

Piecewise Linear Multiple Change Point Model: An Application to Survival Analysis

Sukanta Chakraborty¹ and Soma Chowdhury Biswas²

¹Assistant Director (General)
 Bangladesh Bank
 Email: sukantacu1992@gmail.com

²Professor
 Department Of Statistics, University Of Chittagong
 Email: soma.stat@cu.ac.bd

ABSTRACT

The hazard function is a frequently used function for modeling and evaluating the time related events. In the Cox-proportional hazard model this hazard function assumes to be constant. Earlier many statisticians investigated this assumption and found that this function is not constant over time and it has some change point in survival time of patient. Many authors Matthews and Farewell (1982), Pons(2002), Gijbels and Gurler (2003) and Zhao et al.(2009) investigated about one change point in hazard function through piecewise constant hazard model. Moreover, Goodman et al. (2006, 2007), Qian and Zhang (2014) investigated about multiple change point in hazard function through piecewise linear multiple change point model by Nelder-mead simplex method. Rois et al. (2017) investigated about multiple change point in hazard function through piecewise linear multiple change point model by Cross entropy algorithm. In our paper we calculate the likelihood function for piecewise linear multiple change point model considering three change point. Since this likelihood function is non-linear so we use non-linear minimization method to estimate the parameters of the model. Later we make comparison among the parameters estimated by Non-linear method with those estimated by Nelder-mead simplex method & Cross entropy algorithm in a simulation study. That study shows that the standard error of the estimates estimated by Non-linear minimization method is lowest among the three method. A Wald type test shows that the estimated parameters are significant. So, change point estimation in hazard function by non-linear minimization method can be considered as an improvement in change point analysis.

Keywords: Hazard function, Change point, Non-linear minimization method.

Mathematics Subject Classification: 62N01, 62N02, 62N03

Journal of Economic Literature (JEL) Classification: C34, C41, C61

1 INTRODUCTION

Let $x_1, x_2, \dots, x_n$ be the identically and independently distributed survival times ; $t_1, t_2, \dots, t_n$ be the observation times ; z' be the vector of covariates and β be the regression coefficients then Cox proportional hazard model is defined as

$$h(t|x) = h_0(t)e^{z'\beta}$$

Cox Semi-parametric proportional hazards modeling lies upon two basic assumptions:

- a) the hazard rate to be constant over time across groups and
 - b) the hazard functions of two compared groups (e.g. treatment groups) must run in parallel over time.
- If the latter condition holds, then covariate-adjusted hazard ratios can be produced. Its semi-parametric nature is due to the constant base line hazard which remains unspecified. Therefore, a partial likelihood maximization technique is used for estimation of model parameters.

Primary interest lies in estimating the parameter β using the partial likelihood:

$$L(\beta) = \frac{\prod_1^n \exp(z_{(i)}' \beta)}{\sum_{j \in R(t_i)} \exp(z_j' \beta)}$$

where $R(t_i)$ is the risk set at time t_i . The MLE $\hat{\beta}$ (a vector) is asymptotically distributed $N \sim (\beta, I^{-1})$ where I represents the Fisher information.

Here the first assumption i.e. the hazard rate to be constant over time across groups is not realistic.

To overcome this limitation we have to use "PIECEWISE EXPONENTIAL MODEL".

There are two types of piecewise exponential model

Piecewise linear multiple change point model

Piecewise Constant Multiple Change Point Model

In the field of change point estimation there is enormous erudition about piecewise constant exponential model with one change point. Matthews and Farewell (1982), Nguyen et al. (1984), Matthews et al. (1985), Yao (1986), Worsley (1988), Henderson (1990), Loader (1991), Pons (2002), Gijbels and Gu'rler (2003), and Zhao et al. (2009) investigated extensively the estimation and testing in a piecewise constant model with one change point in parametric and semi-parametric approaches using likelihood-ratio type tests, score test and Bayesian test. Although Matthews and Farewell(1992) did not consider censored data in their analysis. They claimed that dropping those data did not affect significantly the outcome of the likelihood ratio test. Here- after, most of the subsequent work develops theory either by discarding censored data and only considering the observable survival time variable, or by modifying the likelihood function for censored data.

But when we move our angle of vision towards piecewise linear multiple change point model we find meager literature. So far our knowledge there are published works by Goodman et al. (2006, 2011), Qian and Zhang (2014), Rois et al.(2017) in multiple change points hazard problems.

Qian and Zhang (2014) proposed an algorithm to fit both susceptibles and long-term survivors with observed covariates through a grid search weighted least squared method assuming that all potential change points lie in a certain known interval $[B_1, B_2]$ for detecting the number of change point in the hazard functions and estimating those. This is a simulation study without any theoretical contents, moreover, assuming known interval for change points makes the application even more impractical.

