International Journal of Computational and Applied Mathematics.
ISSN 1819-4966 Volume 12, Number 2 (2017), pp. 411-424

http://www.ripublication.com

Comparing Accelerated Failure Time Models with Its Specific Distributions in the Analysis of Esophagus Cancer Patients Data

Rinku Saikia[1]and Manash Pratim Barman[2]

*[1]Research Scholar, Department of Statistics,
Dibrugarh University, Dibrugarh , Assam,India*

*[2]Assistant Professor , Department of Statistics,
Dibrugarh University,Dibrugarh, Assam,India*

Abstract

Accelerated Failure time (AFT) model which is a regression model, used to analyze failure time data to study the reliability of machines can also be considered as a good alternative to Cox Proportional Hazard model to study the survival data with censored observations. In this paper attempt has been made to compare different forms of AFT models in the context of survival data of Esophagus cancer patient of Assam. The present retrospective study was conducted on medical records of 178 patients during 1st January 2007 to 31st December 2009. Akaike Information Criteria (AIC) and Baysian Information Criteria (BIC) and R^2test were used to identify the best fitted model. From the study it was found that Gamma AFT model was better than the other models for the esophagus cancer patients data. Accordingly interpretation were made on the basis of the Gamma AFT model.

Keywords: Accelerated failure time model, AIC, BIC, Esophagus cancer patients, Cox Proportional Hazard model.

INTRODUCTION

Accelerated failure time (AFT) model is a regression model, used to analyzed failure time data in survival analysis. It is called failure time because the event of interest is usually death, disease, remission etc. In case of survival data, though the Cox Proportional Hazard Model (PH) is most widely used but the parametric models are much more useful than the Cox PH model if the functional form of the parametric model is identified correctly (Kleinbaum and Klein).The accelerated failure time models (AFT) (Lawless, 1982) has been found useful in reliability theory and industrial experiments and it is seldom used to analyze survivorship data. The AFT model is also called log-location scale model (Lawless,1982). In AFT model the effect of covariate is proportional with respect to the survival time whereas in Cox PH model the covariate effect is multiplicatively (proportional) with respect to the hazard. In AFT model, to estimate the impact of explanatory variables on survival data, maximum likelihood function is used. In AFT model, the survival time follows some specific distributions. The members of the AFT models which are most commonly used include Exponential, Weibull, Lognormal, Log-logistic, Gamma and many more.

Parametric AFT model is an alternative to Cox PH model in survival analysis (Wei, 1992; Orbe et al., 2002). AFT model could provide more appropriate description of survival data and are more valuable and realistic alternative to Cox PH model in some situations (Venderoff,1988 ; Qi, 2009 ;). Stute (1993,1996)proposed a new methodology to deal with censored observation which can be considered as an AFT model.Orbe et al., (2002)analyzed the performance of Stute's model with PH and AFT parametric models and observed that Stute's method can be successfully applied in a context where Cox assumption also hold. The AFT model for data are size based. For uncensored data, a simple linear regression on the log scale is more natural and provides better estimators (Mandel and Ritov, 2009). AFT model yield plausible and interpretable estimates of the effect of important covariates on survival time. AFT model gives better prediction than the Cox PH model (Ponnuraja and Venkatesan ,2010). AFT models are easier interpretation not only for herpetologists but also for clinicians (Khanal et al., 2012).

The assumption of AFT model can be expressed as

$$s(t/x) = s_0(\exp(\beta' x)t) \quad for\ t \geq 0 \ldots\ldots (1)$$

Where $s(t/x)$ is the survival function at the time t and the $s_0(\exp(\beta' x)t)$ is the baseline survival function at the time t. From this equation (1), AFT model can states that the survival function of an individual with covariate x at the time t is same as the baseline survival function of the at the time $(\exp(\beta' x)t)$. The factor $\exp(\beta' x)$ is known as the acceleration factor. The acceleration factor is the key measure of association obtained in the AFT model. It is a ratio of survival times corresponding to any fixed value of survival time.

The general log-linear representation of AFT model for ith individual is given as

$$logT_i = \mu + \beta_1 x_1 + \cdots + \beta_p x_p + \sigma\varepsilon_i$$

Where $logT_i$ represents the log-transformed survival time, $x_1, \ldots\ldots.. x_p$ are the explanatory variables with the coefficients $\beta_1, \ldots\ldots.. \beta_p$, ε_i is the residual term and assumes a specific distribution and μ is the intercept and σ is the scale parameters respectively.

