

Original Article

A piece-wise additive model of survival data with non-linear rut

Omaku Peter Enesi¹, B. A. Oyejola²

¹Department of Mathematics and Statistics, Federal Polytechnic, Nasarawa, Nigeria, ²Department of Statistics, University of Ilorin, Ilorin, Nigeria

ABSTRACT

A critical step when developing multivariate risk prediction models is to decide which are the predictors that should be included and which ones could be neglected to get an accurate but interpretable model. The general additive models (GAMs) have come as quite an attractive procedure to handle covariate complexities in their various functional forms which is unlikely with the Cox model. In this study, a modification of piece-wise additive hazard model is proposed. Three levels of variance of Weibull distribution were assumed for baseline hazards in generating the data. The sensitivity of the baselines was accessed under four censoring percentages (0%, 25%, 50%, and 75%) and three sample sizes ($n=100$, 500, and 1000), for when models were single additive model (SAM) and when partitioned – piece wise additive model (PAM). A piece-wise Bayesian hazard model with structured additive predictors in which time-varying covariate and the functional form of continuous covariates were incorporated in a non-proportional hazards framework was developed. Hazard function was modeled through additive predictors that are capable of incorporating complex situations in a more flexible framework. Analysis was done utilizing MCMC simulation technique. Results revealed that the PAMs in most situations outperformed the SAMs with smaller DIC values and larger predictive powers with the log pseudo-marginal likelihood criterion.

Keywords: Continuous covariate functions, generalized additive model, penalty splines, piece-wise additive model, proportional hazard, single additive model, time-varying covariates

Submitted: 14-07-2020, **Accepted:** 18-07-2020, **Published:** 30-12-2020

INTRODUCTION

In a ground breaking paper, Cox proposed a model for survival data in the proportional hazard frame work. Sir David Cox observed that if the proportional hazard assumption holds (or, is assumed to hold), then it is possible to estimate the effect of parameter(s) without any consideration of the hazard function.^[1] Developing models in complex analysis may violate this assumption as the model may incorporate time-varying effects and covariates, thereby relaxing the proportional hazard assumption, and allowing the hazard ratio to depend on time t . A common approach to model time-varying effects is by piece-wise constant functions, as these are flexible enough to capture any shape of the baseline hazard or covariate effects,^[2] for a frequentist and^[3] and more recently^[4] for a Bayesian approach.

The hazard model was further extended by Hennerfeind *et al.*, Kneib and Fahrmeir,^[5,6] to incorporate a

flexible spatial generalization of the Cox model, a structured geo-additive predictor, including a spatial component for geographical effects and non-parametric terms for modeling unknown functional forms of the log-baseline hazard rate and non-linear effects of continuous covariates.

Generalized additive models (GAMs) are statistical models in which the conventional multiple linear regression is generalized to permit a much broader class of time-varying and non-linear functional form of continuous covariates and their effects, but still with additive relationships between response and predictor variables. GAMs, derived from the work of Hastie and Tibshirani,^[7,8] provide flexible and effective means of moving out of the “linear rut”^[9] in which a considerable amount of bio-statistical modeling is still located. Conventionally, non-linearity is handled through transformations and then estimation of linear models. Such an approach requires the researcher to have good knowledge of the correct functional

Address for correspondence: Omaku Peter Enesi, Department of Mathematics and Statistics, Federal Polytechnic, Nasarawa, Nigeria.
E-mail: omakupete@gmail.com

form before the model is fitted when, in reality, the choice is very open and theory is usually vague. In contrast, the GAM represents an “adaptive” approach in which the data help guide the choice of appropriate functional form.^[10]

In approaching these complexities in building a model, data are partitioned in bits of intervals and regularized estimation through penalized spline is carried out. The complexity of covariates included in models and the estimation method adopted by different authors such as Marano *et al.*, Lang and Brezger^[4,11] has, however, caught our attention in this study.

