Journal of Advanced Statistics, Vol. 1, No. 1, March 2016
https://dx.doi.org/10.22606/jas.2016.11005

Proportional Hazard Regression Model with Bayesian Approach

Rashmi Aggarwal[1], Suresh Kumar Sharma[1*] and Kanchan Jain[1]

Department of Statistics, Panjab University, Chandigarh, India
Email: ssharma643@yahoo.co.in

Abstract The purpose of survival analysis is to model the underlying distribution of the failure time variable and to assess the dependence of the failure time variable on independent variables. In this paper, we explored PHREG procedure for different models using Bayesian approach. PHREG procedure not only fits COX model but also allows us to fit a piecewise exponential model. The Bayesian analysis treats model parameters as random variables and the inference about these parameters is based on posterior distribution of the parameters. A posterior distribution is a weighted likelihood function of the data with a prior distribution of the parameters using the Bayes' theorem. Generally, for model regression coefficients, normal or uniform prior distributions are used in PHREG procedure. In addition to this, one may specify gamma or improper prior distribution for the scale or variance parameters as well as for hazard parameters in piecewise exponential model. PHREG procedure have been demonstrated with application to real life dataset. Bayesian analysis with PHREG procedure and piecewise constant Bayesian hazard model is also explored along with diagnostic test.

Keywords: Piecewise polynomial, PHREG procedure, Hazard model, Geweke diagnostics.

1 Introduction

The term Survival analysis is used predominantly in biomedical sciences where the interest is in observing time to death of either patient or laboratory animals. An intrinsic characteristic of survival data is the possibility for censoring of observations, that is, the actual time until the event is not observed. Such censoring can arise due to withdrawal from the experiment or termination of the experiment. The time elapsed between enrolment in the study and experiencing one of the events is referred to as patient's survival time. Survival analysis data cannot be analyzed by ignoring censored observations. Therefore, the analysis is carried out using censored as well as uncensored observations. In literature, many censoring situations have been discussed by Kalbfleisch and Prentice [1].

In SAS, there are many procedures for analyzing survival data such as LIFEREG, LIFETEST and PHREG. PROC LIFEREG is a parametric procedure of regression analysis which models the distribution of survival time with a set of covariates. The PROC LIFETEST procedure deals with nonparametric regression for estimating the survival function for comparing the underlying survival curves of two or more samples. Cox proportional hazards model based on semi-parametric approach can be explored by using PROC PHREG procedure.

In this paper, the PHREG procedure for different models using Bayesian approach is discussed. This procedure also allows us to fit a piecewise exponential model. For some basic concepts of Bayesian analysis procedures one may refer to Ibrahim et al. [2], Gelman et al. [3], Gilks et al. [4].

Some basic definitions which are required to cary out PHREG procedure are given below.

Definition 1.1. Markov Chain Monte Carlo (MCMC) Method: The simulation method for sampling from posterior distributions which computes posterior quantities of interest is called the Markov Chain Monte Carlo (MCMC) method. In this method, each sample depends on the previous sample. A Markov chain is a sequence of random variables for which the random variable θ^t depends on all previous θs only through its immediate predecessor θ^{t-1}. Monte Carlo integration is mainly used to approximate an expectation by using the Markov chain samples.

Definition 1.2. Gibbs Sampler: Gibbs sampling requires decomposing the joint posterior distribution into full conditional distributions for each parameter in the model and then sampling from them. If the

parameters are not dependent on each other, then the sampler is regarded as efficient and full conditional distributions are easy to sample from. It may be noted that the conditional distributions can be easily derived but it is not always possible to find an efficient way to sample from these conditional distributions.

Definition 1.3. Burn-in, Thinning and Markov Chain Samples: To discard the initial portion of a Markov chain, so that the effect of initial values on the posterior inference is minimized, we use Burn-in procedure. In practice, if the chain has reached its target distribution after 't' iterations, then one can use good samples for posterior inference by throwing out the early portion. The value of 't' is the burn-in number. The thinning number controls the thinning of the Markov chain samples.

Definition 1.4. Autocorrelations: The sample autocovariance function of lag h for $\{\theta_i^t\}$ is defined by

$$\hat{\gamma}(h) = \frac{1}{n-h}\sum_{t=1}^{n-h}(\theta_i^{t+h} - \bar{\theta}_i), 0 \leq h < n$$

The sample autocorrelation of lag h is defined in terms of the sample autocovariance function as

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, |h| < n$$

Definition 1.5. Effective Sample Size: We can use autocorrelation and trace plots to examine the mixing of a Markov chain. A closely related measure of mixing is the effective sample size (ESS). It is defined as follows:

$$ESS = \frac{n}{\tau} = \frac{n}{1 + 2\sum_{k=1}^{\infty}\rho_k(\theta)}$$

where n is the total sample size and $\rho_k(\theta)$ is the autocorrelation of lag k for θ. The quantity τ is referred to as the autocorrelation time.

Section 2 presents an overview of PHREG procedure with application to real life dataset. Bayesian analysis with PHREG procedure along with diagnostic test is discussed in Section 3. Section 4 includes piecewise constant Bayesian hazard model with practical example.