Goodman et al.(2006,2011) proposed a methodology for estimation of multiple change points using the Nelder-Mead Simplex algorithm and a model selection approach using sequential testing with likelihood ratio test and Wald type test statistics in the piecewise constant hazard model and piecewise linear hazard model for random right censored data. The latter model will easily accommodate the addition of covariates. The estimation of multiple change points using the Nelder-Mead Simplex algorithm is capable of handling an unlimited number of covariates but restricted to an additive nature only. This method also suffers from the number of censored observations. When censoring occurs near the change points valid estimation of these points is not possible.

Rois et al. (2017) in their thesis work estimated two change point in hazard function under random censorship by "Cross-Entropy algorithm". They also showed that change point estimated by their proposed algorithm is better than those estimated by the Nelder-Mead Simplex algorithm. But it has a disadvantage that it doesn't gives us variance and covariance matrix at the meantime.

In our paper we ambush to estimate three change points for piecewise linear multiple change point model by non-linear minimization method. Latter we aim to make a comparison of the estimator that is estimated by the Nelder-Mead Simplex algorithm , Cross-Entropy algorithm, non-linear minimization method.

The residuum of the article is merged as follows: We define piecewise linear multiple change point model with three change points and we also develop the likelihood function for that model in section 2.We describe our data generation procedure in section 3.In section 4 we estimate the parameters of the model by three approximation method of MLE and make a comparison among them. Later we test the significance of the estimated parameters by Wald type statistics in section 5. And we conduct our power analysis in section 6. In section 7 we make a general conclusion.

2. PIECEWISE LINEAR MULTIPLE CHANGE POINT MODEL WITH THREE CHANGE PONITS AND IT'S LIKELIHOOD FUNCTION

Let $X_1, X_2, \dots, X_n$ be the identically and independently distributed survival times ; $t_1, t_2, \dots, t_n$ be the observations times ; z' be the vector of covariates ; β be the regression coefficients then

the piecewise linear exponential model is given as follows

$$h(t/x)=h_0(t) e^{z'\beta} \dots\dots\dots(1)$$

In Cox proportional model, the base line hazard $h_0(t)$ assumes to be fixed . But actually the hazard function is an important component of survival analysis since it describes the immediate risk of failure at a given time point. Although common survival methods such as the Cox proportional hazards model do not require explicit estimation of the hazard function, as they are more concerned with the effects of the covariates on the hazard function, there are several situations where explicit estimation of the hazard function is useful. One such case is change-point hazard rate models. These models assume a function with different hazard rates that change at a few time points. These times points are often referred to as the change points, and are unknown and need to be estimated. For one change point we can write equation (1.1) as follows:

$$\# \ln h(t/x) = \ln h_0(t) + z'\beta$$

$$\# \ln h(t/x) = \alpha_0 + \alpha_1 t + z'\beta$$

$$\# \ln h(t/x) = \alpha_0 + \alpha_1 t + \alpha_2(t - \tau_1) + z'\beta$$

$$\# \ln h(t/x) = \alpha_0 + \alpha_1 t + \alpha_2(t - \tau_1) + \alpha_3(t - \tau_2) + \alpha_4(t - \tau_3) + z'\beta$$

$$\therefore \eta(t) = \alpha_0 + \alpha_1 t + \alpha_2(t - \tau_1) + \alpha_3(t - \tau_2) + \alpha_4(t - \tau_3) + z'\beta \dots \dots \dots (2)$$

The model (2) is called Piecewise exponential model with three change point. It is called piecewise since it describes the hazard rate Here α_0 is initial log hazard rate. α_1 is log hazard rate for time t and α_2 is the log hazard rates for first change point, α_3 is the log hazard rates for first change point. τ_1 is the first change point. τ_2 is the second change point and τ_3 is the third change point.

Again let $c_1, c_2, c_3, \dots, c_n$ be the censoring times and $\delta_1, \delta_2, \dots, \delta_n$ be the censoring indicators then we observe (T_i, δ_i) where $T_i = \min(x_i, c_i)$ is the survival times.

And $\delta_i = 1$ if $x_i \leq c_i$

$= 0$ otherwise

We know the likelihood function is as follows

$$L = \prod_0^n f(T_i) \prod_0^n s(T_i)$$

deaths censored

$$= \prod_0^n h(T_i) s(T_i) \prod_0^n s(T_i)$$

Since

$$f(T_i) = h(T_i) s(T_i)$$

$$= \prod_0^n [h(T_i)]^{\delta_i} s(T_i)$$

Now taking log on the both sides we have

$$\log L = \sum_0^n \{\delta_i h(T_i) + \log s(T_i)\} \dots \dots \dots (3)$$

As we know

$$s(T_i) = e^{(-\int_0^t h(u) du)}$$

$$\log s(T_i) = -\int_0^t h(u) du$$

Then from equation no (2) we have

$$\log L = \sum_0^n \{ \delta_i h(T_i) - \int_0^{T_i} h(u) du \} \dots\dots\dots (4)$$