Motivated by this, we studied and compared the sensitivity of models under different censoring percentages, for several sample sizes within the framework of Weibull distributions whose shape and scale parameters are obtained by varying the variances while keeping the mean at 1; through the unit root function implemented in R.

The study further investigates: The performance of single additive models (SAMs) and the modified piece-wise additive extension or piece-wise additive models (PAMs) under various censoring percentages and sample sizes employing three levels Weibull baseline variance.

METHODOLOGY

The risk data used for this paper were simulated from a Weibull baseline hazard distribution which was used to generate survival times for sample sizes of 100, 500, and 1000, respectively. Various censoring levels or percentages of no censoring “0%,” low “about 25%,” moderate “about 50%,” and high “about 75%” were used.

Model Specification

The Cox hazard model

$$\lambda_i(t, x) = \lambda_0(t) \exp\left(\sum_{j=1}^p \beta_j x_j\right) \quad (1)$$

The baseline hazard rate is unspecified and assumes that covariates $x=(x_1, \dots, x_p)$ act multiplicatively on the hazard rate through the exponential link function.^[1]

Model Extension

$$\lambda_i(t, x_{in}(t)) = \lambda_0(t) \exp\left(\gamma_1 x_{i1} \dots \gamma_p x_{ip} + \gamma_j x_{in}(t)\right) \quad (2)$$

Where, $\lambda_i(t)$ is the hazard function for individual i at time t .

$\gamma_1 x_{i1} \dots \gamma_p x_{ip}$: Are time invariant covariates.

$\gamma_j x_{in}(t)$: Time-dependent covariates

$\lambda_0(t)$ is the baseline hazard function for an individual whose covariates $x_1 \dots x_p$ all have values of 0.

An additive representation of model (1) is given by:

$$\eta(t, z, w, x) = f_0(t) + \sum_{i=1}^p f_i(t) z_i + \sum_{j=1}^{p+1} f_j(w_j) + x' \gamma \quad (3)$$

This is a reparameterization of the Cox model

Where, $f_0(t) = \log \lambda_0(t)$ is the functional form of the baseline hazard, which implies $\exp(f_0(t))$. Other aspects of the model include the functions $f_1(t) z_1 \dots f_p(t) z_p$ which are the functional form of time-varying covariates z_1, \dots, z_p , the functions $f_j(w_1) \dots f_q(w_q)$ are possibly non-linear effects of metrical covariates w_1, \dots, w_q and γ is the usual linear part of the predictor for some categorical covariates.^[5,12]

Modification of Piece-wise Additive Models (PAMs)

The proposed model in bits of intervals is given as

$$\lambda_{PAM}(t; z, w, s, x) = \{I(t) T_h(f_h(t))\} + \sum_{j=1}^p f_j(t) z_{jh} + \sum_{j=p+1}^{p+q} f_j(w_{jh})_{PSplines} + f_{spat}(s_{ih}) + x' \gamma \quad (4)$$

with its various terms defined as:

The function $f_h = \log \lambda_h$ is the baseline effect for the h^{th} interval of PAM

The functions $f_1(t) z_{11} \dots f_p(t) z_{ph}$ are the functional form of time-varying covariates z_{11}, \dots, z_{ph} in the h^{th} interval, the functions $f_j(w_{11}), \dots, f_p(w_{qh})$ are possibly nonlinear effects of metrical covariates w_{11}, \dots, w_{qh} in the h^{th} interval, and $f_{spat}(s_{ih})$ is a structured spatial effect, where $s, s = 1, \dots, S$ is either a spatial index, with $s = s_i$ if subject i in the h^{th} bit (interval) is from area s or it is an exact spatial coordinate $s = (x_i, y_i)$, for example, for centroids of regions or if exact locations of individuals are known.

