eclinm.2022.101622.
- Spuck N, Schmid M, Heim N, Klarman-Schulz U, Hörauf A, Berger M (2023). “Flexible tree-structured regression models for discrete event times.” *Statistics and Computing*, **33**(1), 20. ISSN 0960-3174, 1573-1375. doi:10.1007/s11222-022-10196-x. URL <https://link.springer.com/10.1007/s11222-022-10196-x>.
- Stauffer R, Mayr GJ, Dabernig M, Zeileis A (2009). “Somewhere over the Rainbow: How to Make Effective Use of Colors in Meteorological Visualizations.” *Bulletin of the American Meteorological Society*, **96**(2), 203–216. doi:10.1175/BAMS-D-13-00155.1.
- Sundberg R, Melander E (2013). “Introducing the UCDP Georeferenced Event Dataset.” *Journal of Peace Research*, **50**(4), 523–532. doi: 10.1177/0022343313484347.

- Tucker CJ, Pinzon JE, Brown ME, Slayback DA, Pak EW, Mahoney R, Vermote EF, Saleous NE (2005). “An Extended AVHRR 8-km NDVI Dataset Compatible with MODIS and SPOT Vegetation NDVI Data.” *International Journal of Remote Sensing*, **26**(20), 4485–4498. doi: [10.1080/01431160500168686](https://doi.org/10.1080/01431160500168686).
- Tutz G, Binder H (2004). “Flexible modelling of discrete failure time including time-varying smooth effects.” *Statistics in Medicine*, **23**(15), 2445–2461. ISSN 1097-0258. doi:[10.1002/sim.1824](https://doi.org/10.1002/sim.1824). eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.1824>, URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.1824>.
- Tutz G, Schmid M (2016). *Modeling Discrete Time-to-Event Data*. Springer Series in Statistics. Springer International Publishing, Cham. ISBN 978-3-319-28156-8 978-3-319-28158-2. doi:[10.1007/978-3-319-28158-2](https://doi.org/10.1007/978-3-319-28158-2). URL <http://link.springer.com/10.1007/978-3-319-28158-2>.
- Umlauf N, Klein N, Simon T, Zeileis A (2021). “bamlss : A Lego Toolbox for Flexible Bayesian Regression (and Beyond).” *Journal of Statistical Software*, **100**(4). doi:[10.18637/jss.v100.i04](https://doi.org/10.18637/jss.v100.i04).
- Umlauf N, Klein N, Zeileis A (2018). “BAMLSS: Bayesian Additive Models for Location, Scale, and Shape (and Beyond).” *Journal of Computational and Graphical Statistics*, **27**(3), 612–627. doi:[10.1080/10618600.2017.1407325](https://doi.org/10.1080/10618600.2017.1407325).
- Umlauf N, Seiler J, Wetscher M, Simon T, Lang S, Klein N (2024). “Scalable Estimation for Structured Additive Distributional Regression.” *Journal of Computational and Graphical Statistics*, pp. 1–23. ISSN 1061-8600, 1537-2715. doi:[10.1080/10618600.2024.2388604](https://doi.org/10.1080/10618600.2024.2388604). URL <https://www.tandfonline.com/doi/full/10.1080/10618600.2024.2388604>.
- UNICEF (2020). “UNICEF Conceptual Framework on Maternal and Child Nutrition.” URL <https://www.unicef.org/documents/conceptual-framework-nutrition>.
- United Nations (2015). “Sustainable Development Goal 3: Ensure healthy lives and promote well-being for all at all ages.” URL <https://sdgs.un.org/goals/goal3>.
- United Nations Children’s Fund (2025). “United Nations Inter-agency Group for Child Mortality Estimation (UN IGME), Levels & Trends in Child Mortality: Report 2024, Estimates developed by the United Nations Inter-agency Group for Child Mortality Estimation.” Place: New York, USA.
- United Nations Statistics Division (2025). “Standard Country or Area Codes for Statistical Use (M49).” URL <https://unstats.un.org/unsd/methodology/m49/>.
- Venables WN, Ripley BD (2002). *Modern Applied Statistics with S*. Statistics and Computing. Springer New York, New York, NY. ISBN 978-1-4419-3008-8 978-0-387-21706-2. doi:[10.1007/978-0-387-21706-2](https://doi.org/10.1007/978-0-387-21706-2). URL <http://link.springer.com/10.1007/978-0-387-21706-2>.
- Wagner Z, Heft-Neal S, Bhutta ZA, Black RE, Burke M, Bendavid E (2018). “Armed conflict and child mortality in Africa: a geospatial analysis.” *The Lancet*, **392**(10150), 857–865. doi: [10.1016/S0140-6736\(18\)31437-5](https://doi.org/10.1016/S0140-6736(18)31437-5).

- Weiss DJ, Lucas TCD, Nguyen M, Nandi AK, Bisanzio D, Battle KE, Cameron E, Twohig KA, Pfeffer DA, Rozier JA, Gibson HS, Rao PC, Casey D, Bertozzi-Villa A, Collins EL, Dalrymple U, Gray N, Harris JR, Howes RE, Kang SY, Keddie SH, May D, Rumisha S, Thorn MP, Barber R, Fullman N, Huynh CK, Kulikoff X, Kutz MJ, Lopez AD, Mokdad AH, Naghavi M, Nguyen G, Shackelford KA, Vos T, Wang H, Smith DL, Lim SS, Murray CJL, Bhatt S, Hay SI, Gething PW (2019). “Mapping the Global Prevalence, Incidence, and Mortality of *Plasmodium falciparum*, 2000-17: a Spatial and Temporal Modelling Study.” *Lancet*, **394**(10195), 322–331. doi: [10.1016/S0140-6736\(19\)31097-9](https://doi.org/10.1016/S0140-6736(19)31097-9).
- Weiss DJ, Nelson A, Gibson HS, Temperley W, Peedell S, Lieber A, Hancher M, Poyart E, Belchior S, Fullman N, Mappin B, Dalrymple U, Rozier J, Lucas TCD, Howes RE, Tusting LS, Kang SY, Cameron E, Bisanzio D, Battle KE, Bhatt S, Gething PW (2018). “A Global Map of Travel Time to Cities to Assess Inequalities in Accessibility in 2015.” *Nature*, **553**, 333–336. doi: [10.1038/nature25181](https://doi.org/10.1038/nature25181).
- Weiss DJ, Nelson A, Vargas-Ruiz CA, Gligorić K, Bavadekar S, Gabrilovich E, Bertozzi-Villa A, Rozier J, Gibson HS, Shekel T, Kamath C, Lieber A, Schulman K, Shao Y, Qarkaxhija V, Nandi AK, Keddie SH, Rumisha S, Amratia P, Arambepola R, Chestnutt EG, Millar JJ, Symons TL, Cameron E, Battle KE, Bhatt S, Gething PW (2020). “Global Maps of Travel Time to Healthcare Facilities.” *Nature Medicine*, **26**(12), 1835–1838. doi: [10.1038/s41591-020-1059-1](https://doi.org/10.1038/s41591-020-1059-1).
- Welchowski T, Berger M, Koehler D, Schmid M (2022). *discSurv: Discrete Time Survival Analysis*. R package version 2.0.0, URL <https://CRAN.R-project.org/package=discSurv>.
- Wood SN (2003). “Thin Plate Regression Splines.” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **65**(1), 95–114. ISSN 1369-7412, 1467-9868. doi:[10.1111/1467-9868.00374](https://doi.org/10.1111/1467-9868.00374). URL <https://academic.oup.com/jrsssb/article/65/1/95/7110632>.
- Wood SN (2004). “Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models.” *Journal of the American Statistical Association*, **99**(467), 673–686. doi:[10.1198/016214504000000980](https://doi.org/10.1198/016214504000000980). MAG ID: 2009499611.
- Wood SN (2006). “Low-Rank Scale-Invariant Tensor Product Smooths for Generalized Additive Mixed Models.” *Biometrics*, **62**(4), 1025–1036. doi:[10.1111/j.1541-0420.2006.00574.x](https://doi.org/10.1111/j.1541-0420.2006.00574.x). MAG ID: 1967247922.
- Wood SN (2011). “Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models.” *Journal of The Royal Statistical Society Series B-statistical Methodology*, **73**(1), 3–36. doi:[10.1111/j.1467-9868.2010.00749.x](https://doi.org/10.1111/j.1467-9868.2010.00749.x). MAG ID: 2119160928.
- Wood SN (2017). *Generalized Additive Models: An Introduction with R*. 2 edition. Chapman and Hall/CRC. ISBN 978-1-315-37027-9. doi:[10.1201/9781315370279](https://doi.org/10.1201/9781315370279). URL <https://www.taylorfrancis.com/books/9781498728348>.
- Wood SN, Goude Y, Shaw S (2015). “Generalized Additive Models for Large Data Sets.” *Journal of the Royal Statistical Society Series C: Applied Statistics*, **64**(1), 139–155. ISSN 0035-