2 PHREG Procedure

The PHREG procedure performs regression analysis of survival data based on the Cox proportional hazards model. Cox's Semi-parametric model is widely used in the analysis of survival data to explain the effect of explanatory variables on hazard rates. The survival time of each member of a population is assumed to follow its own hazard function, $h_i(t)$, expressed as

$$h_i(t) = h_i(t : Z_i) = h_0(t)\exp(Z_i'\beta)$$

where $h_0(t)$ is an arbitrary and unspecified baseline hazard function, Z_i is the vector of explanatory variables for the i^{th} individual and β is the vector of unknown regression parameters associated with the explanatory variables. The survivor function can be expressed as

$$S(t : Z_i) = [S_0(t)]^{\exp(Z_i'\beta)}$$

where $S_0(t) = \exp\Big(-\int_0^t h_0(u)du\Big)$ is the baseline survivor function. To estimate β, Cox [5][6] introduced the partial likelihood function, which eliminates the unknown baseline hazard $h_0(t)$ and accounts for censored survival times.

The PHREG procedure provides four selection methods viz. forward selection, backward elimination, stepwise selection, and best subset selection. The best subset selection method is based on the likelihood score statistic. This method identifies a specified number of best models containing one, two or three variables and so on, up to the single model containing all of the explanatory variables. The PHREG procedure uses ODS Graphics to create graphs as part of its output.

2.1 Post PK Glaucoma Analysis Using PHREG Procedure

The retrospective analysis of 233 patients who underwent optical penetrating keratoplasty (PK) and had a minimum follow-up of 3 months, has been carried out using PHREG procedure. Data on post-operative intra-ocular pressure (IOP) was recorded at 3, 6, 9, 12 and 18 months or more. All the patients were followed up for post PK glaucoma and time in months was recorded (follow-up variable). There were 28 censored observations. Various risk factors including age, sex, indications for PK, type of surgical procedures and additional surgical procedures performed have been included in the study. The censored observations are marked "0", if censored and "1", if not censored. The data includes variables follow-up period (in months), the status variable (the censoring indicator variable 0, if censored and 1 if not censored) and the variable group (age$\leq$40, coded as 0 and >40, as 1).

PROC PHREG fits Cox model by maximizing the partial likelihood and computes the base-line survivor function by using the Breslow [7] estimate. Since there are only two groups and the null hypothesis for no difference between the two groups is identical to the null hypothesis that the regression coefficient for group is 0. Thus, the global null hypothesis is $H_0 : \beta = 0$. The hazard ratio estimate is 1.472 which implies that hazard function for group 1 is smaller than group 0. In other words, group 0 patients (age$\leq$40) have less chances of developing post PK glaucoma than those in group 1 (age>40). The analysis is shown in Table 1 and survival curves in Fig. 1.

Table 1. The PHREG procedure.

The PHREG Procedure

Model Information

Data Set	OUT.COX_DATA
Dependent Variable	Followup
Censoring Variable	status
Censoring Value(s)	0
Ties Handling	BRESLOW

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
233	205	28	12.02

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	1846.126	1839.565
AIC	1846.126	1841.565
SBC	1846.126	1844.888

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	6.5606	1	0.0104
Score	6.3332	1	0.0118
Wald	6.2551	1	0.0124

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
Group	1	0.38673	0.15463	6.2551	0.0124	1.472

Note that if there are no ties in the survival times, the likelihood score test in the COX regression analysis is equivalent to the log-rank test.

3 Bayesian Analysis Using PHREG Procedure

Bayesian methods incorporate existing information (based on expert knowledge, past studies, and so on) into your current data analysis. This existing information is represented by a prior distribution and the

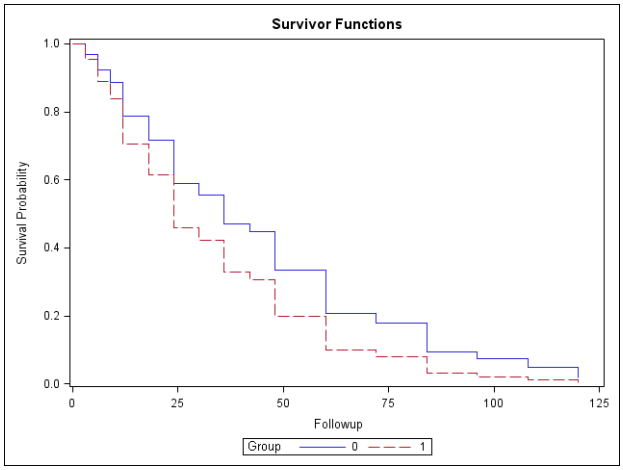

Figure 1. Comparison of two survival curves.

data likelihood is effectively weighted by the prior distribution as the data analysis results are computed. The main outcomes of a Bayesian analysis are the posterior distributions of a model's parameters, rather than point estimates and their standard errors. Access to a model's parameters posterior distributions enables us to address scientific questions of interest directly. This happens since the model parameters are estimated, it is easy to compute the posterior distributions for any functions of the parameters or any quantities of interest. Some statisticians produce Bayesian analyses simply to operate in the Bayesian framework. Before the current data are examined, the uncertainty about the parameters is judged by the prior distribution. Multiplication of likelihood function with prior distribution leads to posterior distribution of the parameter. Posterior distributions are used to carry out all inferences and modelling procedures in Bayesian analysis.