γ is the usual linear part of the predictor.

$$L_{PAM}(\underline{\lambda}, \underline{\beta}; D, \Delta, X, s) = \prod_{i=1}^n \prod_{h=1}^{H_i} (\lambda_h \exp(X_i^T \underline{\beta} + s_{ih})^{d_{ih}} \cdot \exp(\lambda_h \exp(X_i^T \underline{\beta} + s_{ih}) \Delta_{ih}) \Delta_{ih}) \quad (5)$$

Where, for each subject i there is a product of H_i terms, H_i being the number of intervals in which the subject is followed. In the expression above, d_{ih} is the status of the i^{th} subject within the interval T_h ($0 =$ alive or censored, $1 =$ failed); Δ_{ih} is the time spent in T_h by the subject. From expression (5), it may be seen that L_{PAM} is proportional to the product of Poisson likelihoods for D_{ih} with mean parameters: $\mu_{ih} = \lambda_h \exp(X_i^T \underline{\beta} + s_{ih}) \Delta_{ih}$. As a consequence, the expression of the Poisson regression model is:

$$D_{ij} \sim \text{POISSON}(\mu_{ih}); \log(\mu_{ih}) = \underline{\alpha}_h + X_0^T \underline{\beta} + s_{ih} + \log(\Delta_{ih}) \quad (6)$$

Where, $h(i)$ indicate the interval where t_i falls, that is, the interval where individual i died or was censored,

$\alpha_h = \log(\lambda_h)$ are log-hazard parameters, and the term $\log(\Delta_{ih})$ is an offset.

Bayesian P-splines Approach for Modeling the Unknown Functions

A number of competing approaches are available for modeling and estimating non-linear function f_j of continuous covariates. These include smoothing splines,^[8] local polynomials,^[13] regression splines with adaptive knot selection,^[14,15] and P-splines.^[16,17] In this study, Bayesian version of penalized splines is employed following.^[11] P-splines were introduced in a frequentist setting by Eilers and Marx, Marx and Eilers.^[16,17] The basic idea of P-splines is to assume that an unknown smooth function $f_j(x_j)$ of a continuous covariate x_j can be approximated by a linear combination of B-spline basis functions B_m .^[18] That is, denoting the m -th basis function by B_{jm} , we then obtain

$$f_j(x) = \sum_{j=1}^{d_j} \beta_j B_j(x) \quad (7)$$

The basic functions B_m are B-splines of degree l defined over a grid of equally spaced knots $x_{jmin} = \xi_0 < \xi_1 < \dots < \xi_s = x_{jmax}$, $d_j = l + s$, where s is the number of the equally spaced knots.

The Bayesian P-splines method is based on a hierarchical model with non-informative priors for the regression coefficients (β) and a Gaussian random walk (RW) prior of order d for the coefficients of the hazard function (B-spline), conditional to a smoothing parameter τ^2 the general expression of the RW prior as suggested by Lang and Brezger, Kooperberg and Intrator^[11,19] is the following:

$$\beta_j / \tau_j^2 \propto \exp\left(-\frac{1}{2\tau_j^2} \lambda_j \beta_j' K_j \beta_j\right) \quad (8)$$

The penalty matrix K_j is of the form $K_j = D'D$, where D is a first or second order difference matrix.

For an independent and identical random effect, the penalty matrix is the identity matrix, that is, $K_j = I$. The variance parameter τ_j^2 controls the tradeoff between flexibility and smoothing and an inverse gamma prior (the conjugate prior) is assumed, that is, $\tau_j^2 \sim IG(a, b)$.

