Bayesian Analysis of Extended Cox Model with Time-Varying Covariates Using Bootstrap Prior

Oyebayo R. Olaniran  
*University of Ilorin, rid4stat@yahoo.com*

Mohd Asrul A. Abdullah  
*Universiti Tun Hussein Onn Malaysia, afendi@uthm.edu.my*

Follow this and additional works at: [https://digitalcommons.wayne.edu/jmasm](https://digitalcommons.wayne.edu/jmasm)

Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

**Recommended Citation**  
Bayesian Analysis of Extended Cox Model with Time-Varying Covariates Using Bootstrap Prior

Cover Page Footnote
This work was supported by Universiti Tun Hussein Onn, Malaysia [grant number Vot, U607]
A new Bayesian estimation procedure for extended cox model with time varying covariate was presented. The prior was determined using bootstrapping technique within the framework of parametric empirical Bayes. The efficiency of the proposed method was observed using Monte Carlo simulation of extended Cox model with time varying covariates under varying scenarios. Validity of the proposed method was also ascertained using real life data set of Stanford heart transplant. Comparison of the proposed method with its competitor established appreciable supremacy of the method.

Keywords: Bayesian, extended Cox model, time varying covariate, bootstrap, prior

Introduction

The Cox model is a successful method for handling time to event data. Its robustness has been extended to the case of time varying covariate (Therneau & Grambsch, 2000). Time varying covariate also referred to as extended hazard model or extended Cox model allows one or more of the explanatory variables to change over time. The standard Cox proportional hazard model is given in (1):

\[ h(t \mid x, \beta) = h_0(t) \exp[x'\beta], \]

where \( h(t \mid x, \beta) \) is the hazard at time \( t \) given covariate \( x \) and parameter \( \beta \) and \( h_0(t) \) is the baseline hazard. The important feature of (1) is that the hazard ratio
$$\frac{h(t \mid X = x, \beta)}{h(t \mid X = 0, \beta)} = \exp[x'\beta]$$

is independent of time. This stringent assumption is often violated in most real-life scenarios as in the case of heart transplant data. A patient decision towards heart transplant strongly relies on duration of follow up. The changes in decision over time can be recorded as a covariate and model simultaneously with other fixed covariate. This type of model is broadly referred to as extended hazard model (Tseng et al., 2014). Formally, the extended hazard model is defined as

$$h(t \mid x, z(t), \beta) = h_0(t)\exp[x'\beta + \gamma z(t)],$$

where $z(t)$ is the time varying covariate and $\gamma$ is its parameter. The default Cox model assumes hazards are proportional over time and uses partial likelihood instead of maximum likelihood to estimate (1) (Hosmer & Lemeshow, 1999). Estimating (3) using partial likelihood without any data adjustment often results to misspecification error. Misspecification error in estimation give rise to several problems one of which is loss of efficiency (Therneau et al., 2017).

Tseng et al. (2014) develop a semiparametric approach for analyzing (3) using counting process and martingale. They identified the hazard of model misspecification by using (1) instead (3). The result from their Monte Carlo simulation and real-life data analysis showed better performance compared with fixed covariate Cox model (1) and Accelerated Failure Time model (AFT). Apart from Tseng et al., Therneau and Lumley (2016) and Therneau et al. (2017) also used counting techniques a data adjustment procedure to estimate (3). Tseng & Shu (2011) proposed an estimator for (3) using applied kernel smoothing technique. Suissa (2007), Beyersmann et al. (2008), Austin et al. (2006) and Austin (2012) claimed that accounting for time-varying covariates effect is important because it allows one to avoid immortal-time bias.

As a further measure for estimating (1), Bayesian based methods have also been developed. Ibrahim et al. (2001) proposed one of the foremost Bayesian analysis of Cox proportional hazard model using Gamma prior on baseline hazard $h_0(t)$ and Gaussian prior on $\beta$. Their approach can also be extended for estimating (3) but it strongly relies on the piecewise constant hazard assumption. Although, it provides simple framework procedure for solving (1) or (3) but its sensitive to pre-specification of the number and location of hazard function discontinuities along the time axis (Lesaffre & Lawson, 2003; Murray et al., 2016). Other Bayesian
method for analyzing (1) or (3) includes Fahrmeir and Lang (2001), Fahrmeir and Hennerfeind (2003), and Dunson (2005).