Again we know

$$\log h(t) = \eta(t)$$

$$h(t) = \exp(\eta(t))$$

$$\text{So, } h(u) = \exp(\eta(u))$$

So from equation no (3) we have

$$\log L = \sum_0^n \{ \delta_i h(T_i) - \int_0^{T_i} e^{\eta(u)} du \}$$

So finally we have

$$\log L = \sum_{i=1}^n [\delta_i \{ \alpha_0 + \alpha_1 T_i + \alpha_2 (T_i - \tau_1) + \alpha_3 (T_i - \tau_2) + \alpha_4 (T_i - \tau_3) + z' \beta \} - \int_0^{T_i} e^{\{ \alpha_0 + \alpha_1 u + \alpha_2 (u - \tau_1) + \alpha_3 (u - \tau_2) + \alpha_4 (u - \tau_3) + z' \beta \}} du] \dots\dots\dots (5)$$

3. DATA GENERATION PROCEDURE

Survival times T_i from the model (2) can be simulated by the inversion of the cumulative hazard, which is called the inverse transformation method (or so- called inverse cumulative distribution function (CDF) method. Here, we used the inverse CDF method. Since

$$F(t) = 1 - \exp(-\Lambda(t))$$

$$U = 1 - \exp(-\Lambda(T)), \text{ where } U \sim U(0,1) \text{ and}$$

$$\text{Hence, } T = \Lambda^{-1} (-\log(1-U)).$$

As, $-\log(1-U) \sim \text{Exp}(1)$, we can apply the inverse cumulative hazard to an exponential random variable. Hence, survival times T_i are generated for the piecewise linear model (2) using an exponential random variable with rate 1 and then plug it into $\Lambda^{-1}(u)$. Censoring times C_i are generated by using the exponential distribution with rate b . Generated censoring times censoring proportion can be specified by redefining the rate of the generated exponential variate, e.g., using $C_i \sim \text{Exp}(0.45)$ we generated 15% censored observations for the piecewise linear model (2). Then, we calculated observed data $X_i = \min(T_i, C_i)$ and also identified corresponding non-censoring indicator δ_i . We generated the dichotomous covariate by using the binomial random variable with parameters $n = 1$ and $p = 0.7$. The continuous covariate was generated by using the uniform distribution on $[0, 1]$.

4. ESTIMATIONS OF PARAMETER

4.1. Cross-Entropy Algorithm

Cross-Entropy is commonly used to quantify the difference between two probability distributions. Cross-Entropy (CE) method, which is developed by Rubinstein (1997) and Rubinstein and Kroese (2004, 2008) for estimating the multiple change points in the piecewise constant hazard model as well as piecewise linear hazard model to handle the additive covariates with any number of censored observations. More specifically, this is an extension of Goodman et al.(2006,2011) work.

Algorithm 1: CE Algorithm for estimating the change point in the hazard functions

1: Initialize parameter vectors $v_0 = (\alpha_0, \tau_0, \beta_0)$. Set sample size M , rarity parameter ρ , smoothing parameter c and $t = 1$ for model (2).

2: Generate the observed survival times $X_1, \dots, X_M$ where $X_M = X_{m1}, \dots, X_{mn}$ for $m = 1, \dots, M$, from the the density $f(\cdot; v_{t-1})$, which can be done by generating the survival times $T_1, \dots, T_M$ from the density and the censoring times $C_1, \dots, C_M$ and $\text{Exp}(b)$ for some b , respectively, for the models . Hence, we get $X_{m,i} = \min(T_{m,i}, C_{m,i})$ and $\delta_{m,i} = I\{T_{m,i} \leq C_{m,i}\}$ for $m = 1, \dots, M$ and $i = 1, \dots, n$.

3: Calculate the negative log-likelihood functions respectively for the models as $S(X_i) = -\log L$ for all i , and order them from the smallest to the largest, i.e., $S_{(1)} \leq \dots \leq S_{(M)}$. Let y_t be the sample ρ -quantile of performances, i.e., the number of elite samples $y_t = S_{(\lceil \rho M \rceil)}$.

4: Use the same sample $X_1, \dots, X_M$ and solve the stochastic program

$$\hat{v}_t = \max_v \sum_{k=1}^M I\{S(X_k) \leq y_t\} \log f(X_k, v).$$

5: Smooth, $\tilde{v}_t = c\hat{v}_t + (1 - c)\tilde{v}_{t-1}$, where c is a smoothing parameter.

6: If for some $t \geq k$, say $k=5$, $y_t = y_{t-1} = \dots = y_{t-k}$ then stop. Otherwise set $t = t + 1$ and go to Step 2.