Gaussian Random Field (GRF) Priors

For georeferenced data, it is commonly assumed that $v_i = v(s_i)$ arises from a Gaussian random field (GRF) $\{v(s), s \in S\}$

such that $v = (v_1, \dots, v_m)$ follows a multivariate Gaussian distribution as $v \sim N_m(0, \tau^2 R)$, where τ^2 measures the amount of spatial variation across locations and the (i, j) element of R is modeled as $R[i, j] = \rho(s_i, s_j)$. Here $\rho(\cdot, \cdot)$ is a correlation function controlling the spatial dependence of $v(s)$. In “survregbayes” package in R, the powered exponential correlation function $\rho(s_i, s_j) = \rho(s_i, s_j, \phi) = \exp\{-\phi \|s - s'\|^v\}$ is used, where $\phi > 0$ is a range parameter controlling the spatial decay over distance, $v \in (0, 2]$ is a pre-specified shape parameter, and $\|s - s'\|$ refers to the distance (e.g., Euclidean, great-circle) between s and s' . Therefore, the prior GRF (τ^2, ϕ) is defined as

$$V_i / \{v_j\}_{j \neq i} \sim N \left(- \sum_{\{j: j \neq i\}} P_{ij} v_j / P_{ii}, \tau^2 / P_{ii} \right) \quad (9)$$

$i = 1, \dots, m$ where P_{ij} is the (i, j) element of R^{-1} ^[20]

Hazard Model with Regularized Functions

$$\left\{ \begin{array}{l} \eta(t; w, x) = \underline{B}_0^T f_0(t) + \underline{B}_0^T f_j(w_j) + X_0^T \underline{\gamma} \\ \underline{\beta} \gamma \sim \pi_\gamma \\ \left(\underline{f}_0(t) | \tau^2 \right) \sim RW(\tau^2, P_d); \tau^2 \sim \pi_{\tau^2} \\ \left(f_j(w_j) | \tau^2 \right) \sim RW(\tau^2, P_d); \tau^2 \sim \pi_{\tau^2} \end{array} \right. \quad (10)$$

Where, B_0 is the vector form of a B-spline of degree 0 defined for the follow-up period, and π^β and π^{τ^2} are generic prior densities for the regression coefficients and the smoothing parameter.^[20,21]

Regularized piece-wise additive model (RPAM)

$$\left\{ \begin{array}{l} D_{ij} \sim \text{POISSON}(\mu_{ih}) \\ \log(\mu_{ih}) = \underline{B}_0^T \underline{\alpha} + \sum_{j=1}^p Z_{j,i} \underline{B}_0^T \gamma_{1j} + \sum_{i=1}^q \underline{B}(w_{j,i})^T \gamma_{2j} + v_{ij} + \log(\Delta_{ih}) \\ \left(\underline{\alpha} | \tau^2 \right) \sim RW(\tau^2, P_d); \tau^2 \sim \pi_{\tau^2} \\ \left(\gamma_{ij} | \tau_j^2 \right) \sim RW\left(\tau_j^2, P_d^{(j)}\right); \tau_j^2 \sim \pi_{\tau_j^2}; j = 1, \dots, p \\ \left(\gamma_{ij} | \tau_j^2 \right) \sim RW\left(\tau_{p+l}^2, P_d^{(p+l)}\right); \tau_{p+l}^2 \sim \pi_{\tau_{p+l}^2}; l = 1, \dots, q \\ V_i / \{v_j\}_{j \neq i} \sim N \left(- \sum_{\{j: j \neq i\}} P_{ij} v_j / P_{ii}, \tau^2 / P_{ii} \right) \end{array} \right. \quad (11)$$

The time-dependent effects for each covariate are: $Z_{1,j,i} B_0^T \gamma_{1j}$; $j = 1, \dots, p$. Thus, for each Z_j , its values are multiplied by a piece-wise constant function: $B_0^T \gamma_{1j}$; in the parameters. $\gamma_{1j} = (\gamma_{1,j,1}, \dots, \gamma_{1,j,H})^T$. This enables the effect of each Z_j to vary in each interval T_h of the original partition of the follow-up: $B_0^T \alpha + Z_{1,j,i} B_0^T = \alpha_h + z_{1,j,i} \gamma_{1j,h}$ for $t \in T_h$. On the other hand, the effects of w_1, \dots, w_q have been represented through the general expression of B-splines, with the vector $B(w_{j,i})^T$ including the spline basis of w_i calculated for the i^{th} individual. Thus, in this case, the degree and the knots of each spline are fixed following conventional rules. π^β and π_τ^2 are generic prior densities for the regression coefficients and the smoothing parameter.

