On Piece-Wise Modelling of Survival Data with Time Changing Covariate Function

P. E. Omaku & J. S Ibinayin
Department of Mathematics & Statistics, Federal Polytechnic Nasarawa – Nigeria

Abstract

Survival analysis involve the set of statistical techniques or procedures used to study time until an event occurs, these techniques are not without some conditions. One of the basic assumptions is that, to enable a straightforward interpretation of hazard rates of subject's covariate(s) on some reference categories or in situations where variables are continuous in nature, the hazard rates must be constant through time “also known as the proportional hazard assumption” for Cox regression. This assumption is often violated in medical practice where subject's vital statistics or measures are often time varying, as their medical situations changes with time. This paper under study a modification of Piece wise survival model, where three levels of Weibull distribution were assumed for baseline hazards, the sensitivity of the baselines were assessed under four (4) censoring percentages (0%, 25%, 50%, & 75%) and sample sizes (n=100, n=500 & n=1000), for when models were Single parametric (SPM) and when partitioned – Piece wise Parametric Model (PPM). A Piece-wise Bayesian hazard model with structured additive predictors in which the functional form of time varying covariate was incorporated in a non-proportional hazards framework was developed, capable of incorporating complex situations in a more flexible framework. Analysis was done utilizing MCMC simulation technique. Results revealed on comparison that the PPM outperformed the SPM with smaller DIC values and larger predictive powers with the LPML criterion and consistently so throughout all simulations.

Keywords: Time varying covariates, Proportional hazard, Violation, Piece-wise survival model, Piece-wise parametric model, Single parametric model

Corresponding Author: P. E. Omaku
Background to the Study
Survival analysis is a statistical procedure for data analysis for which the outcome variable of interest is time until an event occurs. By time, we mean years, months, weeks, or days from the beginning of follow-up of an individual until an event occurs and by event, we mean death, disease incidence, re-lapse from remission, recovery (e.g., return to work) or any designated experience of interest that may happen to an individual (David and Kleinbaum 2005).

Analysis of survival times data has gained a considerable attention, particularly in the field of medicine, where the conventional denotation 'Survival analysis' arises from (Hennerfeind, 2006). In several other bio-statistical applications on censored follow-up time data, the interest lies mainly on the prognostic role of clinical/biological covariates. To such end, non-parametric and semi-parametric methods have been preferred over parametric ones. The most widely adopted tool is the Cox model, which avoids any assumption of the functional form of the hazard function on time. However, such feature is not useful if the interest lies on investigating the shape of the hazard or in predictive modeling (Kooperberg et al. 1995) when the cox-model is extended to time-varying covariates and time-dependent effects, which combine to give the most general version of the hazard. Again, further progress would require specifying the form of this function of time. In such situation where time is observed to be truly continuous a flexible or semi-parametric strategy is required, where mild assumptions are made about the baseline hazard $\lambda_0(t)$. Specifically, we may subdivide time into reasonably small intervals and assume that the baseline hazard is constant in each interval, leading to a piece-wise survival model.

According to Fabio et al. (2010) the Piecewise Model (PM) arises as a quite attractive alternative to parametric models for the analysis of time to event data. Although parametric in a strict sense, the PM can be thought of as a nonparametric model as far as it does not have a closed form for the hazard function. This nice characteristic of the PM allows us to use this model to approximate satisfactorily hazard functions of several shapes. For this reason, the PM has been widely used to model time to event data in different contexts, such as;

2. Hospital mortality data Clark and Ryan (2002), and cancer studies including leukemia (Breslow, 1974), gastric cancer (Gamerman, 1991), breast cancer (Ibrahim et al. 2001b) (see also Sinha et al., 1999) for an application to interval-censored data), Melanoma (Kim et al. 2006) and nasopharynx cancer (McKeague and Tighiouart, 2000), among others.
3. The PM has also been used in reliability engineering (Kim and Proschan,1991), (Gamerman, 1994), and economics problems (Gamerman, 1991) and (Bastos et al, 2006).

In this paper we shall modify a Piecewise Weibull hazard baseline function of survival model which can cope better with changes in baseline rate over time, leading to a better fit. This paper
investigate, employing three levels of Weibull distributions as baseline; the effects of ignoring time varying effects and regularized estimation of non-linear functions applied often in prognostic factors.