To run algorithm 1 for the piecewise linear model , at first we need to initialize input parameter's value. Hence, we initialized the input vector v_0 with $\alpha_0 = (0.5, 0.5, 0.5, 0.5, 0.5)$, $\tau_0 = (0.5, 1.0, 0.5)$, and $\beta_0 = (0.5, 0.5)$, and used the rarity parameter $\rho = 0.20$, the smoothing parameter $c = 0.98$ and the sample size $M = 200$ for conducting simulation with the CE method. We performed CEoptim function from the **CEoptim** package in **R** to estimate the change points in the hazard function based on our proposed Algorithm1, where we specified the arguments $\text{mean} = c(0.5, 1, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5)$, $\text{sd} = c(.5, 1, .5, .5, .5, .5, .5, .5, .5, .5)$, and a linear constraints on the change points using the **conMat** and **conVec** arguments as **conMat** = **rbind**(**diag**(10), **-diag**(10)) and **conVec** = $c(.8, 1.8, .2, .3, .5, .8, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0)$. Since it can not give us the variance – covariance matrix for the parameter, so to estimate the variance covariance matrix we have to take help of another R package. We can approximate these variances using the R software package **numDeriv** with the function **hessian**.

4.2 Nelder-Mead Simplex Method

Estimates of the parameters in the model are calculated by minimizing the negative log-likelihood function using the optimization function Nelder-Mead Simplex algorithm evaluated at the maximum likelihood estimates of the α_i and finding those values of the τ_j that minimize the function. The maximum likelihood estimates of the τ_j are those values returned by the optimization function that minimize the negative log-likelihood. To implement the Nelder-Mead Simplex algorithm in this context, we used **optim** function in **R**, where we considered the starting value of the parameters in the **par** argument as $(\alpha_0, \tau_0, \beta_0) = (0.1985673, 1.708206, 0.1986111, 0.14416, 0.474127, 0.4802961, 0.6732984, 0.9457613, 0.5335454, 0.9393383)$.

4.3 Non-Linear Minimization Method

To optimizing a nonlinear function, an iterative algorithm starts from some initial value of the argument in that function and then repeatedly calculates next available value according to a particular rule until an optimum is reached approximately. To minimize $Q_T(\beta)$, a generic algorithm can be expressed as

$$\beta(i+1) = \beta(i) + s(i)d(i),$$

so that the $(i+1)$ th iterated value $\beta^{(i+1)}$ is obtained from $\beta^{(i)}$, the value from the previous iteration, by adjusting the amount $s^{(i)}d^{(i)}$, where $d^{(i)}$ characterizes the direction of change in the parameter space and $s^{(i)}$ controls the amount of change. Different algorithms are resulted from different choices of s and d . As maximizing Q_T is equivalent to minimizing $-Q_T$.

In our thesis work we use Newton type algorithm to estimate the parameter of the model and also their variance covariance matrix. To implement an iterative algorithm, one must choose a vector of initial values to start the algorithm and a stopping rule to terminate the iteration procedure. Initial values are usually specified by the researcher or by random number generation; prior information, if available, should also be taken into account. To estimate or conducting the estimation procedure we initialize the parameter as follows:

Initial log hazard rate $\alpha_0 = 0.6069$, log hazard rate along time t $\alpha_1 = 1.62967$, log hazard rate until first change point $\alpha_2 = 0.0634$, log hazard rate until second change point $\alpha_3 = 0.0934$, log hazard rate until third change point $\alpha_4 = 0.050$, first change point $\tau_1 = 0.3737$, second change point $\tau_2 = 0.407$, third change point $\tau_3 = 0.407$, first regression coefficient $\beta_1 = 0.95$, second regression coefficient $\beta_2 = 0.66$. In R program We can easily calculate the negative log likelihood with the above stated initial parameter and can estimate the parameter by using **nlm()** function in **R**.

To compare the three methods we represent the obtain result in the following table along with their variances:

Table 1: Comparison of Parameters estimated for Piecewise linear Exponential model with three change point by above cited three methods

For n=200

Parameter	Non-linear minimization method		Nelder-Mead Simplex Method		Cross-Entropy method	
	Estimated value of parameter	Standard Error of parameter	Estimated value of parameter	Standard Error of parameter	Estimated value of parameter	Standard Error of parameter
α_0	0.1987786	1.39×10^{-05}	0.2269260	3.47×10^{-04}	0.4899067	2.70×10^{-03}
α_1	1.7081084	2.18×10^{-05}	1.7233884	3.93×10^{-04}	1.330608	2.51×10^{-03}
α_2	0.1996600	1.31×10^{-05}	0.1904966	5.35×10^{-04}	0.06516295	3.27×10^{-04}
α_3	0.4747754	2.02×10^{-05}	0.4981802	5.36×10^{-04}	0.26686862	7.75×10^{-04}
α_4	0.6731607	1.54×10^{-05}	0.7816467	2.30×10^{-03}	0.7531103	1.25×10^{-03}
τ_1	0.1448411	1.86×10^{-05}	0.1582665	2.62×10^{-04}	0.05119837	1.88×10^{-04}
τ_2	0.4797921	2.96×10^{-05}	0.4397566	1.47×10^{-03}	0.1686548	1.20×10^{-03}
τ_3	0.9466491	5.37×10^{-05}	0.9431595	1.58×10^{-03}	0.6901867	1.19×10^{-03}
β_1	0.5340583	3.27×10^{-05}	0.5559301	8.94×10^{-04}	0.5390791	9.14×10^{-04}
β_2	0.9398677	2.59×10^{-05}	0.9396408	3.66×10^{-03}	0.8120171	4.56×10^{-03}