This main objectives of this paper are (i) to compare the AFT models with its specific distributions and find the best fitted model (ii)to study the effect of different factors on the survival of esophagus cancer patients by using the best fitted model.

MATERIALS AND METHODS

The study was taken up in a historical cohort and information from the medical charts of patients with esophagus cancer in Assam Medical College Hospital (AMCH) Dibrugarh , Assam. The period of the study was from 1st January 2007 to 31st December 2009. All the patients diagnosed with esophagus cancer during 1st January 2007 to 31st December 2008 were included in the study. Cases diagnosed during 2009 were excluded due to limited follow up (i.e., through 2009). During the inclusion period of the study a total of 178 patients were diagnosed with Esophagus cancer in AMCH. A pre-designed, pretested questionnaire was used for the collection of data. Information about age, sex, extension of the disease at the time of diagnosis, cancer directed treatment were collected from the hospitals records. The patients were considered as censored if they were alive beyond 31st December 2009, died due to other causes or loss to follow up. After collecting the hospitals records, a household survey were conducted to collect the information about the survival status of the patients, date of expired (if he/she expired), continuation of treatment and socioeconomic status of the patients. Also, a re-verification of the information collected from the hospital was made during the household visit. The extension of the disease includes the stages : localized (confined to the esophagus, with no evidence of spread to surrounding organs/tissues or no regional lymph nodes); regional (invasion beyond the organ to surrounding organs/tissues or no regional lymph nodes); distant/metastatic (spread to remote organs/tissues directly or by discontinuous metastasis) and unknown. Survival (in months) was estimated from the month of diagnosis until death, loss to follow up, or the end of 2009. Patients are categorized into three groups based on the cancer directed treatment.

The AFT models:

(i) Exponential and Weibull AFT model:

The exponential distribution was studied 1st inconnection with kinetic theory of gasses (Clausius , 1858). The survival function of T_i can be expressed by the survival function of ε_i. If the ε_i, has an extreme value distribution then T_i follows the exponentialdistribution. The survival function of Gumbel distribution is given by

$$S_{\varepsilon_i}(\varepsilon) = exp(-exp(\varepsilon))$$

The Survival function of Weibull AFT model is given by

$$S_i(t) = \exp\left[-\exp\left(\frac{logt - \mu - \beta_1 x_1 - \cdots - \beta_p x_p}{\sigma}\right)\right]$$

And the cumulative hazard function of Weibull AFT is

$$H_i(t) = -\log S_i(t) = exp\left(\frac{logt - \mu - \beta_1 x_1 - \cdots - \beta_p x_p}{\sigma}\right)$$

(ii) Log-normal AFT model:

If the ε_i, has standard normal distribution then T_i follows the log-normal distribution.The survival function of log-normal AFT model is given by

$$S_i(t) = 1 - \phi\left(\frac{logt - \mu - \beta_1 x_1 - \cdots - \beta_p x_p}{\sigma}\right)$$

The cumulative hazard function of Log-normal AFT model is

$$H_i(t) = -\log S_i(t) = -\,log\left(1 - \phi\left(\frac{logt - \mu - \beta_1 x_1 - \cdots - \beta_p x_p}{\sigma}\right)\right)$$

(iii) Log-logistic AFT model:

If the ε_i, has logistic distribution then T_i follows the log-logistic distribution .The survival function of logistic distribution is given by

$$S_{\varepsilon_i}(\varepsilon) = \frac{1}{1 + e^{\varepsilon}}$$

The survival function of log-normal AFT model is given by

$$S_i(t) = \left\{\frac{1}{1 + e^{\left(\frac{logt - \mu - \beta_1 x_1 - \cdots - \beta_p x_p}{\sigma}\right)}}\right\}$$

The cumulative hazard function of log-logistic AFT is given by

$$H_i(t) = -\log S_i(t) = \log(1 + \exp\frac{logt - \mu - \beta_1 x_1 - \cdots - \beta_p x_p}{\sigma})$$

(iv) Gamma AFT model:

In survival literature, two different gamma models are discussed. The Standard gamma model with 2 parameters and the generalized gamma model with 3 parameters. In this study the standard gamma or gamma model is used.