Bayesian approaches provide solutions to some classical methods priors if genuine priors are used (Yahya et al., 2014). Bayesian analysis with survival data are somewhat difficult to perform because of prior specification. Tractable Bayesian analysis requires prior that relies so much on unrealistic assumption such as the piece-wise exponential hazard assumption. This drawback calls for flexible prior determination technique. The bootstrap prior technique (Olaniran et al., 2016) and (Olaniran & Yahya, 2017) enjoys this important feature. Olaniran and Yahya (2017) demonstrated the flexibility of bootstrap prior for estimation and hypothesis testing with Gaussian model. The results from Monte Carlo simulation conducted revealed that bootstrap priors guarantee unbiased and efficient estimates. Therefore, this paper focuses on estimating (3) with right censoring assumption using bootstrap prior technique.

**Likelihood of Extended Cox Model**

Suppose there are $i$ individual in a study and we have their time to event, event indicator and associated event covariate: $(X_i, s_i, \{Z_i(t), t \in [0, X_i]\})$. The hazard function is already given in (3) but the associated survival function is

$$S(t \mid x, z(t), \beta) = \exp\left[-\int_0^t h_0(t)\exp[x'\beta + \gamma z(t)]\right].$$  \hspace{1cm} (4)

Now assuming $z(t)$ to be a piece-wise function, thus $z(t)$ can be defined as

$$z(t) = \begin{cases} 
0, & t < t_c \\
1, & t \geq t_c 
\end{cases}$$

where $t_c$ is the time at which the covariate $z(t)$ changes. This implies that $S(t \mid x, z(t), \beta)$ is
Following Collett (2015), the likelihood of a survival model with right censored times can be defined as

\[
L(\theta) = \prod_i f(t_i | \theta) \prod_0 S(t_i | \theta),
\]

where 1 denotes uncensored and 0 denotes right censored. The likelihood in (6) can be simplified if a censoring indicator \( s_i \) that takes 0 for censored and 1 for uncensored is assumed. Thus,

\[
L(\theta) = \prod_i h(t_i | \theta)^{s_i} S(t_i | \theta).
\]

Substituting (3) and (5) in (7),

\[
L(\beta, \gamma)
= \prod_{i=1}^n h_0(t_i) \exp[\mathbf{x}'\beta + \gamma z(t)]^{s_i} \exp\left[ - \int_0^{t_i} h_0(u) \exp[\mathbf{x}'\beta + \gamma z(u)] \, du \right] + \int_{t_i}^{t'} h_0(u) \exp[\mathbf{x}'\beta + \gamma z(u)] \, du
\]

\[
L(\beta, \gamma) = \prod_{i=1}^n h_0(t_i) \exp[\mathbf{x}'\beta + \gamma z(t)]^{s_i} \exp\left[ - \left( \exp[\mathbf{x}'\beta] \int_0^{t_i} h_0(u) \, du \right) \right]
\]

\[
+ \exp[\mathbf{x}'\beta + \gamma] \int_{t_i}^{t'} h_0(u) \, du
\]

The corresponding log-likelihood is
Bayesian Analysis of Extended Cox Model

After specifying the likelihood of the model, the next step in Bayesian analysis is to specify the prior distribution and its hyperparameters. The prior determination stage involves bootstrapping the original data a number of times, say \( B \), then estimate the prior parameters using the counting techniques procedure described in Therneau et al. (2017) in R package “survival”. Formally, the prior hyperparameter \( \hat{\theta}_pr \{ \hat{\beta}_{pr}, \hat{\gamma}_{pr} \} \) is based on partial likelihood:

\[
L(\beta_{pr}, \gamma_{pr}) = \prod_{j=1}^{d} P(i_j \mid H_j) \\
L(\beta_{pr}, \gamma_{pr}) = \prod_{j/k} \exp \left[ \sum_{i} \exp \left[ \sum_{i} \left( x_i^* \beta_{pr} + \gamma_{pr} z_i^* (t) \right) \right] \right]
\]

\[
\hat{\theta}_pr \{ \hat{\beta}_{pr}, \hat{\gamma}_{pr} \} = \frac{1}{B} \sum_{k=1}^{B} \sup \left[ L(\beta_{pr}, \gamma_{pr}) \right]
\]

\[
\text{Var} \left[ \hat{\theta}_pr \{ \hat{\beta}_{pr}, \hat{\gamma}_{pr} \} \right] = \frac{1}{B^2} \left[ \sum_{i=1}^{k} \text{var} (\hat{\beta}_{pr}) \text{cov} (\hat{\beta}_{pr}, \hat{\gamma}_{pr}) \right] \left[ \sum_{i=1}^{k} \text{cov} (\hat{\beta}_{pr}, \hat{\gamma}_{pr}) \text{var} (\hat{\gamma}_{pr}) \right]
\]