Table 2: Comparison of Parameters estimated for Piecewise linear Exponential model with three change point by above cited three methods

For n=500

Parameter	Non-linear minimization method		Nelder-Mead Simplex Method		Cross-Entropy method	
	Estimated value of parameter	Standard Error of parameter	Estimated value of parameter	Standard Error of parameter	Estimated value of parameter	Standard Error of parameter
α_0	0.1985048	2.48×10^{-05}	1.2250549	3.14×10^{-04}	0.7264777	1.51×10^{-03}
α_1	1.7081856	2.86×10^{-05}	1.5161140	4.53×10^{-04}	0.9746606	1.50×10^{-03}
α_2	0.1985616	8.62×10^{-05}	0.1089612	2.54×10^{-04}	0.1315243	3.78×10^{-04}
α_3	0.4741045	1.83×10^{-05}	0.7354182	3.20×10^{-04}	0.6872703	3.15×10^{-04}
α_4	0.6732148	3.81×10^{-05}	0.693580	6.34×10^{-04}	0.9145075	8.64×10^{-04}
τ_1	0.1441261	1.09×10^{-05}	2.0734505	1.54×10^{-04}	0.0588428	1.07×10^{-04}
τ_2	0.4802572	1.03×10^{-05}	1.6398214	3.64×10^{-04}	0.3615768	8.01×10^{-04}
τ_3	0.9457755	4.05×10^{-05}	0.9708759	2.87×10^{-04}	0.9590054	1.65×10^{-04}
β_1	0.5335661	1.14×10^{-05}	0.4811854	3.07×10^{-04}	0.1677868	4.14×10^{-04}
β_2	0.9392475	1.54×10^{-05}	0.6655193	3.43×10^{-04}	0.4184561	5.50×10^{-04}

Table 3: Comparison of Parameters estimated for Piecewise linear Exponential model with three change point by above cited three methods

For n=1000

Parameter	Non-linear minimization method		Nelder-Mead Simplex Method		Cross-Entropy method	
	Estimated value of parameter	Standard Error of parameter	Estimated value of parameter	Standard Error of parameter	Estimated value of parameter	Standard Error of parameter
α_0	0.1976923	1.23×10^{-05}	0.49481011	2.39×10^{-04}	0.479907	1.09×10^{-03}
α_1	1.7078165	1.87×10^{-05}	2.11012537	1.97×10^{-04}	1.331608	6.60×10^{-04}
α_2	0.1983563	2.64×10^{-05}	1.41922906	1.81×10^{-04}	0.07016295	3.85×10^{-04}
α_3	0.4732344	2.70×10^{-05}	1.70438129	1.91×10^{-04}	0.2786862	1.21×10^{-04}
α_4	0.6734646	8.07×10^{-05}	2.18275579	2.23×10^{-04}	0.793103	8.30×10^{-04}
τ_1	0.1430526	3.32×10^{-05}	0.43507041	3.58×10^{-04}	0.05919837	8.35×10^{-04}
τ_2	0.4799107	9.01×10^{-05}	0.68612326	1.34×10^{-04}	0.1706548	5.02×10^{-03}
τ_3	0.9453852	1.25×10^{-05}	0.24008532	2.42×10^{-04}	0.7011867	1.11×10^{-03}
β_1	0.5329684	1.15×10^{-05}	0.88910916	1.47×10^{-04}	0.5490791	1.83×10^{-04}
β_2	0.9393167	1.50×10^{-05}	.08117507	2.43×10^{-04}	0.8520171	3.34×10^{-03}

We see from the above table the estimator that are estimated by non linear minimization method has less variance than those estimated by Nelder-Mead simplex method and Cross Entropy method.

That's why Non-linear minimization estimator is more consistent than others.