- 9254, 1467-9876. doi:10.1111/rssc.12068. URL <https://academic.oup.com/jrsssc/article/64/1/139/7067572>.
- Wood SN, Li Z, Shaddick G, Augustin NH (2017). “Generalized Additive Models for Gigadata: modelling the UK Black Smoke Network Daily Data.” *Journal of the American Statistical Association*, **112**(519), 1199–1210.
- Wood SN, Pya N, Säfken B (2016). “Smoothing Parameter and Model Selection for General Smooth Models.” *Journal of the American Statistical Association*, **111**(516), 1548–1563. ISSN 0162-1459, 1537-274X. doi:10.1080/01621459.2016.1180986. URL <https://www.tandfonline.com/doi/full/10.1080/01621459.2016.1180986>.
- World Bank (2022). “World Development Indicators 2022.” Retrieved November 04, 2022 <https://data.worldbank.org/>.
- Zeileis A, Fisher JC, Hornik K, Ihaka R, McWhite CD, Murrell P, Stauffer R, Wilke CO (2020). “colorspace: A Toolbox for Manipulating and Assessing Colors and Palettes.” *Journal of Statistical Software*, **96**(1), 1–49. doi:10.18637/jss.v096.i01.
- Zeileis A, Hornik K, Murrell P (2009). “Escaping RGBland: Selecting Colors for Statistical Graphics.” *Computational Statistics & Data Analysis*, **53**(9), 3259–3270. doi:10.1016/j.csda.2008.11.033.

A. Appendix: Simulation results

A.1. Basic setting: Noise/uninformative variables

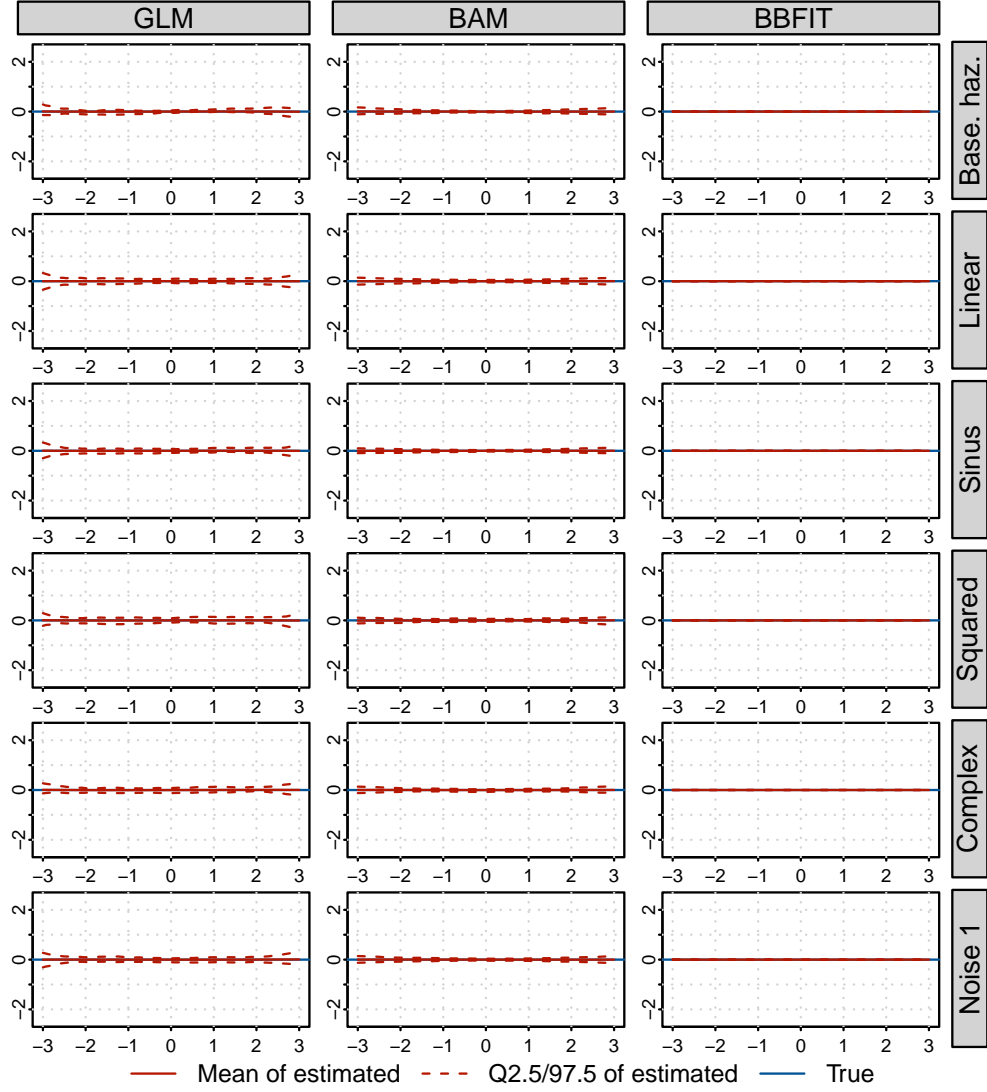


Figure 21: Estimated effects of all noise/uninformative variables in the basic setting (see Section 3.1).