The degree of belief in a random event is attributed to the Bayesian probability measures and these measures are highly subjective. Since long, there has been a desire to obtain results that are objectively valid. Within the Bayesian paradigm, this can be somewhat achieved by using prior distributions that have a minimal impact on the posterior distribution. Such distributions are called objective or noninformative priors, one may refer to DeGroot and Schervish [8] and Press [9]. One may also refer to Berger [10] and Goldstein [11] for information about objective Bayesian versus subjective Bayesian analysis.

A prior $\pi(\theta)$ is noninformative if it has minimal impact on the posterior distribution of θ. The noninformative priors are also called flat priors. In some cases, noninformative priors can lead to improper posteriors. Moreover, noninformative priors are generally variant under transformation, that is, a prior might be noninformative in one parameterization but not necessarily noninformative if a transformation is applied.

If the prior distribution has an impact on posterior distribution and dominates the likelihood, then it is called an informative prior. The power of the Bayesian method is based on the proper use of prior distributions where the current information also includes the information gathered from pervious study, past experience or expert opinion.

3.1 Diagnostic Test for Bayesian Procedure

The PHREG procedure's Bayesian analysis capabilities enable us to do the following:

- fit a Cox proportional hazards model
- fit piecewise constant baseline hazard models (also known as piecewise exponential models)
- fit a superset of the Cox model, known as the multiplicative hazards model (also known as Anderson-Gill model)
- estimate customized hazard ratios
- estimate the survival function

– fit multinomial logit choice models for discrete data

The PHREG procedure supports the following priors

Parameter	Prior
Regression coefficients	Normal, uniform
Baseline hazards (original scale)	Improper, uniform, gamma, independent gamma, Auto Regressive AR(1)
Baseline hazards (log scale)	Uniform, normal
Log-hazards and regression coefficients	Joint multivariate normal

For a Cox model, the model parameters are the regression coefficients. For a piecewise exponential model, the model parameters consist of the regression coefficients and the hazards or log-hazards.

The Geweke test [12] compares values in the early part of the Markov chain to those in the latter part of the chain in order to detect failure of convergence.

The DIC is based on posterior density which means that it takes care of the prior information. It is a model assessment tool and a Bayesian alternative to Akaike's information criterion (AIC) and the Bayesian information criterion (BIC). Calculation of the DIC in Monte Carlo Markov Chain method is trivial and does not require maximization over the parameter space. However, the maximization is required for AIC and BIC criteria. A smaller DIC indicates a better fit to the data set. pD is the effective number of parameters which is equal to the difference between the measure of fit and the deviance evaluated at the estimates.

3.2 PHREG Bayesian Procedure for Post PK Glaucoma

The partial likelihood of the COX model in PHREG procedure generates a chain of posterior distribution samples by Gibbs sampler. The data considered in Section 2.1 has been analysed using Bayesian analysis. Bayesian analysis in SAS can be invoked by BAYES statement. The SEED option is specified to maintain reproducibility and the OUTPOST option saves the posterior distribution samples in a SAS dataset for post processing. By default, a uniform distribution is assumed on the regression coefficient group. However, if we use informative prior on regression coefficients, then COEFPRIOR option is required. The model information is provided in Table 2 along with classical parameter estimates, coefficient prior and fit statistics.

PHREG first fits the Cox model by maximizing the partial likelihood along with 95% confidence intervals for regression parameters. Note that no prior is specified for regression coefficients and therefore default uniform prior is used for regression coefficients. The "Fit Statistics" gives the information about fitted model in terms of DIC (Deviance Information Criterion) and pD (effective number of parameters). The summary statistics of posterior samples and posterior intervals are shown in Table 3.

We observe that in summary statistics, mean and standard deviation of the posterior samples are comparable to MLE and its standard error, due to the use of uniform prior in Table 2. This suggests that the posterior samples, using uniform prior for regression coefficients, produce the same results as given by MLE technique.

PHREG provides diagnostics to access the convergence of the generated Markov Chain. Table 4 shows three diagnostics:

– posterior autocorrelations at Lag1, Lag5, Lag10, Lag50
– Geweke diagnostics, and
– effective sample size (ESS).

From values in Table 4, it can be seen that posterior autocorrelations from Lag1 - Lag50 are almost negligible and it can also be verified by autocorrelation plot shown in Fig. 2.

Trace plots of samples versus the simulation index can be useful in assessing convergence and also shows that the chain is mixing well. It may be noted that if the chain does not converge to its stationary distribution, then there will be long burn-in period. One can observe from a trace plot that there

Table 2. Model information for PHREG procedure.

The PHREG Procedure
Model Information

Uniform Prior for Regression Coefficients

Data Set	OUT.COX_DATA
Dependent Variable	Followup
Censoring Variable	status
Censoring Value(s)	0
Model	Cox
Ties Handling	BRESLOW
Burn-In Size	2000
MC Sample Size	10000
Thinning	1

Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits	
Group	1	0.3867	0.1546	0.0837	0.6898

Fit Statistics

DIC (smaller is better)	1841.538
pD (Effective Number of Parameters)	0.986

Table 3. Summary statistics.