5. TESTING THE SIGNIFICANCE OF ESTIMATED PARAMETERS

Let $\theta' = [\alpha_0, \alpha_1, \dots, \alpha_{k+1}, \tau_1, \dots, \tau_k, \beta]$, a test of the hypothesis for the j th element of θ is,

$$X_w = \frac{(\hat{\theta}_j - \bar{\theta}_0)^2}{\widehat{\text{var}}(\hat{\theta}_j)} \sim \chi^2_1$$

Where $\widehat{\text{var}}(\hat{\theta}_j) = [I(\hat{\theta})^{-1}]_{jj}$. In the piecewise linear model a Wald test for the null hypothesis of no change point versus the alternative of one change point is equivalent to testing the hypothesis that $\alpha_2 = 0$, as the model with one change point reduces to the model with no change points when α_2 is equal to zero. A test for the existence of change point τ_k can be conducted by testing, $H_0 : \alpha_{k+1} = 0$

Versus

$H_1 : \alpha_{k+1} \neq 0$

Using a Wald type test statistic of the form,

$$X_w = \frac{(\alpha_{k+1})^2}{\widehat{\text{var}}(\alpha_{k+1})} \sim \chi^2_1$$

Parameters	Calculated Value	Tabulated Value at 5% level	Comments
α_2	232579437.6	3.84146	Significant
α_3	55247960.89	3.84146	Significant
α_4	1903971966	3.84146	Significant
β_1	269073837.5	3.84146	Significant
β_2	1320997897	3.84146	Significant

6. POWER ANALYSIS

A very important concept in hypothesis testing is the idea of the power of the test. Basically it is the ability of a test to reject a false null hypothesis. Technically the power is defined as $1-\beta$ where β is the type-2 error probability.

Now we are interested to test the null hypothesis

$$H_0 : C\beta = d \text{ against alternative hypothesis}$$

$H_1 : C\beta = b \neq d$, where C is a $c \times p$ matrix of constants imposing c linearly independent constraints on the elements of β and d is a $p \times 1$ vector of constants.

To construct the power function of the Wald test at first we have to know the distribution of Wald test under alternative hypothesis H_1 . PHILLIPS.P.C.B (1986) discussed the exact distribution of the Wald test under alternative hypothesis. According to PHILLIPS.P.C.B (1986), the probability density function of the Wald statistic(w) under H_1 is given by

$$pdf(w) = \frac{\exp(m'm/2)w^{\frac{q}{2}-1}}{N^{\frac{q}{2}}\Gamma(\frac{q}{2})} \sum_{j,k} \frac{(-1)^j (1/2)^k (w/N)^f}{jk(1/2)_k C_{(f)}(I_q)} \bullet \left[\det(L(\partial X \otimes I)L')^{1/2} \cdot C_f^{j,k}(L(\partial X \otimes I)L', (L(\partial X \otimes I)L'), mm') \det(1-X)^{-N/2} \right]_{X=0}$$

$$= \chi_q^2(\delta^2)$$

With the non centrality parameter is

$$\delta^2 = mm' = b'V^{-1}b = b'(C \sum_{\beta} C')^{-1}b$$

Where $L = V^{-1/2}C^{1/2}V^{1/2}C'^{-1/2}$

According to Shieh, G. (2000), if we want calculate the power function, at first we have to find the $100(1-\alpha)$ th percentile of a central chi-square distribution with p d.f., denoted by $\chi_{c,1-\alpha}^2$ and then find

the non centrality parameter δ^2 of the non-central chi-square distribution with c d.f. such that the $100 \cdot \beta$ th percentile ,denoted by $\chi_{c,\beta}^2(\delta^2)$ is equal to the $\chi_{c,1-\alpha}^2$, mathematically it can be written as

$$\psi(\delta^2, \alpha) = 1 - F_{\chi_{c,\beta}^2(\delta^2)}(p_{1-\alpha})$$

Where $p_{1-\alpha}$ denotes the $(1-\alpha)$ th percentile of the central chi-square $\chi_{c,1-\alpha}^2$ and $F_{\chi_{c,\beta}^2(\delta^2)}(p_{1-\alpha})$ is the cumulative density of the non-central chi-square distribution with c degrees of freedom and non centrality parameter δ^2 at $p_{1-\alpha}$.

Let us consider a piecewise linear exponential model with three change point

$$\ln h(t/x) = \alpha_0 + \alpha_1 t + \alpha_2(t-\tau_1) + \alpha_3(t-\tau_2) + \alpha_4(t-\tau_3) + z'\beta$$

Let $\hat{\alpha}_2$ be the estimator of the piecewise linear exponential model with three change point. Now if we are interested to test the null hypothesis $H_0 : \alpha_2 = 0$ against the alternative hypothesis $H_1 : \alpha_2 \neq 0$. Then the Wald test statistic in the matrix form under the null hypothesis is

$$Q_c = (C\hat{\alpha} - d)[C\hat{V}(\hat{\alpha})C']^{-1}(C\hat{\alpha} - d)$$

Where $C=(0,0,1,0,0,0,0,0,0)$, $d=0$, $\hat{\alpha} = \hat{\alpha}_2$, $\hat{V}(\hat{\alpha}) = V(\hat{\alpha}_2)$ and $Q_c = Q_1$ is distributed as Chi square with 1 degrees of freedom.

To calculate the power of the above test at first we have to know the distribution of the test under alternative hypothesis H_1 . Again let us consider the hypothesis that $H_0 : \alpha_2 = 0$ against the alternative hypothesis $H_1 : \alpha_2 = l \neq 0$.