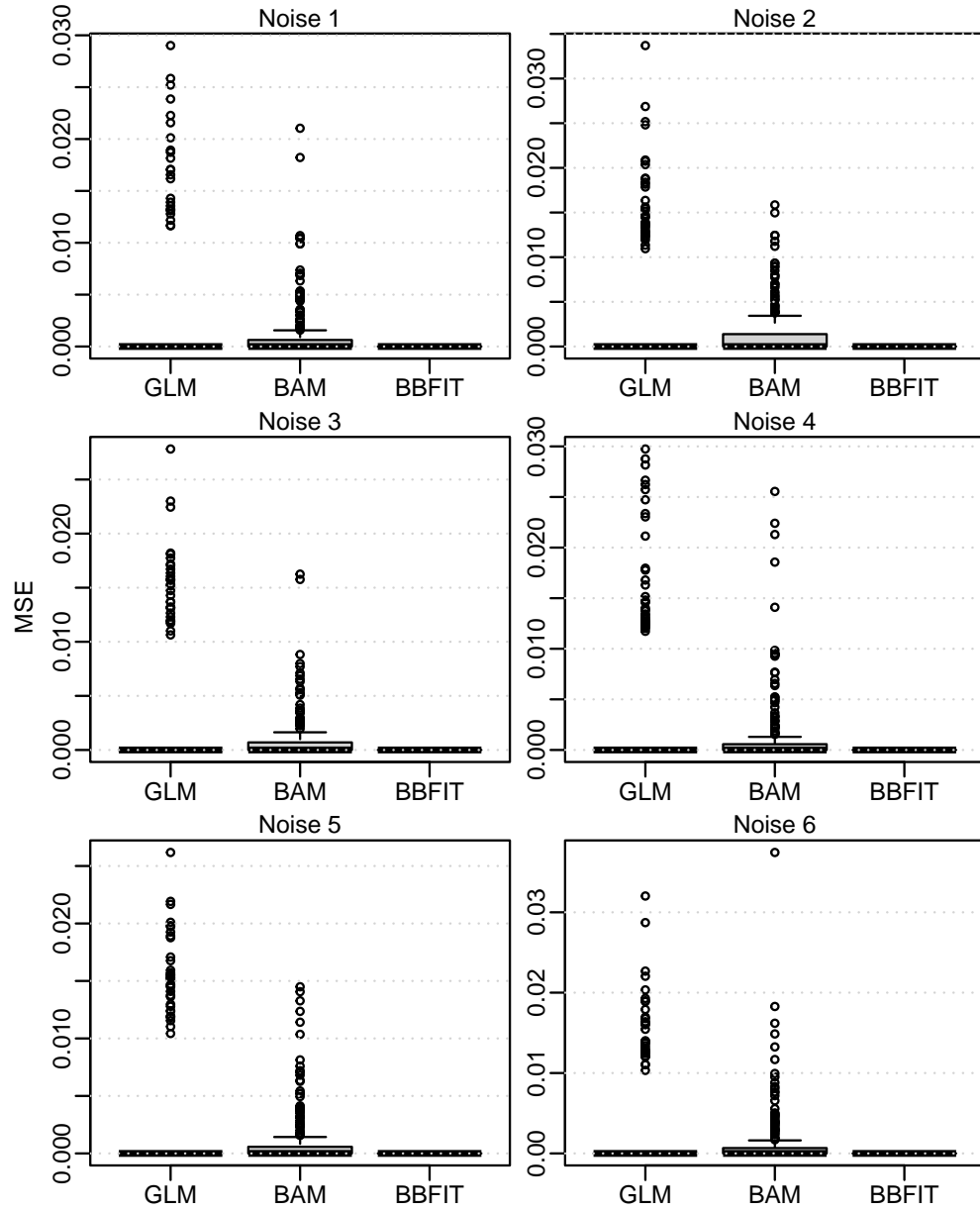


Figure 22: MSE of all noise/uninformative variables in the basic setting (see Section 3.1).

A.2. Further settings: Mean MSE by number of individuals

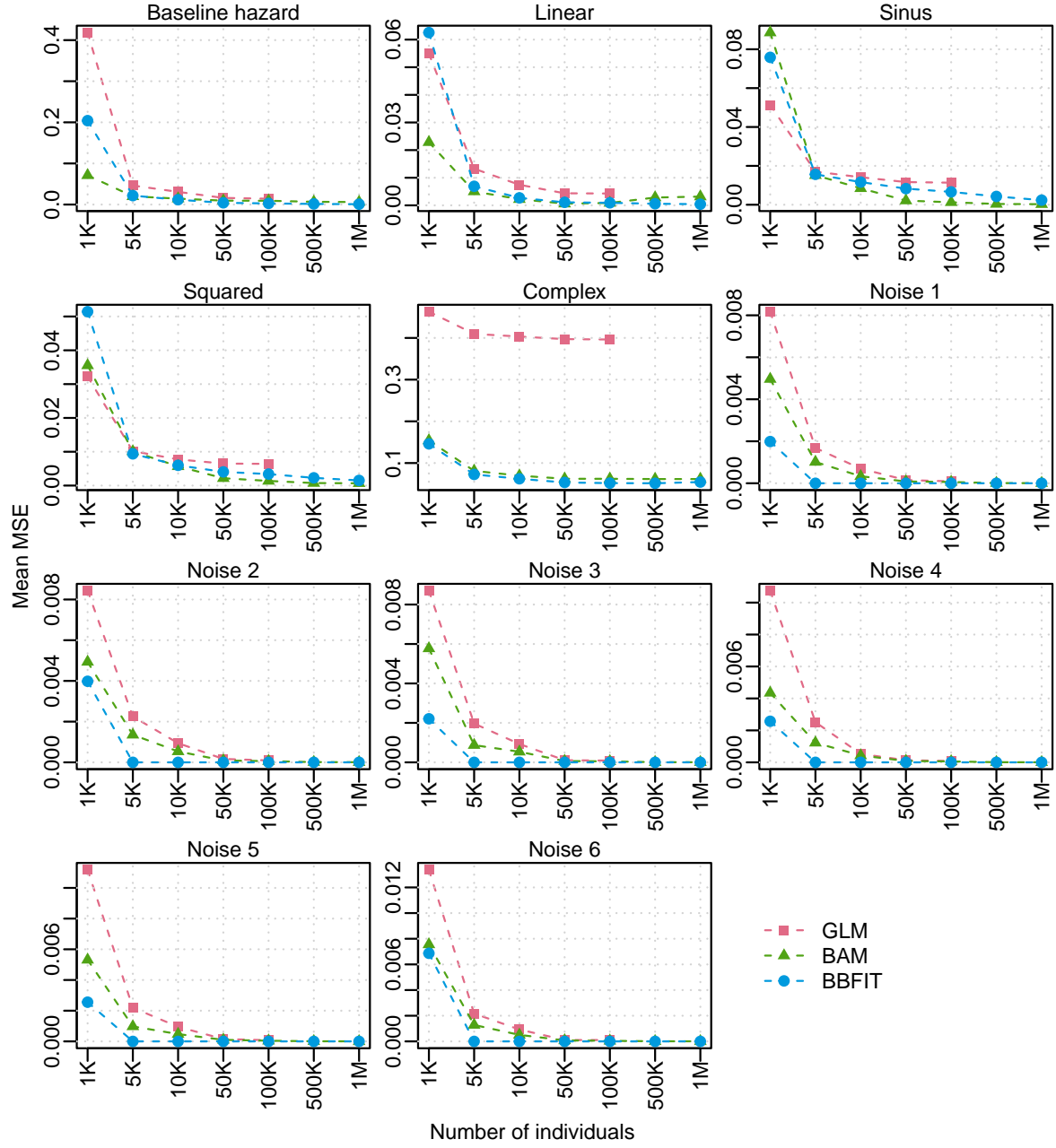
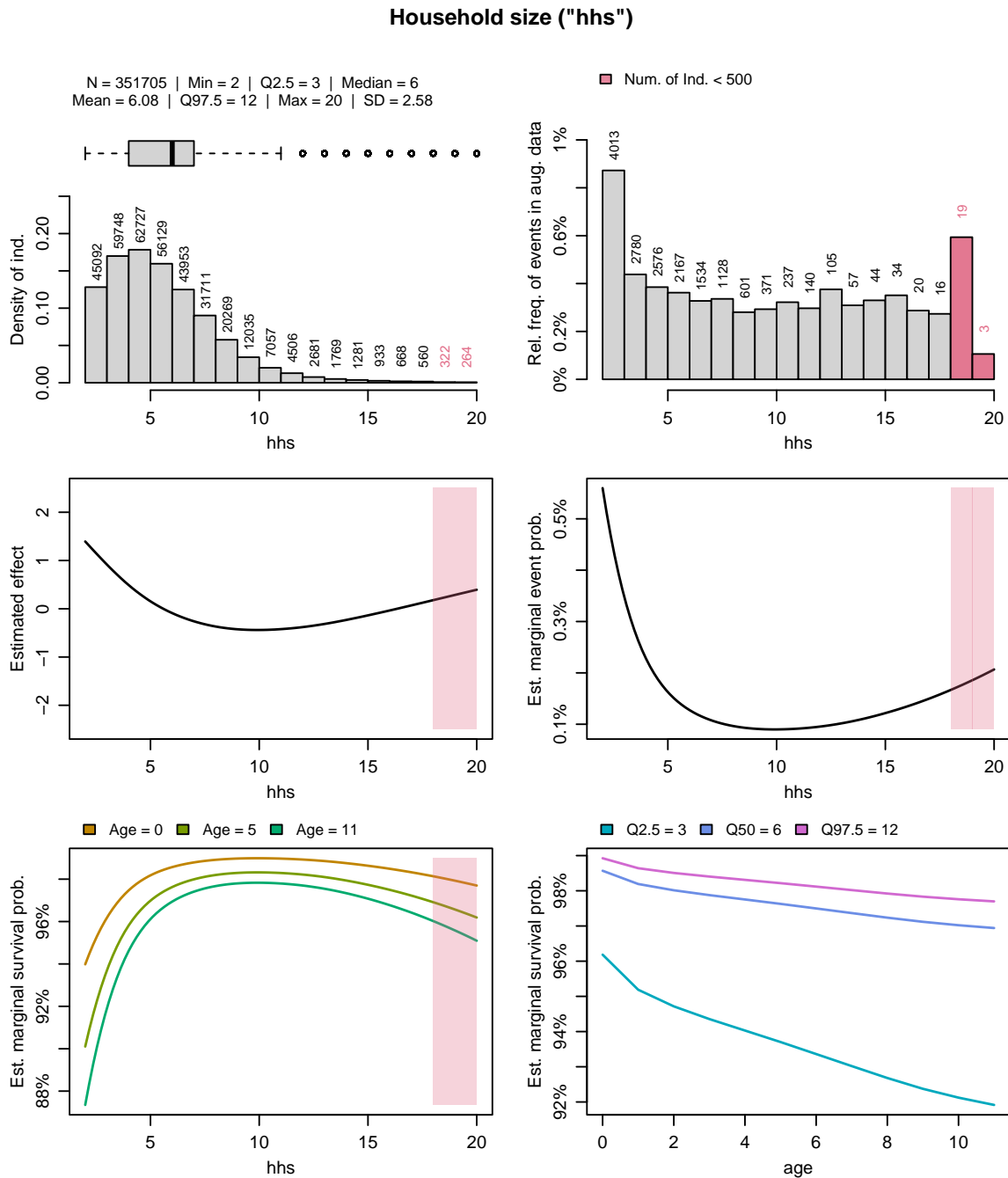
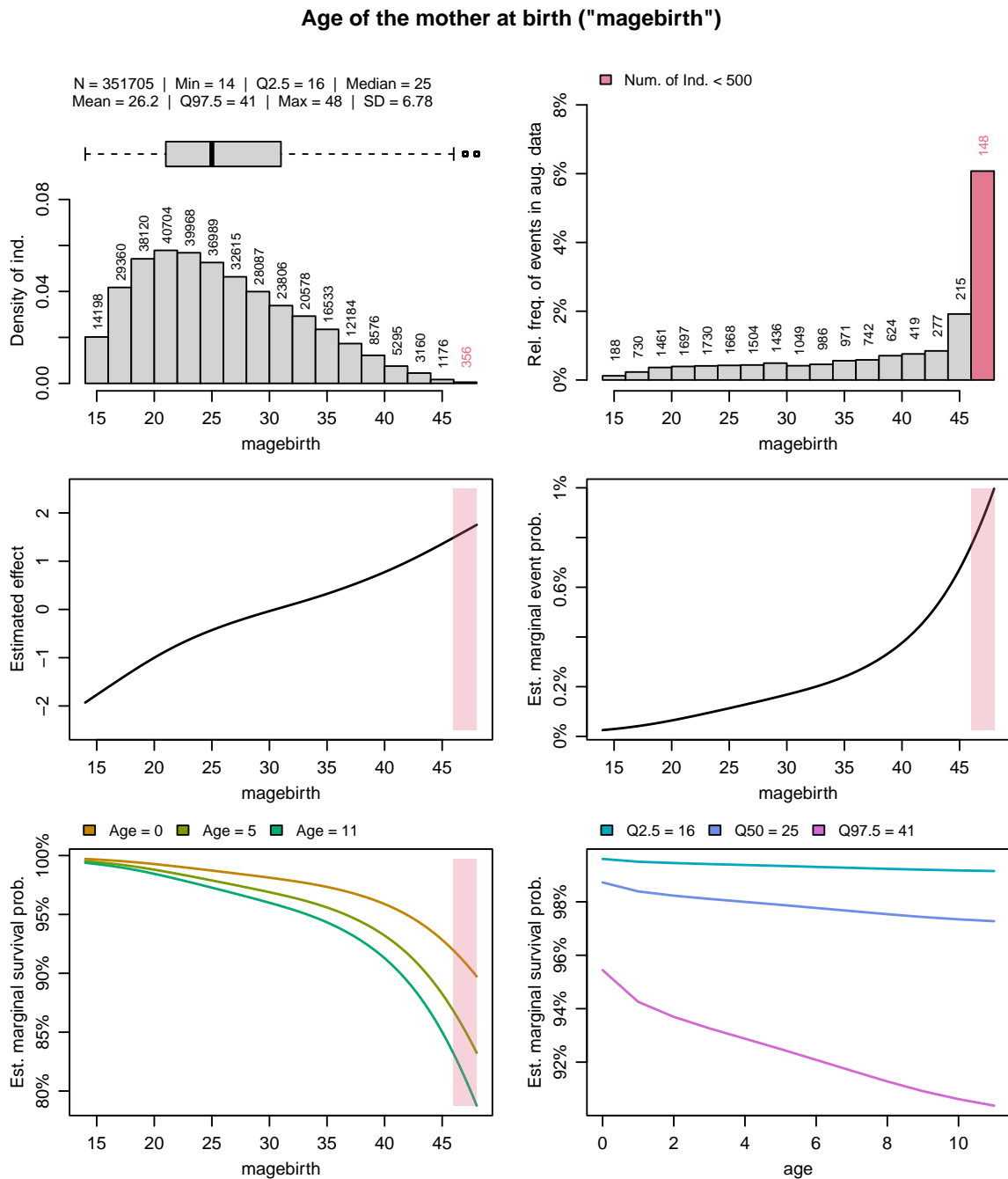