Model Specification to Advance Simulation

Model 1:

$$\lambda_{SAM}(t; z, w, s) = f_0(t) + f(t)z_j + f(w_i) + f_{spat}(s_{ih})$$

Model 2:

$$\lambda_{PAM}(t; z, w, s) = \{I(t \in T_h(f_h(t)))\} + f_h(t)z_j + f_{h(\text{pslines})}(w_i) + f_{spat}(s_{ih})$$

Where, λ_{PI} is the hazard function when partitioning is ignored (PI) or SAM

Where, λ_{PD} is the hazard function when partitioning is done (PD) or piece wise additive model (PAM).

Test for Non-proportionality

To test the hypothesis that the proportional hazard assumption is valid, the following statement of hypothesis is made.

- $H_0: \delta_1 = \delta_2 = \dots = \delta_p$ (Assumption is valid)
- $H_1: \text{At least one of the } \delta_i \text{'s is not equal to zero (Assumption violated)}$

Decision rule: Reject H_0 if $P \leq \alpha$ (level of significance)

Residual measures are used to investigate the departure from the proportional hazard assumption. Schoenfeld residuals are used to test the assumption of proportionality. Schoenfeld residuals are usually calculated at every failure of time under the proportional hazard assumption and usually not defined for censored observations. The overall significance test is called the global test of the model, sighted in Adeniyi and Akinrefon.^[22]

Data Analysis

The data are simulated using the functional form of time-varying covariate by Bender *et al.*^[23] given as

$$f(t) = 0.5\sqrt{t} * y \quad y \sim \text{binom}(N, 1, 0.5)$$

The functional form of the continuous covariates as in Brezger^[24] is given as:

$$f_1(w_i) = 1.0 / 1.75w_i$$

where $x_i \sim U(-3, 3)$.

For spatial frailty, we propose, $S = \text{pnorm}(v)$ and $v \sim \text{mvnorm}(I, \Sigma)$; if $S = \text{pnorm}(v)$ then $S \sim \text{mvnorm}$, where, Σ is the covariance matrix for spatial correlation.

Coordinates for spatial correlations follow the uniform distribution. $s_1 = \text{runif}(N, 0, 40)$ and $s_2 = \text{runif}(N, 0, 100)$,^[25] obtained the shape and scale parameters of the Weibull distribution from the following

$$\eta = \frac{1}{\Gamma(1 + \frac{1}{\alpha})} \tag{12}$$

And

$$\left(\frac{\Gamma\left(1 + \frac{2}{\alpha}\right)}{\left(\Gamma\left(1 + \frac{1}{\alpha}\right)\right)^2} - 1 \right) = 0.5 \tag{13}$$

for a convenience choice of mean 1 and variance 0.5. Using the uniroot function in R. Parameters were given to be approximately $\alpha = 1.435523$ and $\eta = 1.101321$. We considered studying the impact of increasing and decreasing the variance of the Weibull distribution while keeping the mean at 1. The result is displayed in Table 1.

Data simulation and analysis were carried out in R using the coda package for spBayesSurv, version 3.6.2. Comparisons were done using deviance information criterion (DIC) (smaller is better) which places emphasis on the relative quality of model fitting and log pseudo-marginal likelihood (LPML) (larger is better) focuses on the predictive performance. Both criteria are readily computed from the MCMC output.