Efron and Tibshirani (1993) proved the limiting form of bootstrap estimates is Normal with location and scale parameters as described above. Therefore, the bootstrap prior distribution for Bayesian extended cox model is

\[
f(\beta_{pr}, \gamma_{pr}) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \left( \left\{ \beta_{pr}, \gamma_{pr} \right\} - \mu \right)^{\top} \Sigma^{-1} \left( \left\{ \beta_{pr}, \gamma_{pr} \right\} - \mu \right) \right\},
\]
where \( \mu = \hat{\theta}_p \left\{ \hat{\beta}_p, \hat{\gamma}_p \right\} \) and \( \Sigma = \text{Var}\left[ \hat{\theta}_p \left\{ \hat{\beta}_p, \hat{\gamma}_p \right\} \right] \) are prior hyperparameters and \( x_i^* \) and \( z_i^* \) are bootstrapped covariates.

Once the prior hyperparameters and likelihood are determined, the posterior distribution can be sampled via MCMC (Gelman, 2006). Metropolis Hasting (MH) algorithm (Gelman et al., 2014) was used to sample from the posterior distribution in this paper.

**Simulation Studies**

The simulation strategy used in Austin (2012) was adapted with the following parameters: \( \lambda_0(t) = 0.02 \), \( \beta = 2 \) and \( \gamma = 1 \). The covariate \( x_1 \sim \text{Normal}(0, 1) \), \( t_{ci} \sim \text{Exponential}(\lambda(t) = 0.01) \). The baseline hazard assumed is exponential. Also, sample sizes \( n = 50, 100, 200 \) were used to study the effect of sample size on parameter estimates. Performance metrics used to assess the estimating methods are standard error (SE), bias, and Mean square error (MSE). The formulas based on 1000 repetition of the study are

\[
\text{SE}\left( \hat{\theta} \right) = \sqrt{\frac{\sum_{j=1}^{1000} \left( \hat{\theta}_j - \bar{\theta} \right)^2}{999}} \\
\text{bias}\left( \hat{\theta} \right) = \frac{\sum_{j=1}^{1000} \left( \hat{\theta}_j - \theta \right)}{1000} \\
\text{MSE}\left( \hat{\theta} \right) = \frac{\sum_{j=1}^{1000} \left( \hat{\theta}_j - \theta \right)^2}{1000}
\]

The estimates of the proposed method obtained were compared with estimates using Cox proportional hazard implemented in R via "coxph." The results based on performance metrics are presented in the Tables 1 through 3.

Compiled in Table 1 are the results for small sample size 50. For all the performance metrics used, the results of the proposed method are better than the competing. Specifically, the proposed method is more stable in terms of low average standard error, consistent in terms of low average bias as well as efficient in terms of low mean square error. The performance of time dependent Cox model also enjoys stable estimate but not as Bayesian extended Cox model. It was also
observed that using fixed covariate Cox model largely resulted in unstable and inconsistent estimates. Sample size consistency was observed using size 100 and 200 with results shown in Tables 2 and 3.

Increasing the sample size in Tables 2 and 3 to 100 and 200, improves the performance of all the entire methods. The Bayesian extended Cox model is relatively closer to the true parameter at any reasonable sample size.

Table 1. Simulation results for average estimate \( \hat{\theta} \), average bias (bias(\( \hat{\theta} \))), average standard error (SE(\( \hat{\theta} \))) and average mean square error (MSE(\( \hat{\theta} \))) based on 1000 replications for sample size \( n = 50 \); the true parameter is \( \theta = \{ \beta = 2, \gamma = -1 \} \)

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Parameter</th>
<th>Fixed covariate Coxph</th>
<th>Time dependent covariate Coxph</th>
<th>Bayes extended Cox</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\theta} )</td>
<td>( \beta )</td>
<td>1.7962</td>
<td>2.0935</td>
<td>2.0626</td>
</tr>
<tr>
<td></td>
<td>( \gamma )</td>
<td>-4.1858</td>
<td>-2.0613</td>
<td>-1.4403</td>
</tr>
<tr>
<td>SE(( \hat{\theta} ))</td>
<td>( \beta )</td>
<td>0.4297</td>
<td>0.3967</td>
<td>0.3734</td>
</tr>
<tr>
<td></td>
<td>( \gamma )</td>
<td>5.7721</td>
<td>4.3269</td>
<td>3.4515</td>
</tr>
<tr>
<td></td>
<td>( \beta, \gamma )</td>
<td>3.1009</td>
<td>2.3618</td>
<td>1.9125</td>
</tr>
<tr>
<td>bias(( \hat{\theta} ))</td>
<td>( \beta )</td>
<td>-0.2038</td>
<td>0.0935</td>
<td>0.0626</td>
</tr>
<tr>
<td></td>
<td>( \gamma )</td>
<td>-3.1858</td>
<td>-1.0613</td>
<td>-0.4403</td>
</tr>
<tr>
<td></td>
<td>( \beta, \gamma )</td>
<td>-0.7843</td>
<td>-0.4839</td>
<td>-0.1889</td>
</tr>
<tr>
<td>MSE(( \hat{\theta} ))</td>
<td>( \beta )</td>
<td>0.2262</td>
<td>0.1660</td>
<td>0.1433</td>
</tr>
<tr>
<td></td>
<td>( \gamma )</td>
<td>43.4663</td>
<td>19.8486</td>
<td>12.1069</td>
</tr>
<tr>
<td></td>
<td>( \beta, \gamma )</td>
<td>21.8463</td>
<td>10.0073</td>
<td>6.1251</td>
</tr>
</tbody>
</table>