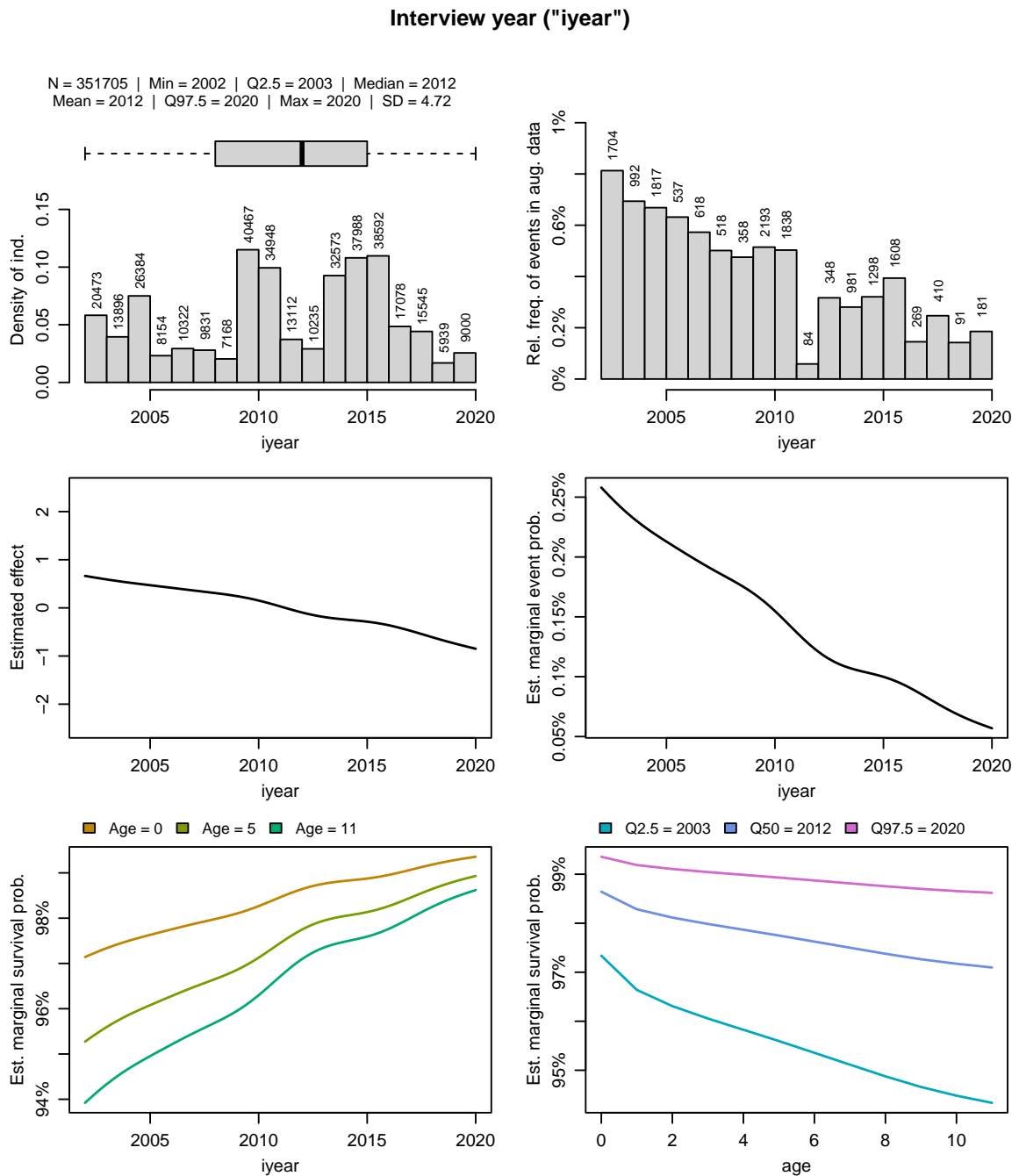
Figure 23: Mean MSE for different number of individuals of all variables.