**Materials and Method**
The risk data used for this paper was simulated from a Weibull baseline hazard distribution which was used to generate survival times for sample sizes of 100, 500 & 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

\[ \lambda_i(t, X) = \lambda_0(t) \exp\left( \sum_{j=1}^{p} \beta_j x_j \right). \]

The baseline hazard rate is unspecified, and assumes that covariates \( x = (x_1, \ldots, x_p) \) act multiplicatively on the hazard rate through the exponential link function (Abiodun, 2007).

This is a re-parameterization of the cox model

\[ \eta(t; w, z, x, s) = f_0(t) + \sum_{j=1}^{p} f_j(t)z_j + x' \gamma \]

Where \( f_0(t) = \log \lambda_0(t) \) which implies, \( \exp(f_0(t)) \), is the baseline function, other aspects of the models include the functions \( f_1(t)z_1 \ldots f_p(t)z_p \) are possibly functional form of time varying covariate \( z_1, \ldots, z_p \) and \( \gamma \) is the usual linear part of the predictor for some categorical covariates (Abiodun, 2009) and (Hennerfeind et al., 2006)

\[ \lambda_{PE}(t; v, x, s) = \{t(t) \exp\{f_0(t)\} + \sum_{j=1}^{p} f_j(t)z_j + f_{spat}(s_{ih}) \}. \]

With its various terms defined as

The function \( f_0 = \log \lambda_0 \), is the baseline effect for the kth interval of PEM

The functions \( f_j(z_{ik}), \ldots, f_p(z_{ip}) \) are functional forms of time varying covariates \( z_{ik}, \ldots, z_{ip} \) in the \( h \)th interval and \( f_{spat}(s_{ih}) \) is a structured spatial effect, where \( s, s = 1, \ldots, S \) is either a spatial index, with \( s = s \) if subject \( i \) in the \( h \)th bit (interval) is from area \( s \) or it is an exact spatial coordinate \( s = (x_s, y_s) \), e.g. for centroids of regions or if exact locations of individuals are known.

**Model Likelihood Function**

\[ L_{PE}\left(\tilde{\gamma}; \beta, D, \Delta, X, s\right) = \prod_{i=1}^{n} \prod_{h=1}^{H} \left( \lambda_h \exp\left( X_i' \beta + s_{ih} \right)^{D_{ih}} \exp\left( \lambda_h \exp\left( X_i' \beta + s_{ih} \right) \Delta_{ih} \right) \right). \]

where for each subject \( i \) there is a product of \( h \) terms, \( H \) 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 \( T_s \) (0 = alive or censored, 1 = failed); \( \Delta_{ih} \) is the time spent in \( T_s \) by the subject. From
expression (3) it may be seen that \( L \) is proportional to the product of Poisson likelihoods for \( D_n \) with mean parameters: \( \mu_{ni} = \lambda_i \exp(\mathbf{x}_i^T \mathbf{\beta} + \mathbf{z}_i \mathbf{a}_i) \). As a consequence, the expression of the Poisson regression model is:

\[
D_{ij} \sim \text{POISSON}(\mu_{ni}); \log(\mu_{ni}) = \alpha_n + \mathbf{x}_i^T \mathbf{\beta} + \mathbf{z}_i \mathbf{a}_i + \log(\Delta_{ni}) \quad \text{5}
\]

Where \( h(i) \) indicate the interval where \( t_i \) falls, i.e. the interval where individual \( i \) died or was censored.

where \( \alpha_n = \log(\lambda_n) \) are log-hazard parameters, and the term \( \log(\Delta_{ni}) \) is an offset.

The expression of the Piecewise model with regularized effects is the following:

\[
\begin{align*}
V_{ij} & \sim \text{POISSON}(\mu_{ij}) \\
\log(\mu_{ij}) & = B_0^T \mathbf{a} + \sum_{j=1}^{p} \mathbf{z}_{ij} \mathbf{B}_0^T \mathbf{r}_{ij} + v_{ij} + \log(\Delta_{ij}) \\
& = (\mathbf{a}^T \mathbf{r}_{ij}) \sim \text{RW}(\tau^2, \mathbf{p}_d); \tau^2 \sim \pi_{\tau^2} \\
(\mathbf{y}_{ij} | \mathbf{r}_{ij}) & \sim \text{RW}(\mathbf{r}_{ij}, \mathbf{p}_d[[j]]); \mathbf{r}_{ij} \sim \pi_{\mathbf{r}_{ij}} j = 1, \ldots, p \\
V_i / \{v_{ij} | j \neq i\} & \sim N\left(-\sum_{\{j \neq i\}} P_{ij} V_{ij} / P_{ii}, \tau^2 / P_{ii}\right)
\end{align*}
\]