Table 2. Simulation results for average estimate \( \hat{\theta} \), average bias (bias(\( \hat{\theta} \))), average standard error (SE(\( \hat{\theta} \))) and average mean square error (MSE(\( \hat{\theta} \))) based on 1000 replications for sample size \( n = 100 \); the true parameter is \( \theta = \{ \beta = 2, \gamma = -1 \} \)

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Parameter</th>
<th>Fixed covariate Coxph</th>
<th>Time dependent covariate Coxph</th>
<th>Bayes extended Cox</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\theta} )</td>
<td>( \beta )</td>
<td>1.7633</td>
<td>2.0536</td>
<td>2.0417</td>
</tr>
<tr>
<td></td>
<td>( \gamma )</td>
<td>-2.3318</td>
<td>-1.1067</td>
<td>-0.8575</td>
</tr>
<tr>
<td>SE(( \hat{\theta} ))</td>
<td>( \beta )</td>
<td>0.0822</td>
<td>0.0689</td>
<td>0.0686</td>
</tr>
<tr>
<td></td>
<td>( \gamma )</td>
<td>2.3986</td>
<td>1.5671</td>
<td>0.7976</td>
</tr>
<tr>
<td></td>
<td>( \beta, \gamma )</td>
<td>1.2404</td>
<td>0.8180</td>
<td>0.4331</td>
</tr>
</tbody>
</table>

Table 2 (continued).
Table 3. Simulation results for average estimate $\hat{\theta}$, average bias ($\text{bias}(\hat{\theta})$), average standard error ($\text{SE}(\hat{\theta})$) and average mean square error ($\text{MSE}(\hat{\theta})$) based on 1000 replications for sample size $n = 200$; the true parameter is $\theta = \{\beta = 2, \gamma = -1\}$

<table>
<thead>
<tr>
<th>Metrics $\hat{\theta}$</th>
<th>Parameter $\beta$</th>
<th>Fixed covariate $\text{coxph}$</th>
<th>Time dependent covariate $\text{coxph}$</th>
<th>Bayes extended Cox</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\theta}$</td>
<td>$\beta$</td>
<td>1.7333</td>
<td>2.0173</td>
<td>2.0060</td>
</tr>
<tr>
<td></td>
<td>$\gamma$</td>
<td>-2.1450</td>
<td>-1.0261</td>
<td>-0.8305</td>
</tr>
<tr>
<td>SE($\hat{\theta}$)</td>
<td>$\beta$</td>
<td>0.1827</td>
<td>0.1686</td>
<td>0.1780</td>
</tr>
<tr>
<td></td>
<td>$\gamma$</td>
<td>0.3635</td>
<td>0.3810</td>
<td>0.2452</td>
</tr>
<tr>
<td></td>
<td>$\beta, \gamma$</td>
<td>0.2731</td>
<td>0.2748</td>
<td>0.2116</td>
</tr>
<tr>
<td>bias($\hat{\theta}$)</td>
<td>$\beta$</td>
<td>-0.2667</td>
<td>0.0173</td>
<td>0.0060</td>
</tr>
<tr>
<td></td>
<td>$\gamma$</td>
<td>-1.1450</td>
<td>-0.0261</td>
<td>0.1695</td>
</tr>
<tr>
<td></td>
<td>$\beta, \gamma$</td>
<td>-0.7059</td>
<td>-0.0044</td>
<td>0.0878</td>
</tr>
<tr>
<td>MSE($\hat{\theta}$)</td>
<td>$\beta$</td>
<td>0.1045</td>
<td>0.0287</td>
<td>0.0317</td>
</tr>
<tr>
<td></td>
<td>$\gamma$</td>
<td>1.4432</td>
<td>0.1459</td>
<td>0.0888</td>
</tr>
<tr>
<td></td>
<td>$\beta, \gamma$</td>
<td>0.7739</td>
<td>0.0873</td>
<td>0.0603</td>
</tr>
</tbody>
</table>