B. Appendix: Application results

B.1. Single variable analysis

Figure 24: Single variable analysis of covariate *household size* (hhs).

Figure 25: Single variable analysis of covariate *age of the mother at birth* (magebirth).

Figure 26: Single variable analysis of covariate *interview year* (*iyear*).

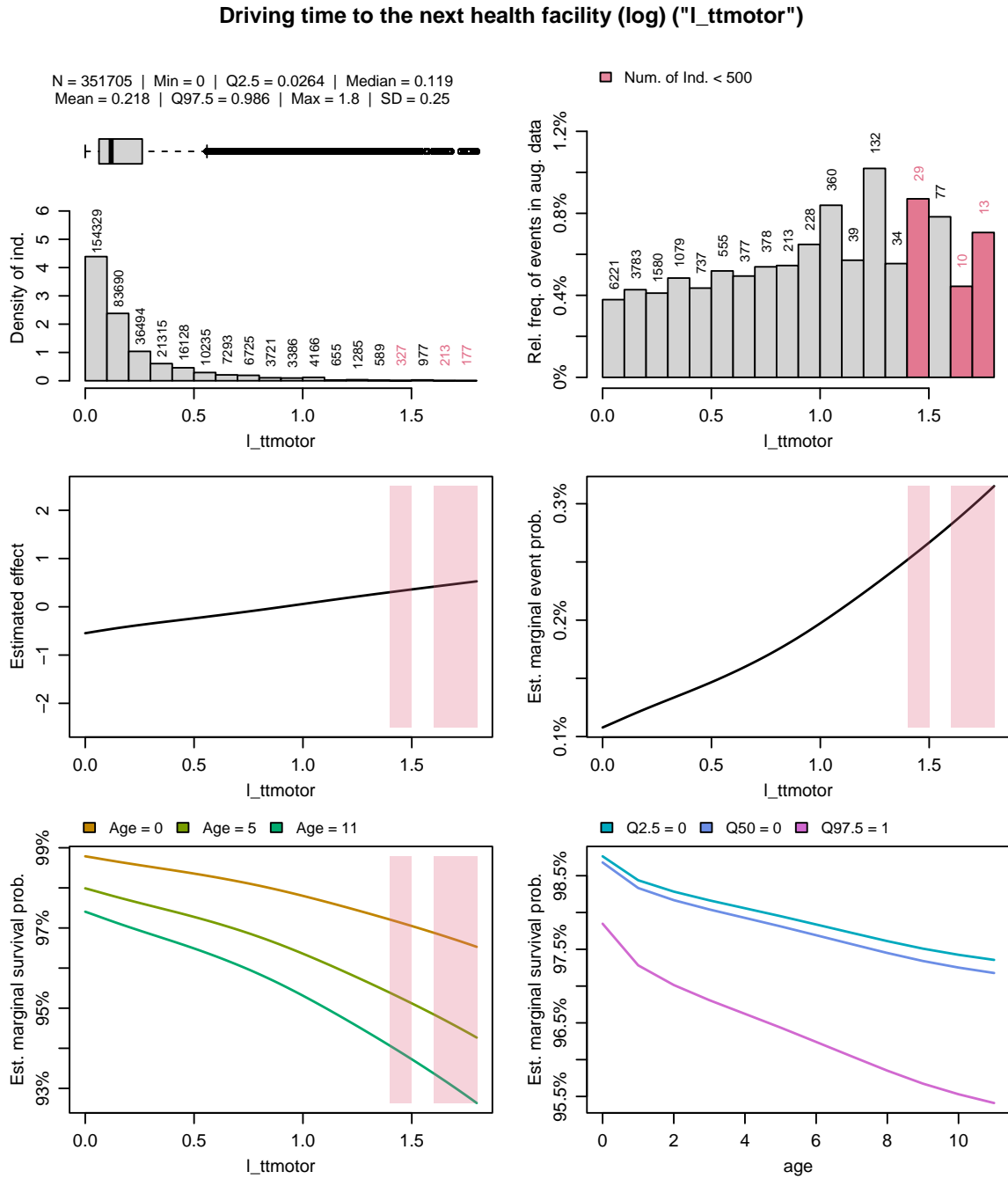
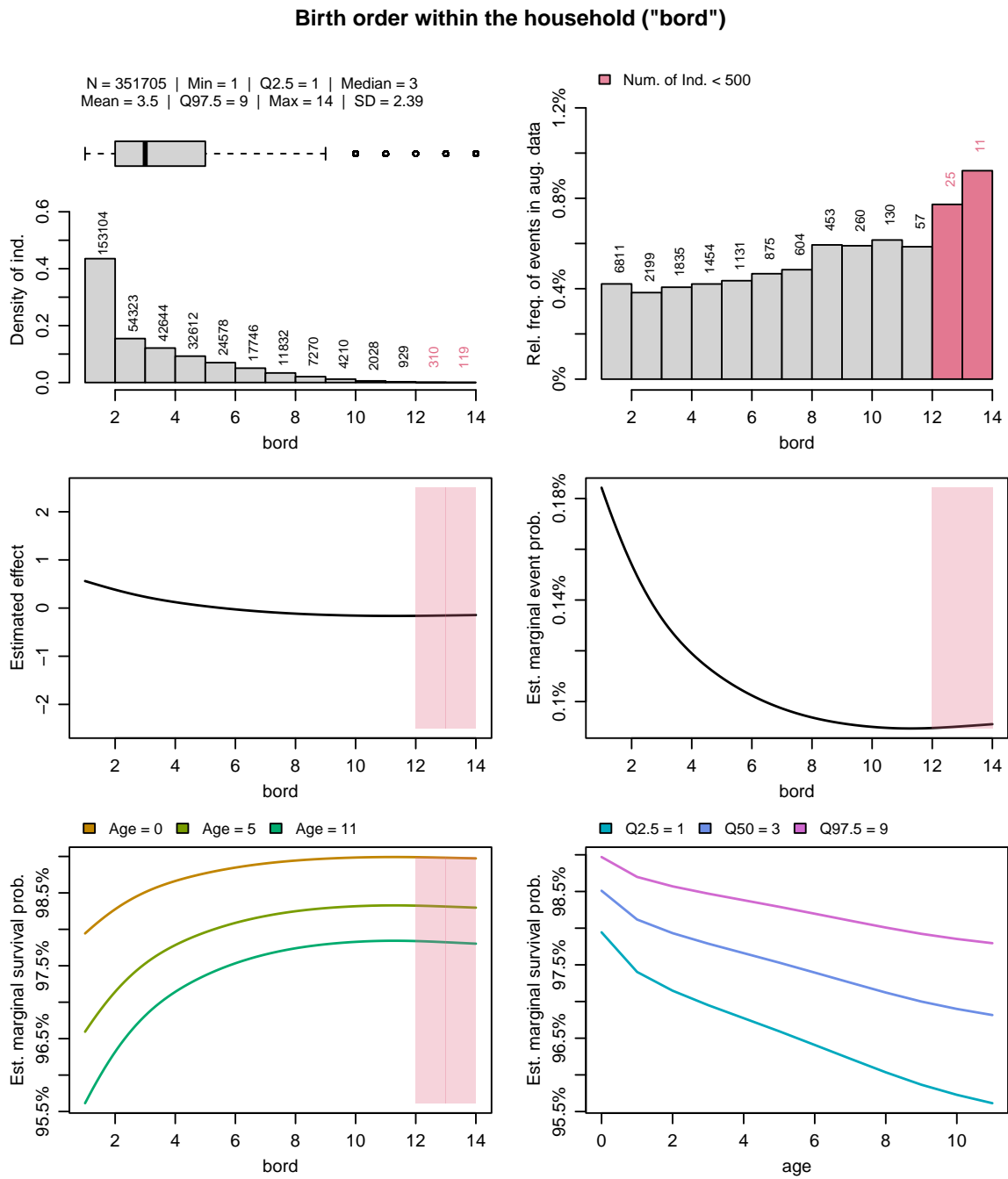
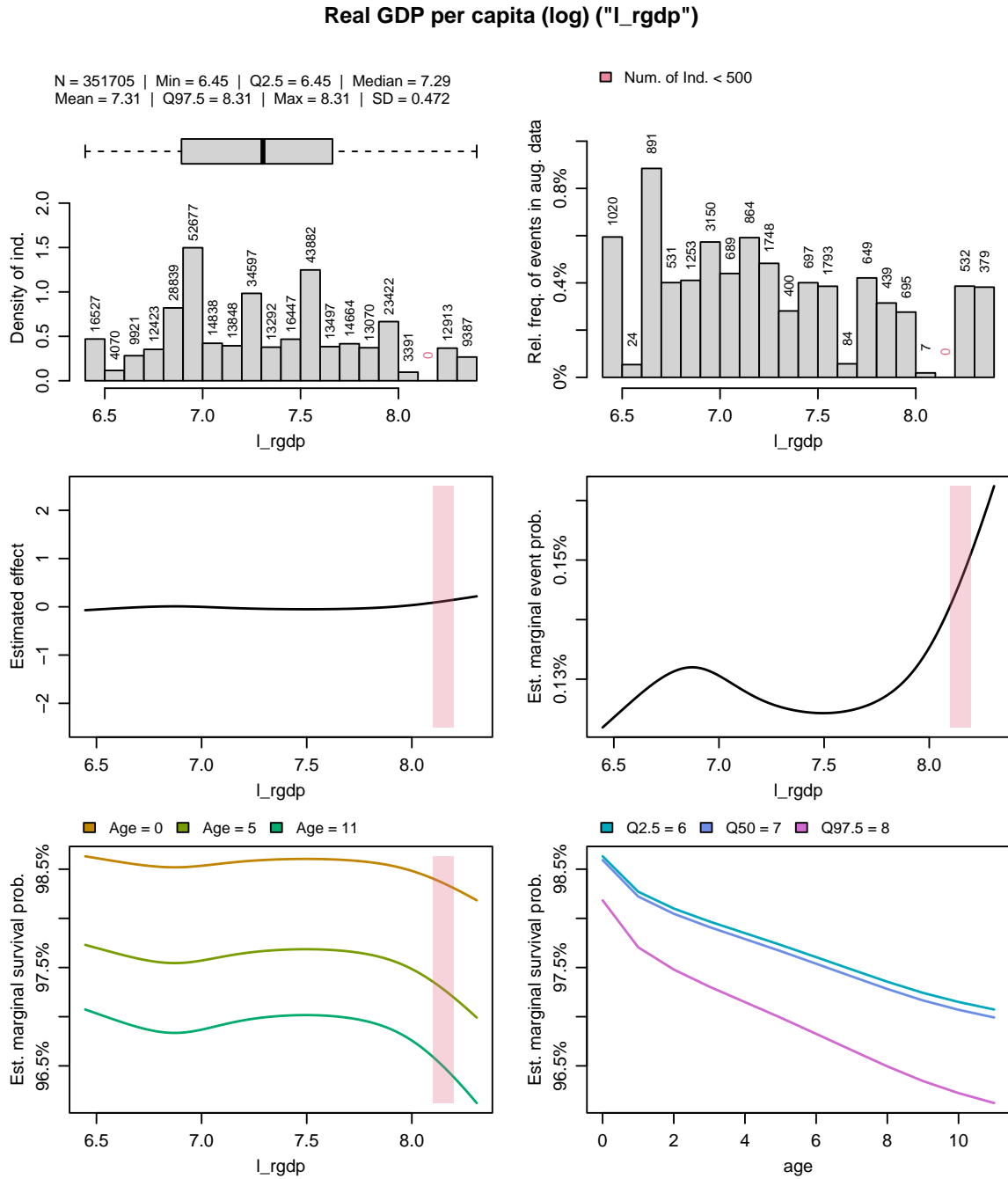
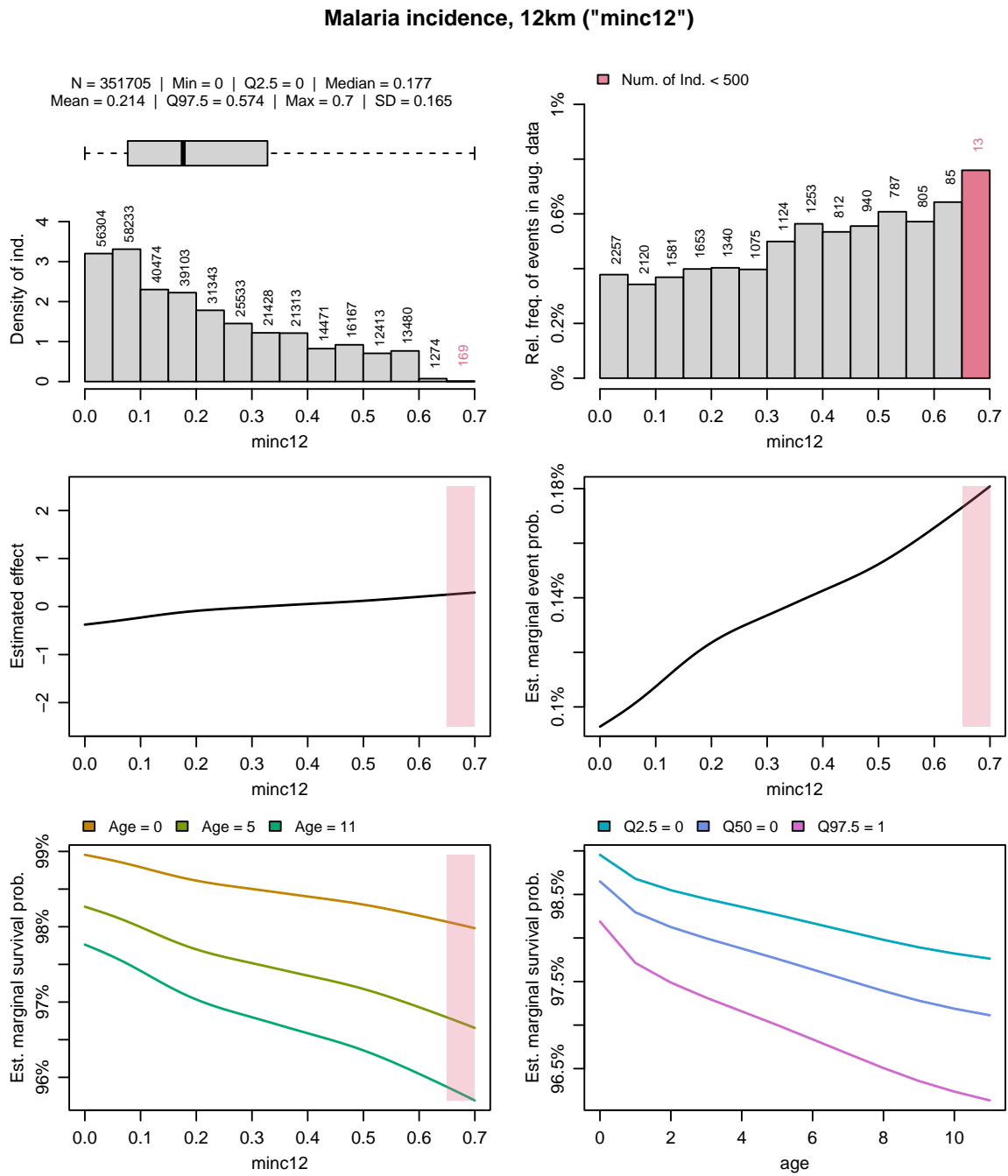
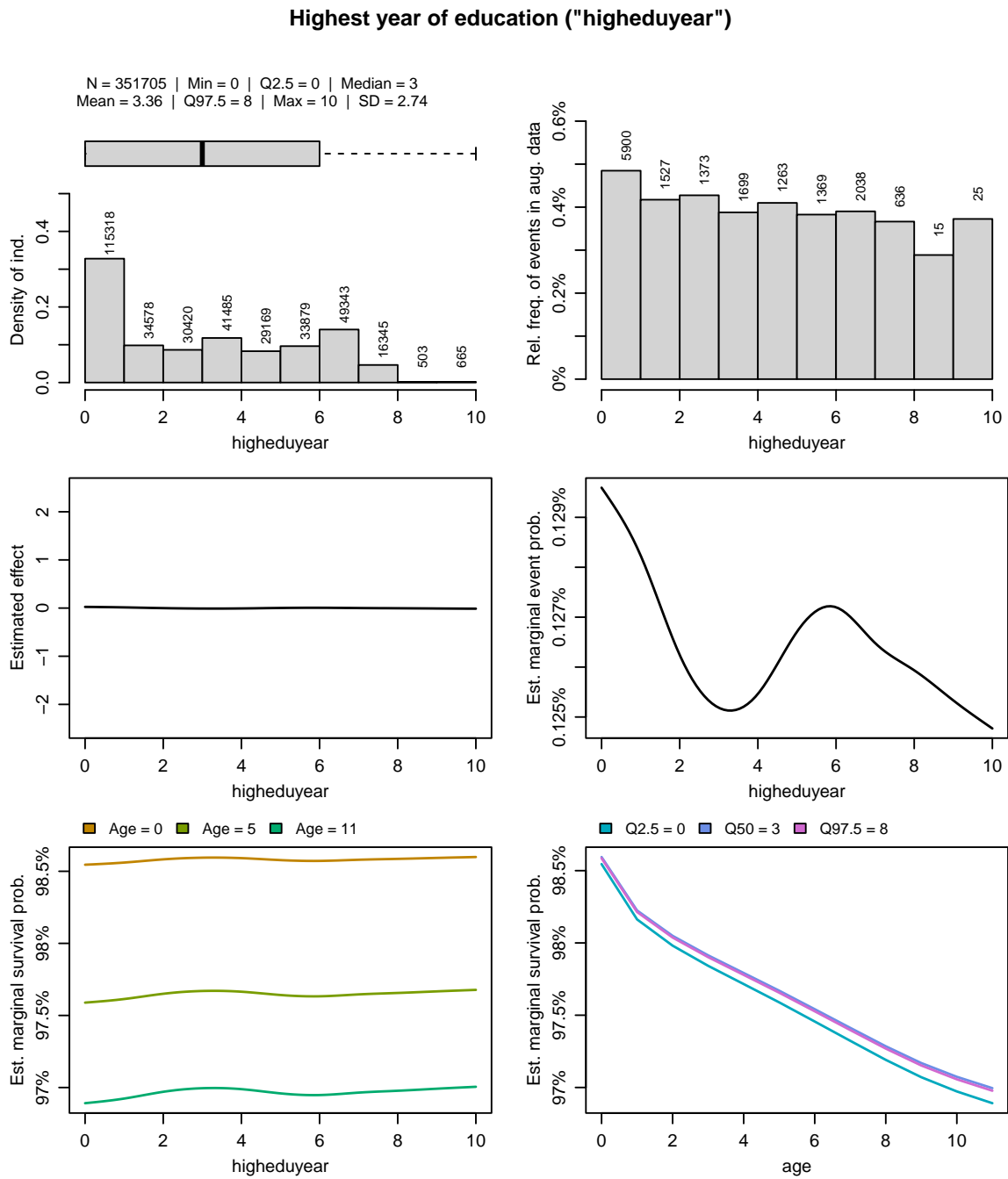