The PHREG Procedure
Posterior Summaries

Parameter	N	Mean	Standard Deviation	Percentiles 25%	50%	75%
Group	10000	0.3898	0.1538	0.2870	0.3898	0.4913

Posterior Intervals

Parameter	Alpha	Equal-Tail Interval		HPD* Interval	
Group	0.050	0.0874	0.6884	0.0904	0.6898

*HPD: Highest Posterior Density

Table 4. Convergence diagnostics and mean procedure.

The PHREG Procedure

Bayesian Analysis
Posterior Autocorrelations

Parameter	Lag 1	Lag 5	Lag 10	Lag 50
Group	-0.0062	0.0024	-0.0056	0.0104

Geweke Diagnostics

Parameter	z	Pr > |z|
Group	0.2149	0.8298

Effective Sample Sizes

Parameter	ESS	Correlation Time	Efficiency
Group	10000.0	1.0000	1.0000

P(hazard(group 1) < hazard(group 0))

The Means Procedure
Analysis Variable : Indictor Group < 0

N	Mean
1000	0.982

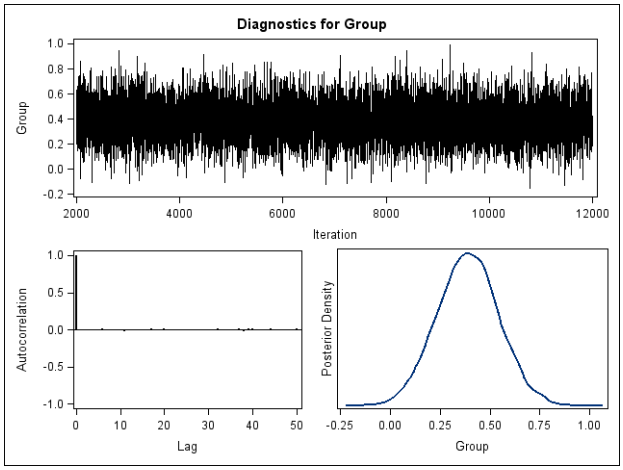

Figure 2. Diagnostic plots.

is a relatively constant mean and variance in case of stationarity. Fig. 2 displays the trace plot, the autocorrelation function plot and posterior density plot generated by Markov Chain.

On the basis of Fig. 2, we conclude that

- Markov Chain (MC) has reached convergence
- MC has reached stationarity as the distribution of points is not changing as the chain progresses
- the burn-in size is 2000 which means that the initial 2000 observations have been discarded
- trace plot is perfect and the centre of the chain is around 0.4 with small fluctuations which indicates that the MC has reached the right distribution
- the chain is mixing well as it is exploring the distribution by traversing to areas where its density is very low.

The proportional hazards model for comparing the two groups is

$$h(t) = \begin{cases} h_0(t) & \text{if } group = 0 \\ h_0(t)e^{\beta} & \text{if } group = 1 \end{cases}$$

The probability that the hazard of group = 1 is less than that of group = 0 is written as

$$P(h_0(t)e^{\beta} < h_0(t)) = P(\beta < 0).$$

The probability for posterior distribution samples can be worked out by taking into consideration those samples whose coefficient is less than 0. For our model, $P(hazard(group = 1) < hazard(group = 0))$, computed by using PROC MEANS procedure is 0.982 (Table 4). Thus, there is 98.2% chance that the hazard rate of $group = 0$ is less than that of $group = 1$. This is also consistent with the fact that the average survival time of $group = 0$ is less that that of $group = 1$.

4 Piecewise Constant Baseline Hazard Model

We consider different priors for piecewise exponential model. For Cox model, the model parameters are regression coefficients. For piecewise exponential model, parameters are the regression coefficients and hazard functions. The priors for the hazard functions and the regression coefficients are assumed to be independent.

The Bayesian analysis treats model parameters as random variables and the inference about these parameters is based on posterior distribution of the parameters. Although it is difficult to obtain the closed form of the posterior distribution, however MCMC method is used to simulate samples from the posterior distribution. The procedure for conducting Bayesian analysis is described below:

For single failure time variable, let $\{(t_i, x_i, \delta_i), i = 1, 2, \ldots, n\}$ be the observed data, where t_i is the failure time associated with the covariate x and δ_i indicates whether the failure time is censored or not. Let $a_0 = 0 < a_1 < \cdots < a_{J-1} < a_J = \infty$. For piecewise baseline hazard model, we consider hazard function in original scale and estimate the parameters using maximum likelihood technique and also the information matrix.

The hazard function for subject i is

$$h(t|x_i;\beta) = h_0(t)\exp(\beta' x_i), i = 1, 2, \ldots, n$$

where $h_0(t) = \lambda_j, a_{j-1} \le t < a_j (j = 1, \ldots, J)$.

In fact, we assume that the baseline hazard is constant within each interval. The baseline hazard is characterized using J parameters, that is, $\lambda = (\lambda_1, \ldots, \lambda_J)'$.

The baseline cumulative hazard function is

$$H_0(t) = \sum_{j=1}^{J} \lambda_j \Delta_j(t)$$

where $\Delta_j(t) = \begin{cases} 0 & t - a_{j-1} \\ t - a_{j-1} & a_{j-1} \le t < a_j \\ a_j - a_{j-1} & t \ge a_j \end{cases}$.