RESULTS

Table 1: Shape and scale parameters of the Weibull distributions

E(T)	Var(T)	α	η
1	0.25	2.101377	1.129063
1	0.5	1.435523	1.101321
1	0.75	1.157975	1.052847

Table 2: DIC values for levels of censoring, three levels of Weibull baseline variance and sample sizes for partitioned and no partition models

Censoring %	Variance	n=100		n=500		n=1000	
		No partition	PAM	No partition	PAM	No partition	PAM
0	0.25	853.6004	832.1385	4068.979	3999.343	8779.792	8597.94
	0.5	962.164	943.7934	4707.551	4686.069	10,137.67	10,085.62
	0.75	1032.945	1005.326	5138.187	5076.758	11,028.65	10,970.23
25	0.25	653.4635	644.8024	3210.955	3181.119	7007.161	6939.78
	0.5	716.3681	702.3201	3623.79	3613.921	7820.168	7813.243
	0.75	756.2604	746.4749	3848.511	3829.991	8272.722	8263.318
50	0.25	512.0976	488.0027	2407.989	2356.165	5000.534	4942.05
	0.5	520.4511	515.8235	2620.238	2604.272	5512.023	5514.923
	0.75	543.8143	540.5304	2607.02	2552.504	5698.33	5707.325
75	0.25	255.7324	287.7841	1280.065	1275.756	2695.128	2677.361
	0.5	312.231	299.871	1338.623	1350.044	2850.098	2848.829
	0.75	269.8102	267.6131	1350.597	1353.066	2831.845	2841.659

Table 3: LPML values for levels of censoring, three levels of Weibull baseline variance, and sample sizes for partitioned and no partition models

Censoring %	Variance	n=100		n=500		n=1000	
		No partition	PAM	No partition	PAM	No partition	PAM
0	0.25	-430.1177	-426.613	-2036.033	-2000.033	-4392.223	-4280.412
	0.5	-483.6684	-475.1374	-2355.252	-2344.109	-5071.276	-5044.9
	0.75	-519.1246	-506.7498	-2570.389	-2539.7	-5516.488	-5485.61
25	0.25	-329.0801	-314.399	-1606.389	-1593.613	-3504.767	-3474.337
	0.5	-360.4299	-352.6802	-1812.87	-1809.941	-1416.636	-1426.327
	0.75	-380.336	-376.0484	-1925.237	-1918.535	-4137.215	-4132.988
50	0.25	-289.8239	-246.5223	-1204.68	-1178.775	-2500.928	-2473.196
	0.5	-262.7259	-261.5076	-1310.931	-1304.996	-3910.985	-3908.749
	0.75	-274.6984	-274.7674	-1304.353	-1277.373	-2849.581	-2854.882
75	0.25	-128.6635	-148.3942	-641.0598	-638.736	-1348.426	-1341.252
	0.5	-159.6845	-157.4531	-670.0822	-678.3336	-2757.031	-2758.394
	0.75	-135.9174	-134.4134	-676.0527	-681.0287	-1416.071	-1423.263

Interpretation

From Tables 2 and 3: The PAMs perform better than the SAMs – by Hennerfeind *et al.*, and Abiodun.^[5,12] This suggests that smoothing with smaller bit of intervals is better than when partitioning is ignored. It was observed that P-spline smoothing for continuous covariates in smaller sample sizes is better than in larger sizes. Increments in variance parameter for the baseline distribution reduce the precision and predictive power of the models (high values of DIC and low values of LPML). This was particularly observed in PAMs for intermediate and high levels of Weibull baseline variance at 75% censoring when sample size is $n = 500$ and 1000, and for intermediate levels of variance at 50% and 75% censoring, when $n = 1000$.

Precision and predictive power (based on DIC and LPML) improved with increase in censoring percentages. PAM outperformed the SAMs for reduced amount of the spread of event times and censoring percentages. Increase in sample size induces increase in DIC and a decrease in LPML values.

In general when dealing with survival analysis data in a non-proportional framework, with complexities associated with covariate combinations, smaller sample sizes produce a better model representation for SAMs and much better when the models are represented within bits of intervals in constant hazards – PAMs. Increase in the variability in the baseline

distribution seems to increase the spread of survival times which further reduces model precision.