The probability density function of gamma model

$$f(t) = \frac{\alpha\lambda^{\alpha\gamma}}{\Gamma(\gamma)} t^{\alpha\gamma-1} \exp[-(\lambda t)^{\alpha}] \quad t > 0, \gamma > 0, \lambda > 0, \alpha > 0,$$

Where γ is the shape parameter of the distribution. The exponential, Weibull and log-normal models are all special cases of the generalized gamma model. The generalized gamma distribution becomes the exponential distribution if $\alpha = \gamma = 1$,the Weibull distribution if $\gamma = 1$, and the log-normal distribution if $\gamma \rightarrow \infty$.

Various goodness of fit Test:

(i) **AIC:** To compare various semi-parametric and parametric models Akaike

Information Criterion (AIC) is used. The AIC is proposed by Akaike (Akaike, 1974). It is a measure of goodness of fit of an estimated statistical model. In this study, AIC is computed as follows

$$AIC = -2(log - likelihhod) + 2(P + K)$$

Where P is the number of parameters and K is the number of coefficients (excluding constant) in the model. For P=1, for the exponential, P=2, for Weibull, Log-logistic, Lognormal etc. The model which as smallest AIC value is considered as best fitted model.

(i) BIC: The Baysian Information Criteria (BIC) is given by Schwarz (Schwarz, 1978). It is computed as follows

$$BIC = -2(log - likelihood) + (P + K) * \log(n)$$

Where P is the number of parameters in the distribution, K is the number of coefficients and log(n) is the number of observations. The distribution which has the lowest BIC value is considered as best fitted model.

(ii) $\boldsymbol{R^2}$**:** It isalso a goodness offit test. This statistic is calculated as follows:

$$R_p^2 = 1 - \{\exp[\frac{2}{n}(L_0 - L_p)]\}$$

Where L_p is the log-likelihood for the fitted model with p covariates and L_0 is the log likelihood for the model with no covariates.

On the basis of Akaki's Information Criteria (AIC), Baysian Information Criteria (BIC) and R^2, the best fitted model is identified. After identifying the best fitted model the effect of different variables such as age, sex, location, socio-economic status, status of the patients ,stage of the patient and cancer directed treatments are used to fit the models on the survival for this esophagus cancer patient. Cox- Snell residuals (Cox and Snell, 1968) plot is also used to fit the goodness of fit graphically. All the data are analyzed with the help of computer software package R version 3.3.1.

RESULTS

A total of 178 individuals diagnosed with esophagus cancer during the study period were included in the study. The average age of the patients were 59.13 years (s.d. 1.16 years). There were male preponderance in the sample with 67.4% and 32.6% are female. Among the patients about 27% are diagnosed at Distant stage while the stage at the time of diagnosis could not specify for 18% of the patients. The detail demographic, treatment and disease profile of the esophagus cancer patients are presented in the table I .

A parametric models including Exponential, Weibull , Log-logistic, Log-normal, Gamma were used to fit the same data to assess the best fitted model which explained the survival data of esophagus cancer. The various explanatory variables considered were sex, location, socio-economic status, stage of the patients and cancer directed treatment.

Different statistical measures such as AIC, BIC and R^2 as discussed in the methodology section were estimated for the models on the consideration. Cox-Snell residual plots were also drawn for the fitted models. By observing these plots, one can have idea about the best fitted model. The values of AIC, BIC and R^2were presented in table II . From table II, could be observed that AIC and BIC values of AFT models assuming Exponential, Weibull , Log-logistic, Log-normal and Gamma distribution were more or less similar but for the Gamma model was better less then the other parametric models . In case of R^2, all the fitted models registered more or less similar values. By observing the Cox-Snell residuals, also the Gamma AFT model was found to be best fitted which was presented in Fig1 to Fig 5. Thus we can conclude that the

Gamma AFT model is the best fitted to studying the survival time of esophagus cancer data with different covariates.