The time-dependent effects for each covariate are: \( Z_{1j, i} B_0^T \mathbf{y}_{ij} j = 1, \ldots, p \). Thus, for each \( Z_{0j} \), its values multipled for a piecewise constant function: \( b_j^T \mathbf{y}_{ij} \) in the parameters.

\( Y_{1j} = (Y_{1j, 1}, \ldots, Y_{1j, p})^T \). This enables the effect of each \( Z_{0j} \) to vary in each interval \( T_n \) of the original partition of the follow-up:

\[
B_0^T \mathbf{a} + Z_{1j, i} B_0^T = \alpha_n + Z_{1j, i} Y_{1j, i} \quad \text{for} \quad t \in T_n
\]

**Gaussian Random Field (GRF) priors**

For georeferenced data, it is commonly assumed that \( v_t = v(s_t) \) arises from a Gaussian random field (GRF) \( \{v(s) : s \in S\} \) such that \( v = (v_1, \ldots, v_m) \) follows a multivariate Gaussian distribution as \( v \sim N_m(0, \tau^2 \mathbf{R}) \), where \( \tau^2 \) measures the amount of spatial variation across locations and the \((i, j)\) element of \( \mathbf{R} \) is modeled as \( R[i, j] = \rho(s_i, s_j) \). Here \( \rho(\ldots) \) is a correlation function controlling the spatial dependence of \( v(s) \). In “survregbayes” package in R, the powered exponential correlation function \( \rho(s, s') = \rho(s, s', \phi) = \exp\{-\phi ||s - s'||^2\} \) is used, where \( \phi > 0 \) is a range parameter controlling the spatial decay over distance, \( \phi \in (0, 2] \) is a prespecified shape parameter, and \( ||s - s'|| \) refers to the distance (e.g., Euclidean, great-circle) between \( s \) and \( s' \). Therefore, the prior GRF(\( \tau^2, \phi \)) is defined as

\[
v_i / \{v_{ij} | j \neq i\} \sim N\left(-\sum_{\{j \neq i\}} P_{ij} V_{ij} / P_{ii}, \tau^2 / P_{ii}\right) \quad \text{(7)}
\]

\( l = 1, \ldots, m \) where \( P_{ij} \) is the \((i, j)\) element of \( \mathbf{R}^{-1} \) (Zhou et al, 2017).
Data Analysis

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 (sighted in Adeniyi and Akinrefon, 2018).

Test for Non-Proportionality
To test the hypothesis that the proportional hazard assumption is valid, the following statement of hypothesis is made.

\[ H_0: \delta_1 = \delta_2 = \cdots = \delta_p \text{ (Assumption is valid)} \]
\[ H_1: \text{at least one of the } \delta_i \text{ is not equal to zero (Assumption violated)} \]

Decision rule: Reject \( H_0 \) if \( p-value \leq \alpha \) (level of significance)

For spatial frailty we propose, S= npnorm(v) and v~norm(1, S); if S=npnorm(v) then S~norm, enhanced in simulations via the Mass package in R. Where S is the covariance matrix for spatial correlation in the frailty model.

(Ulviya, 2011), obtained the shape and scale parameters of the Weibull distribution from the formulas below

\[ \eta = \frac{1}{\Gamma(1+\frac{2}{\eta})} \]  
(9)

And

\[ \left( \frac{\Gamma(1+\frac{2}{\eta})}{\Gamma(1+\frac{2}{\eta})} - 1 \right) = 0.5 \]  
(10)

for a convenience choice of mean 1 and variance 0.5. Using the uniroot function in R. parameters were given to be approximately \( \eta = 1.435523 \) and \( \rho = 1.101321 \). 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 below.
2. Investigate the performance of Single hazard models or Single Parametric models (SPM) and the modified Piece-wise model extension or Piece-wise Parametric models (PPM) under various censoring percentages and sample sizes employing their levels of Weibull distributions as baseline.