4.1 Estimation of λ and β

The log likelihood function involving parameters λ and β is given by

$$\begin{aligned} l(\lambda, \beta) &= \sum_{i=1}^{n} \delta_i \left[\sum_{j=1}^{J} I(a_{j-1} \le t_i < a_j) log\lambda_j + \beta' x_i \right] - \sum_{i=1}^{n} \left[\sum_{j=1}^{J} \Delta_j(t_i)\lambda_j \right] \exp(\beta' x_i) \\ &= \sum_{j=1}^{J} d_j log\lambda_j + \sum_{i=1}^{n} \delta_i \beta' x_i - \sum_{j=1}^{J} \lambda_j \left[\sum_{i=1}^{n} \Delta_j(t_i)\exp(\beta' x_i) \right] \end{aligned} \tag{1}$$

where $d_j = \sum_{i=1}^{n} \delta_i I(a_{j-1} \le t_i < a_j)$.

For fixed $\beta, \frac{\partial l}{\partial \lambda} = 0$ gives

$$\widetilde{\lambda_j}(\beta) = \frac{d_j}{\sum_{i=1}^{n} \Delta_j(t_i)\exp(\beta' x_i)} \quad (j = 1, \ldots, J).$$

It is difficult to obtain the estimate of parameter β, but one may use the profile log likelihood for β by substituting the values of $\widetilde{\lambda}$ into the likelihood function $l(\lambda, \beta)$. Thus, the profile likelihood of β will be

$$l_p(\beta) = \sum_{i=1}^{n} \delta_i \beta' x_i - \sum_{j=1}^{J} d_j log \left[\sum_{l=1}^{n} \Delta_j(t_l)\exp(\beta' x_l) \right] + c$$

where $c = \sum_j (d_j log d_j - d_j)$.

Since the constant c does not depend on β and hence, it can be discarded from $l_p(\beta)$.

The MLE $\hat{\beta}$ of β is obtained by maximizing

$$l_p(\beta) = \sum_{i=1}^{n} \delta_i \beta' x_i - \sum_{j=1}^{J} d_j log \left[\sum_{l=1}^{n} \Delta_j(t_l)\exp(\beta' x_l) \right]$$

with respect to β and the MLE $\hat{\lambda}$ of λ is given by $\hat{\lambda} = \widetilde{\lambda}(\hat{\beta})$.

Priors for Piecewise Exponential Model

By assuming $\lambda = (\lambda_1, \ldots, \lambda_J)'$ to be the constant baseline hazards, the following priors have been used for regression parameters.

Improper Prior: The joint prior density is given by

$$p(\lambda_1, \ldots, \lambda_J) = \prod_{j=1}^{J} \frac{1}{\lambda_j} \quad \forall \lambda_j > 0.$$

This prior is improper (nonintegrable), but the posterior distribution is proper as long as there is at least one event time in each of the constant hazard intervals.

Uniform Prior: The joint prior density is given by

$$p(\lambda_1, \ldots, \lambda_J) \propto 1 \quad \forall \lambda_j > 0$$

This prior is improper (nonintegrable), but the posteriors are proper as long as there is at least one event time in each of the constant hazard intervals.

Gamma Prior: The Gamma distribution $G(a, b)$ has a pdf

$$f_{a,b}(t) = \frac{b(bt)^{a-1}e^{-bt}}{\Gamma(a)}, t > 0$$

where a is the shape parameter and b^{-1} is the scale parameter. The mean and variance are $\frac{a}{b}$ and $\frac{a}{b^2}$ respectively.

Independent Gamma Prior: For $j = 1, \ldots, J$, λ_j has an independent $G(a_j, b_j)$ prior and the joint prior density is given by

$$p(\lambda_1, \ldots, \lambda_J) \propto \prod_{j=1}^{J} \left\{ \lambda_j^{a_j - 1} e^{-b_j \lambda_j} \right\} \quad \forall \lambda_j > 0.$$

Auto Regressive (AR) Prior: $\lambda_1, \ldots, \lambda_J$ are correlated as follows:

$$\begin{aligned}
&\lambda_1 \sim G(a_1, b_1) \\
&\lambda_2 \sim G\Big(a_2, \frac{b_2}{\lambda_1}\Big) \\
&\ldots \\
&\lambda_J \sim G\bigg(a_J, \frac{b_J}{\lambda_{J-1}}\bigg).
\end{aligned}$$

The joint prior density is given by

$$p(\lambda_1, \ldots, \lambda_J) \propto \lambda_1^{a_1 - 1} e^{-b_1 \lambda_1} \prod_{j=2}^{J} \Big(\frac{b_j}{\lambda_{j-1}}\Big)^{a_j} \lambda_j^{a_{j-1}} e^{-\frac{b_j}{\lambda_{j-1}} \lambda_j}.$$

4.2 Bayesian Analysis of Piecewise Exponential Model for Post PK Glaucoma

To illustrate the Bayesian analysis of piecewise exponential model, we again consider the same data as described in Section 4.1 with 28 censored observations. But in this case, the results are given for two groups viz. group1 as 'age' (age ≤ 40 coded '0' and age > 40 coded '1') and group2 as 'sex' (male, coded '0' and female, coded '1') separately. The results have been compared by using DIC and pD criteria, by considering 4 different priors viz. improper, uniform, Gamma and AR Gamma. The option PIECEWISE = HAZARD in SAS is used for modelling of constant hazard in original scale according to the procedure defined in Section 4.