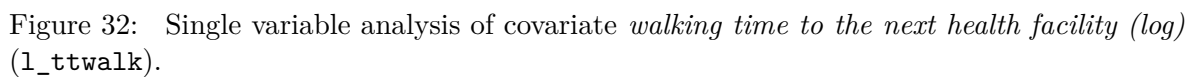
Figure 27: Single variable analysis of covariate *driving time to the next health facility (log)* ($l_ttmotor$).

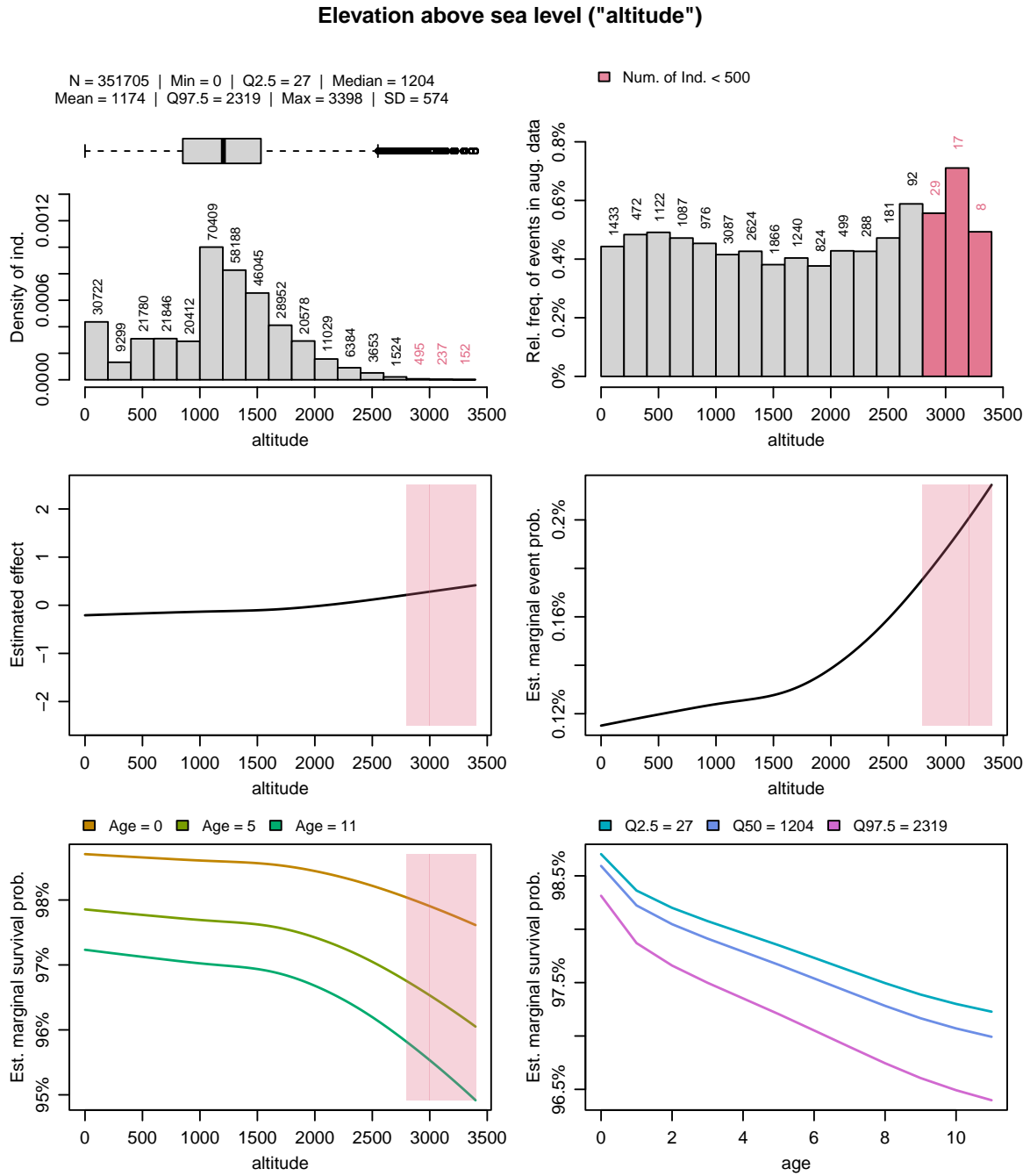
Figure 28: Single variable analysis of covariate *birth order within the household* (bord).

Figure 29: Single variable analysis of covariate *real GDP per capita (log)* (*l_rgdg*).

Figure 30: Single variable analysis of covariate *malaria incidence, 12km* (minc12).

Figure 31: Single variable analysis of covariate *highest year of education* (higheduyear).



Figure 33: Single variable analysis of covariate *elevation above sea level (altitude)*.