### Table 1: Shape and scale parameters of the Weibull distributions

<table>
<thead>
<tr>
<th>E(T)</th>
<th>Var(T)</th>
<th>α</th>
<th>η</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.25</td>
<td>2.101377</td>
<td>1.129063</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>1.435523</td>
<td>1.101321</td>
</tr>
<tr>
<td>1</td>
<td>0.75</td>
<td>1.157975</td>
<td>1.052847</td>
</tr>
</tbody>
</table>

The simulation study is to investigate:

1. How the baseline hazards behave under functional forms of time varying effect and continuous covariates in the presence of spatial correlations and
2. Investigate the performance of Single hazard models or Single Parametric models (SPM) and the modified Piece-wise model extension or Piece-wise Parametric models (PPM) under various censoring percentages and sample sizes employing their levels of Weibull distributions as baseline.

#### Model Specification to advance Simulation

**Model 1:** \( \lambda_{PR}(t; z) = f_0(t) + f(t)z_j + f_{z\text{ip}}(s_{ih}) \).

**Model 2:** \( \lambda_{PD}(t; z) = \{l(tT_h(f_0(t))) + f_0(t)z_j + f_{z\text{ip}}(s_{ih}) \} \).

Where \( \lambda_{PR} \) is the hazard function when Partitioning is Ignored (PI) or Single Parametric model (SPM)

Where \( \lambda_{PD} \) is the hazard function when Partitioning is done (PD) or Piece wise Parametric model (PPM)

Simulations 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 and Interpretation of Simulation Study**
Results and Interpretation of Simulation Study

**Table 1**: DIC and LPML of $\beta(t)$ by three (3) levels of Weibull baseline hazard and level of censoring for all sample sizes and $\beta=0.5/\sqrt{t}$ executed for models I & II