CONCLUSION

This research brings to knowledge the essence of performing survival data analysis in bits of time, especially in scenarios with covariate complexities, which makes interpretation of hazard rates difficult. PAM has been shown in this study to be better than SAM at all levels of variance, censoring percentages, and sample sizes.

REFERENCES

1. Abiodun AA. Analyzing competing risk survival time data using cox and parametric proportional hazards models. *JNSA* 2007;19:74-9.
2. Verweij PJ, van Houwelingen HC. Time-dependent effects of fixed covariates in cox regression. *Royal Stat Soc Series B* 1995;34:187-220.
3. Gamerman D. Bayes estimation of the piece-wise exponential distribution. *IEEE Trans Reliab* 1994;43:128-31.
4. Marano G, Boracchi P, Biganzoli EM. Estimation of the piecewise exponential model by Bayesian P-splines via Gibbs sampling: Robustness and reliability of posterior estimates. *Open J Stat* 2016;6:451-68.
5. Hennerfeind A, Brezger A, Fahrmeir L. Geoadditive survival models. *J Am Stat Assoc* 2006;101:1065-75.
6. Kneib T, Fahrmeir L. A mixed model approach for geo additive hazard regression. *Scand J Stat* 2007;34:207-28.
7. Hastie T, Tibshirani R. Generalized additive models. *Stat Sci* 1986;1:297-318
8. Hastie T, Tibshirani R. *Generalized Additive Models*. London: Chapman and Hall; 1990.
9. Jones K, Almond S. Moving out of the linear rut the possibilities of generalized additive models. *Trans Inst Br Geogr* 1992;17:434-47.
10. Marra G, Wood SN. Practical variable selection for generalized additive models. *Comput Stat Data Anal* 2011;55:2372-87.
11. Lang S, Brezger A. Bayesian P-splines. *J Comput Graphical Stat* 2004;13:183-212.
12. Abiodun AA. A Bayesian approach to exploring unobserved heterogeneity in clustered survival and competing risk data. *JNSA* 2009;20:1-13.
13. Fan J, Gijbels I. *Local Polynomial Modelling and its Applications*. London: Chapman and Hall; 1996.
14. Friedman J, Silverman B. Flexible parsimonious smoothing and additive modelling (with discussion). *Technometrics* 1989;31:3-39.
15. Stone C, Hamsen M, Kooperberg C, Truong Y. Polynomial splines and their tensor products in extended linear modelling. *Ann Stat* 1997;25:1371-470.
16. Eilers PH, Marx BD. Flexible smoothing using B-splines and penalized likelihood (with comments and rejoinder). *Stat Sci* 1996;11:89-121.
17. Marx DB, Eilers HC. Direct generalized additive modelling with penalized likelihood. *Comput Stat Data Anal* 1998;26:93-209.
18. De Boor C. *A Practical Guide to Splines*. New York: Springer-Verlag; 1978.
19. Kooperberg C, Intrator N. Trees and splines in survival analysis. *Stat Methods Med Res* 1995;4:237-61.
20. Zhou H, Hanson T. A unified framework for fitting bayesian semiparametric models to arbitrarily censored survival data, including spatially-referenced data. *J Am Stat Assoc* 2017;113:571-81.
21. Omaku PE, Ibinayin JS, Tanko N, Braimah JO. A modified additive hazard model for some risk factors associated with hypertensive condition. *AJMS* 2020;3:1-8.
22. Adeniyi OI, Akinrefon AA. First birth interval: Cox regression model with time varying covariates. *CJPL* 2018;6:1-7.
23. Bender R, Augustin T, Blettner M. Generating survival times to simulate cox proportional hazards models. *Stat Med* 2005;24:1713-23.
24. Brezger A. *Bayesian P-splines in Structured Additive Regression Models*. PhD Thesis. Munich: Springer-Verlag, LMU; 2004.
25. Ulviya A. *Frailty Models for Modelling Heterogeneity*. Canada: McMasters University; 2013.



This work is licensed under a Creative Commons Attribution Non-Commercial 4.0 International License.