B.2. Estimated effects and marginal probability of survival

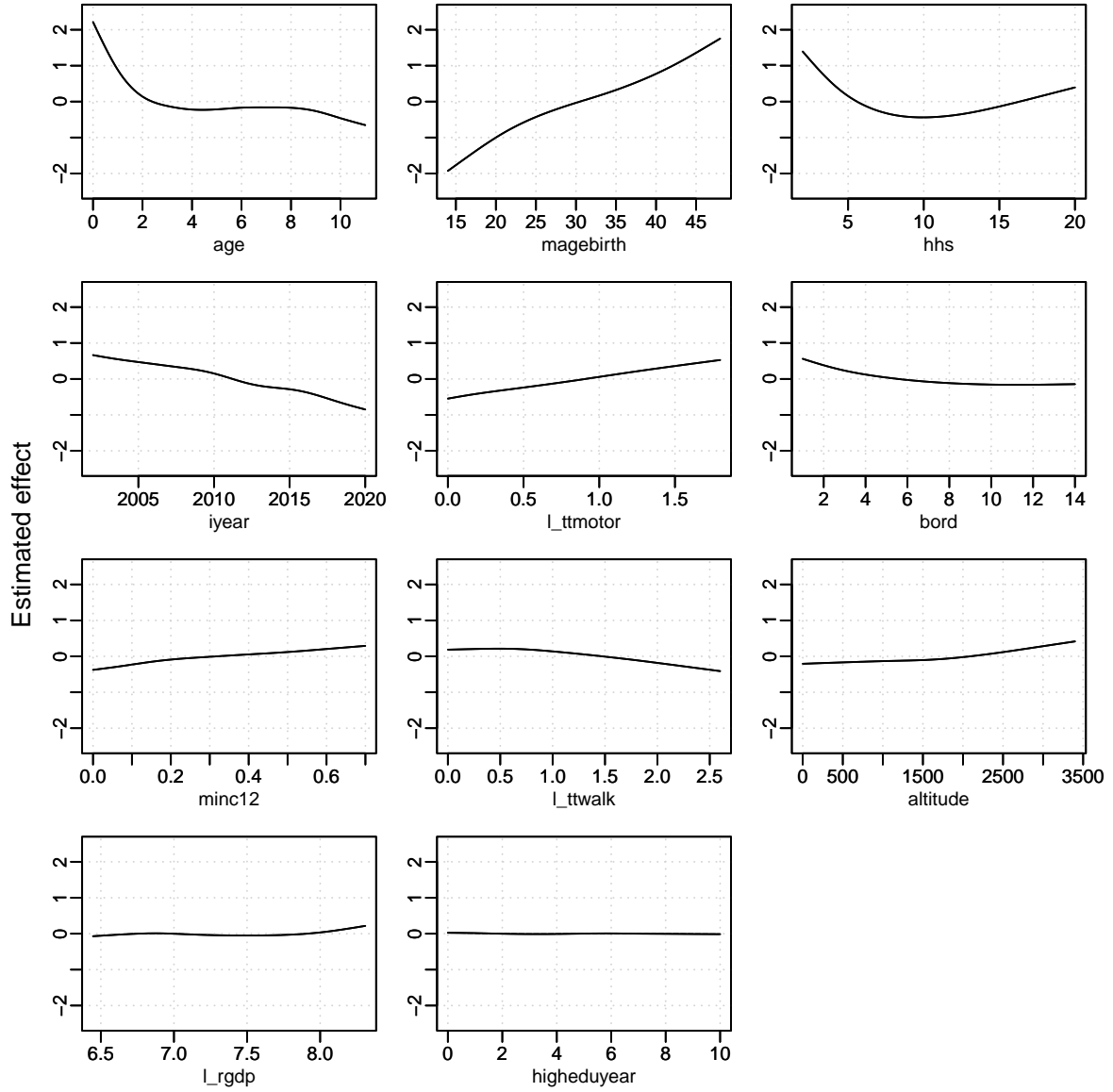


Figure 34: Estimated effects of all selected variables (see Figure 16) centred at zero. See Table 2 for variable description.

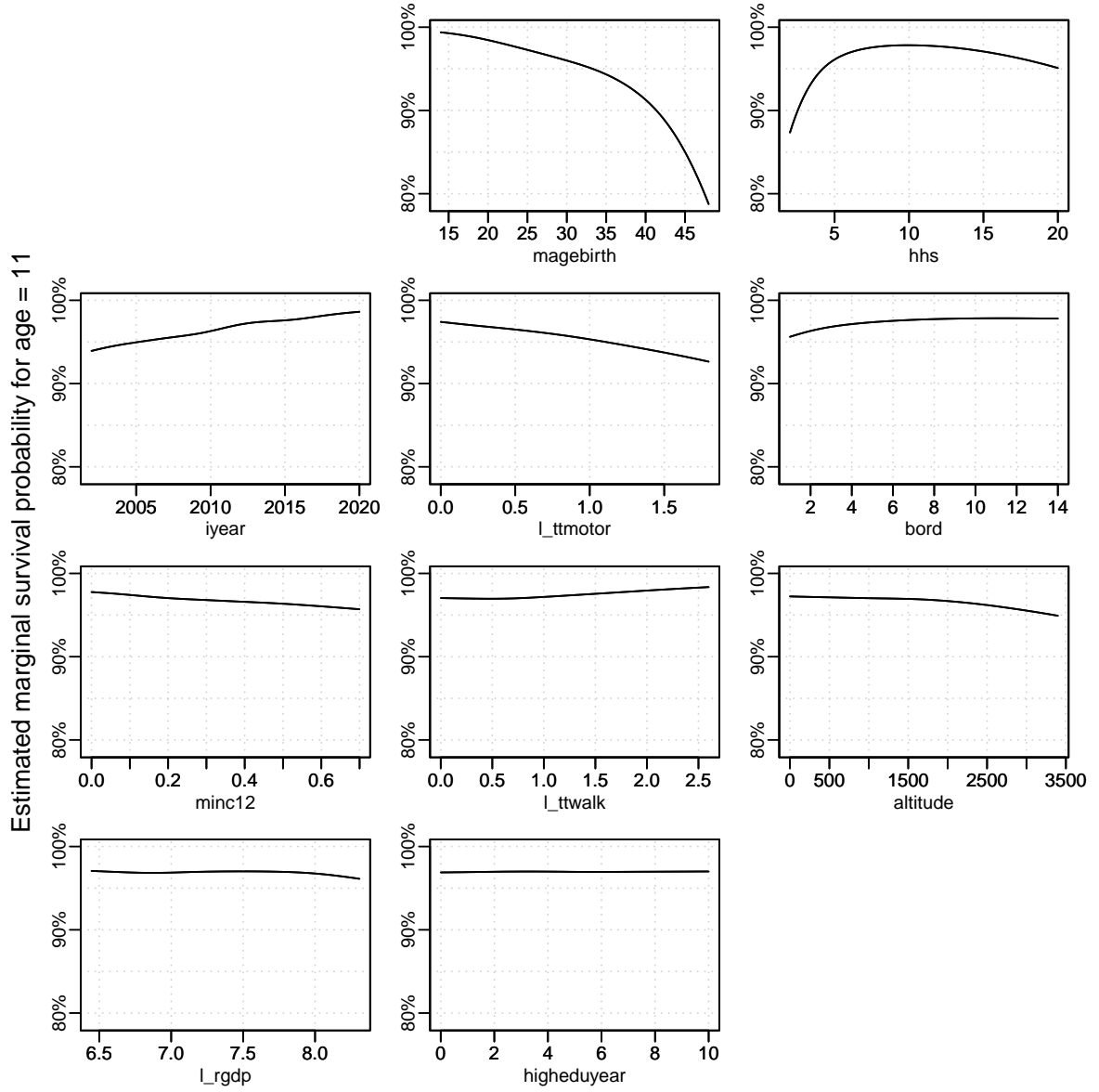


Figure 35: Estimated marginal survival probability for the first year of life (Age = 11) of all selected variables (see Figure 16). See Table 2 for variable description.

Affiliation:

Benjamin Müller, Nikolaus Umlauf, Johannes Seiler, and Stefan Lang
Department of Statistics, University of Innsbruck, Innsbruck, Austria

E-mail: benjamin.mueller@uibk.ac.at,
nikolaus.umlaut@uibk.ac.at,
johannes.seiler@uibk.ac.at,
stefan.lang@uibk.ac.at

Kenneth Harttgen

*Development Economics Group, ETH Zurich, Zurich, Switzerland and
NADEL Center for Development and Cooperation, ETH Zurich, Zurich, Switzerland*

E-mail: kenneth.harttgen@nadel.ethz.ch