<table>
<thead>
<tr>
<th>Partitioning is ignored (PI)</th>
<th>Weibull baseline with low variance of 0.25</th>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>n=100</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No censoring</td>
<td>$\beta$ DIC LPML</td>
<td>$\beta$ DIC LPML</td>
</tr>
<tr>
<td>1.083</td>
<td>900.3891 -450.8276</td>
<td>2.1120 -1.6191 -2.8592 -29.978</td>
</tr>
<tr>
<td>25%</td>
<td>0.9342 671.157 -336.9246</td>
<td>0.8049 -0.0379 0.8172 -9.0184</td>
</tr>
<tr>
<td>50%</td>
<td>1.062 503.4064 -252.7583</td>
<td>0.7330 0.6288 1.1786 -88.591</td>
</tr>
<tr>
<td>75%</td>
<td>1.147 284.3667 -143.1534</td>
<td>0.4906 1.0942 2.4062 -14.162</td>
</tr>
<tr>
<td>Weibull baseline with intermediate variance of 0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PI</td>
<td>$\beta$ DIC LPML</td>
<td>$\beta$ DIC LPML</td>
</tr>
<tr>
<td>1.388</td>
<td>1008.247 -504.8157</td>
<td>2.4477 -1.3555 -2.7600 -3.4787</td>
</tr>
<tr>
<td>25%</td>
<td>1.305 732.8461 -367.927</td>
<td>1.2463 -0.2178 0.5533 -91.743</td>
</tr>
<tr>
<td>50%</td>
<td>1.296 540.0572 -271.5071</td>
<td>0.9155 0.5045 1.6572 -107.66</td>
</tr>
<tr>
<td>75%</td>
<td>1.665 296.9573 -149.734</td>
<td>0.4918 0.4660 10.515 5.8117</td>
</tr>
<tr>
<td>PD</td>
<td>$\beta$ DIC LPML</td>
<td>$\beta$ DIC LPML</td>
</tr>
<tr>
<td>1.598</td>
<td>1078.553 -540.2496</td>
<td>2.5353 -1.1581 -2.5135 -8.2776</td>
</tr>
<tr>
<td>25%</td>
<td>1.542 772.0216 -387.4007</td>
<td>1.6048 -0.4414 1.4104 1.2518</td>
</tr>
<tr>
<td>50%</td>
<td>1.432 570.3035 -286.657</td>
<td>1.0935 -0.1834 2.9965 2.1824</td>
</tr>
<tr>
<td>75%</td>
<td>1.586 311.8024 -157.3363</td>
<td>1.1233 1.1324 2.8323 2.3014</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Partitioning is done</th>
<th>Weibull baseline with low variance of 0.25</th>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>n=500</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No censoring</td>
<td>$\beta$ DIC LPML</td>
<td>$\beta$ DIC LPML</td>
</tr>
<tr>
<td>0.7422</td>
<td>4033.102 -2017.044</td>
<td>0.8626 -0.5374 -0.1370 -2.0944</td>
</tr>
<tr>
<td>25%</td>
<td>0.99 3278.719 -1639.99</td>
<td>1.0717 -0.0804 -0.6797 0.9899</td>
</tr>
<tr>
<td>50%</td>
<td>1.113 2364.521 -1182.72</td>
<td>1.2679 -0.1969 0.30105 -0.6474</td>
</tr>
<tr>
<td>75%</td>
<td>1.201 1243.873 -622.8476</td>
<td>0.9743 0.2131 0.1425 1.2370</td>
</tr>
<tr>
<td>Weibull baseline with intermediate variance of 0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PI</td>
<td>$\beta$ DIC LPML</td>
<td>$\beta$ DIC LPML</td>
</tr>
<tr>
<td>1.061</td>
<td>4686.231 -2343.574</td>
<td>1.2189 -0.7703 0.7253 -2.9976</td>
</tr>
<tr>
<td>25%</td>
<td>1.35 2519.796 -1260.203</td>
<td>1.6592 -0.5801 -1.2489 0.60127</td>
</tr>
<tr>
<td>50%</td>
<td>1.297 2638.277 -1319.477</td>
<td>1.5359 -0.6076 -0.8941 3.3212</td>
</tr>
<tr>
<td>75%</td>
<td>1.397 1364.274 -682.8361</td>
<td>1.4953 -0.3659 -0.9157 15.9160</td>
</tr>
<tr>
<td>PD</td>
<td>$\beta$ DIC LPML</td>
<td>$\beta$ DIC LPML</td>
</tr>
<tr>
<td>1.314</td>
<td>5117.817 -2559.707</td>
<td>1.7013 -0.7955 -1.1526 -0.5874</td>
</tr>
<tr>
<td>25%</td>
<td>1.398 3898.273 -1949.766</td>
<td>1.5475 -0.3632 -0.8892 4.3118</td>
</tr>
<tr>
<td>50%</td>
<td>1.544 2651.088 -1326.259</td>
<td>1.9363 -1.0053 -1.0276 0.3476</td>
</tr>
<tr>
<td>75%</td>
<td>1.739 1390.316 -696.2893</td>
<td>1.8493 -1.1994 1.0844 1.8478</td>
</tr>
</tbody>
</table>
When variance parameters for the Weibull baseline hazard were examined for low at 0.25, moderate or Intermediate at 0.5 & high at 0.75, estimates become worse with increase in variance and sample sizes, reflective in high DIC values and weak predictive power. In all of these, the Piece wise models out-performed the single ones; we again, noticed that the mean posterior estimates were better with increase in censoring percentages.

Table 2, present the mean posterior estimates, DIC and LPML across all sample sizes and censoring percentages for single models and for the modified Piece wise models in the presence of the functional form of Time changing covariate, we observed that the values of estimates when models were fitted with data partitioning having observed the graph of beta against time for appropriate cut points are different (not constant), which indicate a change of effect parameters over time. We observed that the PPMs perform better than the single models throughout the simulations, for all censoring percentages & sample sizes.

**Interpretation**

When variance parameters for the Weibull baseline hazard were examined for low at 0.25, moderate or Intermediate at 0.5 & high at 0.75, estimates become worse with increase in variance and sample sizes, reflective in high DIC values and weak predictive power. In all of these, the Piece wise models out-performed the single ones; we again, noticed that the mean posterior estimates were better with increase in censoring percentages.
Conclusion
We observed that the mean posterior estimates when the PPM - Model II was fitted, indicates change in effect parameters over time in all four intervals, with DIC and LPML values suggesting that PPM performs better than the Single model, for all censoring percentages, sample sizes & for the three (3) levels Weibull baseline. When the Weibull baseline hazard gain spread estimates were worse. In all of these, the PPM out-performed the SPM.

Recommendations
The researcher recommends that:
1. other life distributions should be assumed as baseline to study the behavior of the models
2. combinations of baseline distributions to study competing risk problems

References