The comparison in terms of DIC and effective number of parameters (pD) for both the groups (age and sex) are presented in Table 5.

Table 5. DIC and pD for different priors.

Group	Status							
	Improper		Uniform		Gamma		AR Gamma	
	pD	DIC	pD	DIC	pD	DIC	pD	DIC
Group1(Age)	8.82	1808.81	8.14	1811.59	8.66	1808.91	8.44	1808.15
Group2(Sex)	8.92	1815.99	8.26	1818.02	8.75	1815.90	8.55	1815.28

Table 6. The PHREG procedure for AR Gamma prior.

Bayesian Analysis
For Group1 (age) and Group2 (sex)

Model Information		Constant Hazard Time Intervals				
Data Set	OUT.COX_DATA	Interval				Hazard
Dependent Variable	Followup	[Lower,	Upper)	N	Event	Parameter
Censoring Variable	status	0	15	86	60	Lambda1
Censoring Value(s)	0	15	21	19	18	Lambda2
Model	Piecewise Exponential	21	27	34	33	Lambda3
Burn-In Size	2000	27	33	7	7	Lambda4
MC Sample Size	10000	33	45	23	23	Lambda5
Thinning	1	45	54	24	24	Lambda6
		54	66	24	24	Lambda7
		66	Infty	16	16	Lambda8

Since DIC (Deviance Information criterion) is lower for AR Gamma prior and hence the results for this prior are presented for both the groups, that is, age and sex in Table 6.

The table gives the information for piecewise exponential model along with constant hazard time intervals (for both groups). By default, the time axis is partitioned into eight intervals of constant hazards. The number of events and observations are also shown in the Table 6. Note that the constant hazard parameters are named as Lambda1,. . . , Lambda8.

The maximum likelihhod estimates of Lambda1, . . . , Lambda8 for both the groups are presented in Table 7.

The model parameters include eight hazard parameters Lambda1,. . . , Lambda8, and the AR Gamma prior for regression coefficients for age as well as sex. The maximum likelihood estimates (Ref. Section 4.1) obtained in Table 7, are used as the starting values for simulation of the posterior distribution.

Summary statistics for all model parameters along with first and third quartile are shown in Table 8.

Table 9 shows posterior intervals along with highest posterior density (HPD) intervals.

The posterior autocorrelations namely, Lag1, Lag5, Lag10, Lag50 for both the groups are shown in Table 10. We observe that as the lag increases, the values of autocorelations decrease and becomes negligible for Lag50 in both the groups.

The Geweke diagnostics result are presented in Table 11.

The values in the table show non-significant results for group2, for all hazard parameters. This means that there is good mixing of the Markov chain. However, the results are significant for group1 which means that the chain has not reached the right distribution and the chain is not mixing well. One may conclude that we may need higher burn-in size for convergence for group1.

The effective sample size (ESS) along with correlation time and efficiency are shown in Table 12.

The corelation time for sex is 15.1724 with efficiency 0.0659, however, for age the correlation time is 25.8846 with efficiency 0.0386. The efficiency is high and correlation time is low for sex as compared to age. One of the reasons for this is that the distribution of sex has reached convergence while the distribution of age has not reached convergence.

Fig. 3 displays the diagnostic plots for age and sex.

It is evident from trace plot that the centre of the chain is around 0.0 for sex, with small fluctuations. However, chain does not converge to a proper point for age. The autocorrelations are high for age as compared to sex at least up to Lag10.

Table 7. Maximum likelihood estimates for AR Gamma prior.

Maximum Likelihood Estimates (For Group1, Age)					
Parameter	DF	Estimate	Standard Error	95% Confidence Limits	
Lambda1	1	0.0106	0.0032	0	0.02
Lambda2	1	0.0111	0.0039	0	0.02
Lambda3	1	0.0253	0.008	0.01	0.04
Lambda4	1	0.0066	0.003	0	0.01
Lambda5	1	0.0139	0.0046	0	0.02
Lambda6	1	0.029	0.0095	0.01	0.05
Lambda7	1	0.0376	0.0122	0.01	0.06
Lambda8	1	0.0267	0.0093	0.01	0.05
Age	1	0.4009	0.1538	0.1	0.7

Maximum Likelihood Estimates (For Group2, Sex))					
Parameter	DF	Estimate	Standard Error	95% Confidence Limits	
Lambda1	1	0.0201	0.0047	0.01	0.03
Lambda2	1	0.0207	0.0064	0.01	0.03
Lambda3	1	0.0471	0.0124	0.02	0.07
Lambda4	1	0.0123	0.0052	0	0.02
Lambda5	1	0.0254	0.0073	0.01	0.04
Lambda6	1	0.0528	0.0151	0.02	0.08
Lambda7	1	0.0679	0.0194	0.03	0.11
Lambda8	1	0.047	0.0148	0.02	0.08
Sex	1	0.0376	0.1459	-0.25	0.32

Table 8. Posterior summary for AR Gamma prior.