Application to Stanford Heart Transplant Data

The Stanford heart transplant data were used to explain the time-varying covariate effect of patient decision towards heart transplant (Tseng et al., 2014). The dataset contains survival times of 103 patients accepted into the heart transplant program along with ages at the time of the first transplant and T5 mismatch scores. Tseng et al. (2014) analyzed the dataset using Accelerated Failure Time (AFT), Extended hazard (EH) and time dependent Cox for 99 patients. They excluded 4 patients’
outcomes because of incomplete observations. However, direct estimation of the proposed method could not be achieved in the presence of missing observation.

First, impute the missing observations using multiple imputation techniques via the Expectation Maximization (EM) algorithm. The three-time dependent covariate \( z_1(t) \), \( z_2(t) \), and \( z_3(t) \) used by Tseng et al. (2014), Lin and Ying (1995) among others are defined as

\[
z_1(t) = \begin{cases} 
0, & t < t_c \\
1, & t \geq t_c 
\end{cases}
\]

\[
z_2(t) = \begin{cases} 
0, & t < t_c \\
age at transplant - 35, & t \geq t_c 
\end{cases}
\]

\[
z_3(t) = \begin{cases} 
0, & t < t_c \\
mismatch score - 0.5, & t \geq t_c 
\end{cases}
\]

where \( z_1(t) \) is the transplant status, \( z_2(t) \) is the patient’s age at the transplant, \( z_3(t) \) is the mismatch score, and \( t_c \) is the waiting time for accepting a heart transplant.

The result in Table 4 reiterate efficiency property of the proposed Bayesian method. In all the three covariates, the standard error of Bayesian extended Cox model are clearly lower than time dependent Cox model. In addition, the 95% credible region for the parameter estimates are reasonably narrower than the Cox model.

Table 4. Parameter estimation of the Stanford heart transplant data for time dependent Cox model and Bayesian extended Cox model

<table>
<thead>
<tr>
<th>Model</th>
<th>Covariate</th>
<th>Coef.</th>
<th>se(Coef.)</th>
<th>95% lower</th>
<th>95% upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time dependent Cox</td>
<td>( z_1(t) )</td>
<td>-4.1089</td>
<td>0.5520</td>
<td>-5.1908</td>
<td>-3.0270</td>
</tr>
<tr>
<td></td>
<td>( z_2(t) )</td>
<td>0.0244</td>
<td>0.0168</td>
<td>-0.0085</td>
<td>0.0572</td>
</tr>
<tr>
<td></td>
<td>( z_3(t) )</td>
<td>0.5296</td>
<td>0.2525</td>
<td>0.0347</td>
<td>1.0246</td>
</tr>
<tr>
<td>Bayesian extended Cox</td>
<td>( z_1(t) )</td>
<td>-1.8103</td>
<td>0.2043</td>
<td>-2.2107</td>
<td>-1.4098</td>
</tr>
<tr>
<td></td>
<td>( z_2(t) )</td>
<td>0.0392</td>
<td>0.0135</td>
<td>0.0128</td>
<td>0.0656</td>
</tr>
<tr>
<td></td>
<td>( z_3(t) )</td>
<td>0.7333</td>
<td>0.1621</td>
<td>0.4157</td>
<td>1.0510</td>
</tr>
</tbody>
</table>
Conclusion

An alternative modelling strategy was presented for extended Cox model with time-varying covariate within the Bayesian framework. The analysis is simple in terms of prior specification that is based on basic bootstrapping principle. Furthermore, the method stands as a hidden extension of Cox model as the upper bound of its standard error is that of standard Cox model. This is because the standard error of Bayesian extended Cox is the average of data variance and bootstrap variance. In general, the proposed method is clearly more efficient than the traditional Cox model at any reasonable sample size which is an important issue in medicine.

Acknowledgements

This work was supported by Universiti Tun Hussein Onn, Malaysia [grant number Vot, U607].

References


Fahrmeir, L., & Hennerfeind, A. (2003). *Nonparametric Bayesian hazard rate models based on penalized splines* (Sonderforschungsbereich 386, Paper
BAYESIAN ANALYSIS OF EXTENDED COX MODEL

361. Ludwig Maximilian University of Munich: Munich, Germany. doi: 10.5282/ubm/epub.1736


12