Then the distribution of the test statistic under H_1 is the non central chi-square with one degrees of freedom and the non-centrality parameter

$$\delta^2 = l'(C\hat{V}(\hat{\alpha}_1)C')^{-1}l$$

We know from the above discussion the power function of the above test is

$$\psi(\delta^2, \alpha) = 1 - F_{\chi_{c,\beta}^2(\delta^2)}(p_{1-\alpha})$$

Where $p_{1-\alpha}$ denotes the $(1-\alpha)$ th percentile of the central chi-square $\chi_{1,1-\alpha}^2$ and $F_{\chi_{c,\beta}^2(\delta^2)}(p_{1-\alpha})$ is the cumulative density of the non-central chi-square distribution with 1 degrees of freedom and non centrality parameter δ^2 at $p_{1-\alpha}$.

Again for the hypothesis concerning other estimator it can be easily find the power function of the above test by the similar way.

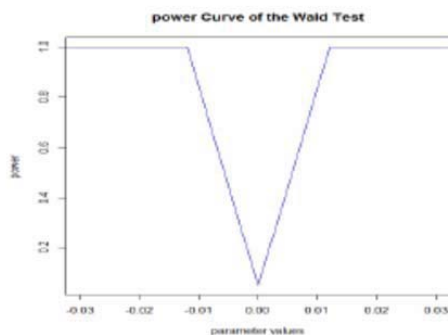


Figure 1 : power curve of α_1

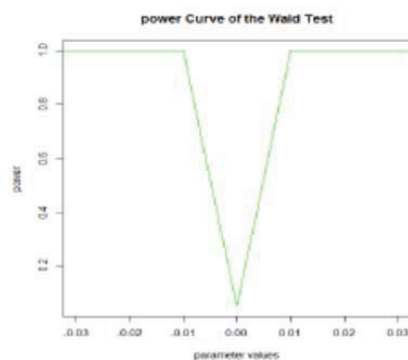


Figure 2: power curve of τ_2

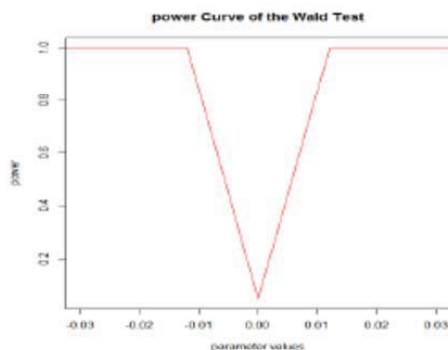


Figure 3 : power curve of τ_1

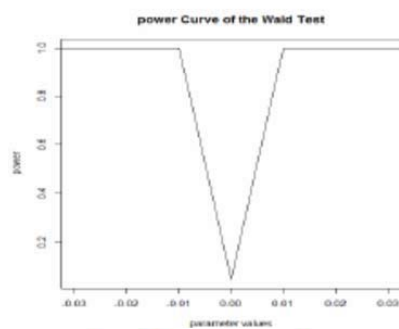


Figure 4 : power curve of α_4

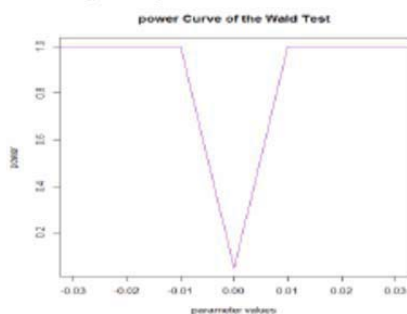


Figure 5 : power curve of α_3

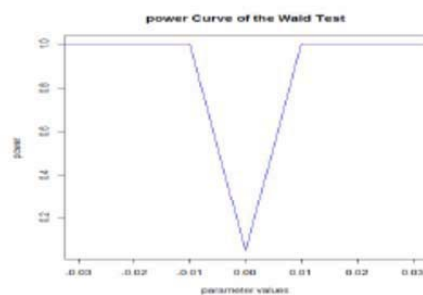


Figure 6 : power curve of τ_3

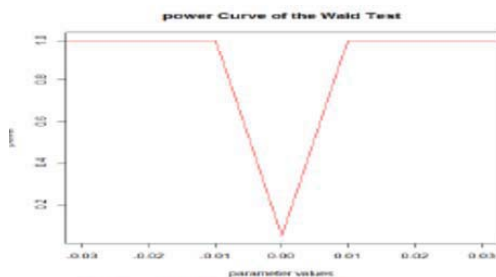


Figure 7: power curve of β_1

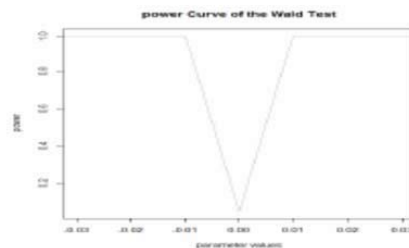


Figure 8: power curve of β_2

7. CONCLUSIONS

In our paper we consider Piecewise Linear Multiple Change Point Model with three change point. We estimate the parameter by Non-linear minimization of negative log likelihood function which is a Newton type algorithm.