Posterior Summary (For Group1, Age)						
Parameter	N	Mean	Standard Deviation	Percentiles		
				25%	50%	75%
Lambda1	10000	0.0111	0.0032	0.0088	0.0107	0.0131
Lambda2	10000	0.0117	0.004	0.0088	0.0112	0.014
Lambda3	10000	0.0259	0.008	0.0201	0.0249	0.0308
Lambda4	10000	0.0074	0.0032	0.005	0.0068	0.0091
Lambda5	10000	0.0145	0.0047	0.0111	0.0139	0.0172
Lambda6	10000	0.0299	0.0096	0.0231	0.0286	0.0355
Lambda7	10000	0.0387	0.0124	0.0299	0.0372	0.0458
Lambda8	10000	0.0278	0.0096	0.0209	0.0265	0.0333
Age	10000	0.3938	0.1491	0.2889	0.3899	0.4922

Posterior Summary (For Group2, Sex)						
Parameter	N	Mean	Standard Deviation	Percentiles		
				25%	50%	75%
Lambda1	10000	0.0205	0.0046	0.0173	0.0201	0.0232
Lambda2	10000	0.0215	0.0064	0.0169	0.0207	0.0252
Lambda3	10000	0.0472	0.0122	0.0384	0.0458	0.0544
Lambda4	10000	0.0133	0.0054	0.0094	0.0125	0.0163
Lambda5	10000	0.0258	0.0072	0.0208	0.025	0.03
Lambda6	10000	0.0533	0.0147	0.0426	0.0515	0.0622
Lambda7	10000	0.0687	0.0192	0.0547	0.0664	0.08
Lambda8	10000	0.0479	0.015	0.0371	0.0462	0.0565
Sex	10000	0.0345	0.138	-0.055	0.0358	0.1255

Table 9. Posterior intervals for AR Gamma prior.

Posterior Intervals (For Group1, Age)					
Parameter	Alpha	Equal-Tail Interval		HPD Interval	
Lambda1	0.05	0.0058	0.0182	0.0054	0.0177
Lambda2	0.05	0.0055	0.0213	0.005	0.02
Lambda3	0.05	0.0131	0.0442	0.0116	0.0416
Lambda4	0.05	0.0027	0.0151	0.0023	0.0139
Lambda5	0.05	0.0071	0.0253	0.0063	0.0238
Lambda6	0.05	0.0149	0.0523	0.0127	0.0487
Lambda7	0.05	0.0193	0.0678	0.0176	0.0642
Lambda8	0.05	0.013	0.0502	0.0109	0.0468
Age	0.05	0.1143	0.6998	0.1091	0.6889

Posterior Intervals (For Group2, Sex)					
Parameter	Alpha	Equal-Tail Interval		HPD Interval	
Lambda1	0.05	0.0129	0.0312	0.0123	0.0299
Lambda2	0.05	0.0114	0.0363	0.0102	0.0343
Lambda3	0.05	0.0272	0.075	0.0254	0.0712
Lambda4	0.05	0.0054	0.0258	0.0044	0.0239
Lambda5	0.05	0.0144	0.0427	0.0128	0.0401
Lambda6	0.05	0.0294	0.0867	0.0275	0.0832
Lambda7	0.05	0.038	0.1123	0.0343	0.106
Lambda8	0.05	0.0241	0.0822	0.0222	0.0784
Sex	0.05	-0.2416	0.3055	-0.2494	0.2958

Table 10. Posterior autocorrelations for AR Gamma prior.

Posterior Autocorrelations (For Group1, Age)				
Parameter	Lag 1	Lag 5	Lag 10	Lag 50
Lambda1	0.7935	0.5925	0.4184	-0.0244
Lambda2	0.543	0.4066	0.2872	-0.0215
Lambda3	0.6578	0.4936	0.3439	-0.0314
Lambda4	0.3395	0.231	0.1695	-0.0088
Lambda5	0.5757	0.4343	0.3002	-0.0212
Lambda6	0.5942	0.4346	0.3087	-0.0263
Lambda7	0.5653	0.4234	0.2945	-0.0213
Lambda8	0.4497	0.3356	0.2257	-0.0122
Age	0.9301	0.7047	0.506	-0.0386

Posterior Autocorrelations (For Group2, Sex)				
Parameter	Lag 1	Lag 5	Lag 10	Lag 50
Lambda1	0.67	0.3954	0.2122	0.0277
Lambda2	0.4113	0.2393	0.1238	0.0134
Lambda3	0.5208	0.307	0.1575	0.0175
Lambda4	0.22	0.1317	0.0768	0.0135
Lambda5	0.443	0.2685	0.1391	0.0307
Lambda6	0.4695	0.271	0.1403	0.0213
Lambda7	0.466	0.2757	0.1422	0.0159
Lambda8	0.3401	0.2052	0.1156	0.0048
Sex	0.8784	0.523	0.277	0.0242

Table 11. Geweke diagnostics for AR Gamma prior.