We also showed the parameter estimated by Non-linear minimization method is more consistent than the estimator estimated by "Nelder-Mead simplex algorithm" and "Cross-Entropy algorithm".

We also test the significance of parameter using Wald statistics which shows that all the estimated parameter is significant. We also perform the power analysis for the estimated parameter.

Finally we can say that parameter estimated for Piecewise Linear Multiple Change Point Model with three change point by Non-linear minimization method can be considered as an improvement.

In future research one can add a spatial component to the model. By adding frailty terms to the model we can model spatial dependence across regions.

Another interesting field of further research is an implementation of the sequential (online) change point method in the hazard functions, which allows detecting change as soon as possible while the data are collected. As in survival analysis, the sequential method is more reasonable.

8. REFERENCES

- Antoch, J., Huskova', M., Jaruskov'a, D., 2000, Change point detection. In: Lecture Notes of the 5th IASC Summer School. pp. 1–76. ISI, Voorburg (2000)
- Bhore, R., Huque, M., 2008 Estimation and testing for change in hazard for time-to- event endpoints. In: Design and Analysis of Clinical Trials with Time-to-Event Endpoints. pp. 195–226. CRC Press
- Csorgo, M., Horva'th, L., 1997, Limit theorems in change-point analysis. John Wiley & Sons Inc, New York
- Dupuy, J.F., M., 2009: Detecting change in a hazard regression model with right- censoring. Journal of Statistical Planning and Inference 139(5), 1578–1586
- Evans, G.E., Sofronov, G.Y., Keith, J.M., Kroese, D.P., 2011, Estimating change- points in biological sequences via the cross-entropy method. Annals of Operations Research 189(1), 155–165
- Ferger, D., 1998, Change-point analysis of censored data. Economic Quality Control 13(3), 167–177
- Gijbels, I. and Gu'rler, U. ,2003, Estimation of a change point in a hazard function based on censored data. Lifetime Data Analysis 9, 395–411.
- Goodman, M.S., Li, Y., Tiwari, R.C., 2006, Survival analysis with change point hazard functions. Tech. rep., Harvard University Biostatistics Working Paper Series, Working Paper 40, <http://biostats.bepress.com/cgi/viewcontent.cgi?article=1043&context=harvardbiostat>
- Goodman, M.S., Li, Y., Tiwari, R.C., 2011, Detecting multiple change points in piecewise constant hazard functions. Journal of applied statistics 38(11), 2523– 2532

- Henderson, R. ,1990, A problem with the likelihood ratio test for a change-point hazard rate model. *Biometrika* 77, 835–843.
- Loader, C. R. ,1991, Inference for a hazard rate change point. *Biometrika* 78, 749–757.
- Matthews, D. E. and Farewell, V. T. ,1982, On testing for a constant hazard against a change-point alternative (Corr: V41 p1103). *Biometrics* 38, 463–468.
- Matthews, D. E. and Farewell, V. T. ,1985, On a singularity in the likelihood for a change-point hazard rate model. *Biometrika* 72, 703–704.
- Nguyen, G. S., Rogers, G. and Walker, E. A. ,1984, Estimation in change-point hazard rate models. *Biometrika* 71, 299–304.
- Pons, O. ,2002, Estimation in a Cox regression model with a change-point at an unknown time. *Statistics* 36, 101–124.
- Qian, L., Zhang, W.,2014, Multiple change-point detection in piecewise exponential hazard regression models with long-term survivors and right censoring. In: *Contemporary Developments in Statistical Theory*, pp. 289–304. Springer
- Rumana Rois, Prof. Dr. Ju'rgen Franke, Technische Universit'at Kaiserslautern Prof. Dr. Claudia Kirch, Otto-von-Guericke-Universit'at Magdeburg: *Nonparametric Tests for Change Points in Hazard Functions under Random Censorship in Survival Analysis*.
- Rubinstein, R.Y.,1997, Optimization of computer simulation models with rare events. *European Journal of Operational Research* 99(1), 89–112
- Rubinstein, R.Y., Kroese, D.P., 2004, *The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation and machine learning*. Springer-Verlag
- Rubinstein, R.Y., Kroese, D.P.,2008, *Simulation and the Monte Carlo method*. John Wiley & Sons, New York, 2nd edn.
- Worsley, K. J. ,1988 Exact percentage points of the likelihood-ratio test for a change- point hazard-rate model. *Biometrics* 44, 259–263
- Yao, Y.-C.,1986, Maximum likelihood estimation in hazard rate models with a change- point. *Communications in Statistics: Theory and Methods* 15, 2455–2466.
- Zhao, X., Wu, X., Zhou, X.,2009, A change-point model for survival data with long-term survivors. *Statistica Sinica* 19(1), 377–390.