Geweke Diagnostics					
(For Group1, Age)			(For Group2, Sex)		
Parameter	z	Pr > |z|	Parameter	z	Pr > |z|
Lambda1	-2.1878	0.0287	Lambda1	0.1356	0.8921
Lambda2	-2.4979	0.0125	Lambda2	-0.3773	0.706
Lambda3	-2.1979	0.028	Lambda3	0.2583	0.7962
Lambda4	-2.1204	0.034	Lambda4	-0.2774	0.7815
Lambda5	-1.9887	0.0467	Lambda5	-0.1486	0.8818
Lambda6	-2.0735	0.0381	Lambda6	0.054	0.957
Lambda7	-2.6589	0.0078	Lambda7	-0.1637	0.8699
Lambda8	-2.1701	0.03	Lambda8	-0.0607	0.9516
Age	2.2752	0.0229	Sex	-0.1352	0.8924

Table 12. Effective sample sizes for AR Gamma prior.

Effective Sample Sizes							
(For Group1, Age)				(For Group2, Sex)			
Parameter	ESS	Correlation Time	Efficiency	Parameter	ESS	Correlation Time	Efficiency
Lambda1	461.4	21.6716	0.0461	Lambda1	854.2	11.7075	0.0854
Lambda2	677.5	14.7608	0.0677	Lambda2	1352.4	7.3944	0.1352
Lambda3	530.7	18.8436	0.0531	Lambda3	1042.8	9.5897	0.1043
Lambda4	1032.1	9.6891	0.1032	Lambda4	2314.4	4.3208	0.2314
Lambda5	641.2	15.5956	0.0641	Lambda5	1228.1	8.1424	0.1228
Lambda6	631.5	15.8349	0.0632	Lambda6	1158.6	8.6308	0.1159
Lambda7	665.2	15.034	0.0665	Lambda7	1185.6	8.4344	0.1186
Lambda8	806.7	12.3956	0.0807	Lambda8	1463	6.8355	0.1463
Age	386.3	25.8846	0.0386	Sex	659.1	15.1724	0.0659

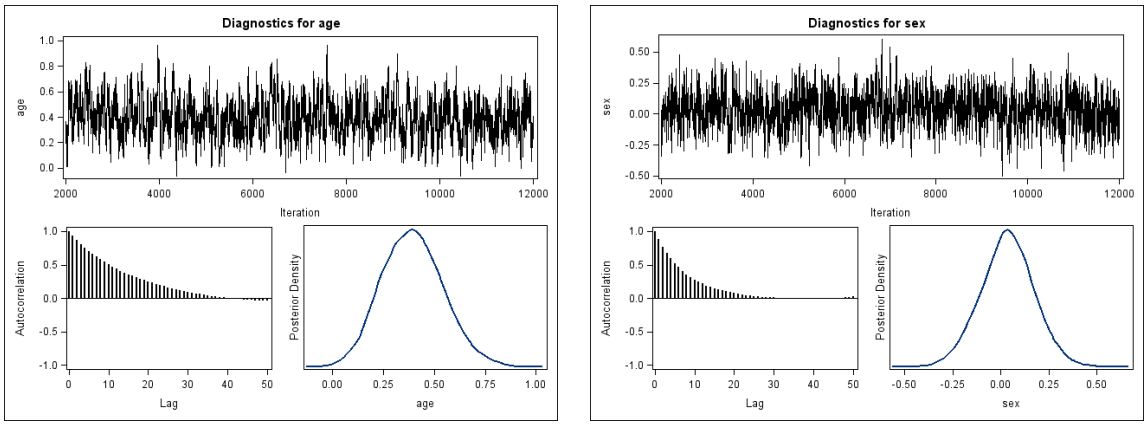

Figure 3. Diagnostic plots for age and sex.

It may be concluded that based on DIC and pD criteria, the AR Gamma prior performs better than all other priors viz. improper, uniform and gamma in PHREG Bayesian analysis for piecewise exponential model. Also trace plots, posterior autocorrelations, lags, correlation time and efficiency shows that group2 (sex) has reached the right distribution.

References

1. J. D. Kalbfleisch and R. L. Prentice, *The Statistical Analysis of Failure Time Data.* John Wiley & Sons, New York, 1980.
2. J. G. Ibrahim, M. H. Chen, and D. Sinha, *Bayesian Survival Analysis.* Springer-Verlag, New York, 2001.

3. A. Gelman, J. Carlin, H. Stern, and D. Rubin, *Bayesian Data Analysis, Second Addition*. Chapman & Hall, London, 2004.
4. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London, 1996.
5. D. R. Cox, "Regression models and life tables," *Journal of the Royal Statistical Society, Series B*, vol. 20, pp. 187–220, 1972.
6. ——, "Partial likelihood," *Biometrika*, vol. 62, pp. 269–276, 1975.
7. N. E. Breslow, "Discussion of professor cox's paper," *Journal of Royal Statistical Society B*, vol. 34, pp. 216–217, 1972.
8. M. H. DeGroot and M. J. Schervish, *Probability and Statistics, 3rd Edition*. Addison-Wesley, MA, Boston, 2002.
9. S. J. Press, *Subjective and Objective Bayesian Statistics*. John Wiley & Sons, New York, 2003.
10. J. O. Berger, "The case for objective bayesian analysis," *Bayesian Analysis*, vol. 3, pp. 385–402, 2006.
11. M. Goldstein, "Subjective bayesian analysis: Principles and practice," *Bayesian Analysis*, vol. 3, pp. 403–420, 2006.
12. J. Geweke, *Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments*. Oxford University Press, Oxford, 1992.