As Gamma AFT model was found to be the best fitted one, the interpretation about the effect of different independent variables on the survival times wasmade by using Gamma AFT model. The results were presented in table III. From the table it was seen that, the middle and higher socio-economic group had lower risk of dying than the lower socio- economic group. The patients underwent the cancer directed treatment other than surgery and the patients who had not taken any treatment experiencing a significantly higher risk of 1.49 times (95% C.I. 1.03 to 2.14) and 2.66 times(95% C.I. 1.59 to 4.45) respectively of dying than that of patients who underwent surgery and others treatment. The stage at the time of diagnosed was a prominent factor for better survival of esophagus cancer patients of Assam. The risk of dying among patients diagnose with Regional, Distant stage were 1.73 (95% C.I. 1.11 to 2.71) and 3.57 (95% C.I. 2.23 to 5.71) times more than that of patients diagnosed in Localized stage. The patients whose stage remain unknown at the time of diagnosis were experiencing a significantly higher role of 2.39 (95% C.I. 1.43 to 3.98). From the result presented in table III, it could be observed that residential status, sex, and age of the patients had no significant influence on the survival of esophagus cancer patients. Cancer directed treatment had a significant role.

DISCUSSION AND CONCLUSION

In this paper, attempt has been made to find the best fitted model for studying the survival time of esophagus cancer patients. To meet the objectives various AFT models following distributions Exponential, Weibull, Log-normal, Log-logisitic, Gamma are fitted to survival data of esophagus cancer patients collected from AMC, Dibrugarh , Assam. Different statistical measures such as AIC, BIC, R^2, Cox- Snell residuals plots are used to find the best fitted model. From the results, the Gamma AFT model is found to be the best fitted model.

Thought the AFT models are widely used in industrial data and seldom used in survival data but from some past research works it is seen that AFT models are more useful in various situations. Sayehmir et al.,(2008) studied prognostics factors of survival time after hematopoietic stem cell transplant in acute lymphoblastic leukemia patients in Shariati Hospital , Teharan and found that Weibull AFT model was superior to Cox PH model. Ravangard et al.,(2011) compared Cox PH model and the parametric models in studying the length of stay in a Tertiary Teaching Hospital in Tehran and showed that Gamma AFT model was best fitted for that data. Khanel et al., (2012) identified the important prognostic factor of Acute Liver Failure patients in

India by applying AFT models and found that Log-normal AFT was well fitted for that data in comparison to log-logistics. Vallinayagam et al., (2014) compared parametric models including Weibull and Log-normal with Cox PH model for bone-marrow transplantation data and Log-normal model was better fit than the other models. Vallinayagam et al., (2014) compared the performance of parametric models including the Exponential, Weibull, Log-normal, Gompertz and Log-logistic using Breast Cancer data and the study revealed that Log-normal model was better than the other models. Nawumbeni et al., (2014) compared Cox PH and AFT models in HIV/TB Co-infection survival data and revealed that Gamma model was well fitted to the Co-infection data.In the present study a number of AFT models(Exponential, Weibull, Lognormal, Log-logistic, Gamma) were tested and based on the AIC, BIC and R^2 type statistics we came to the conclusion that the AFT models with Gamma distribution was the best explained the survival data of esophagus cancer. In the study it was observed that socio-economic status, stage of the patients, cancer directed treatment had a significant impact on survival. Socio- economic status had an important role in the survival of esophagus cancer patients.

In this study, the result shows that Gamma AFT model is better than the other models in case of explaining the survival esophagus cancer data. The factor, cancer directed treatments, has a significant role in case of survival of esophagus cancer patients. The patients who undergo the cancer directed treatment other than surgery has lower risk of dying than the patients who has underwent the treatment of surgery and its combinations. Socio- economic status has also play the significant role on the survival of esophagus cancer patients. With reference to lower socio economic status patients, the middle and higher socio economic patients have higher chances of survival. The stage at diagnosis of patients is also a responsible factor in case of survival of esophagus cancer patients. The probability of survival of a patient diagnose in early stage is significantly higher than patients diagnose in advance stages. From the observations it is found that, the age of the patients at the time of diagnosis has no significant impact on the esophagus cancer patients. Also, the patients belonging to both rural and urban areas are experiencing more or less similar risk of time. The sex of the patients is also not found to be a significant factor which can influence the survival. If the distributional form of the model is correctly specified then AFT model can be provide a good alternative to Cox PH model for modeling survival data.

REFERENCES

[1]] Akaike H.(1974) : A New Look at the Statistical Model Identification. *IEEE. Transaction and Automatic Control AC*-19. 716-23.

[2] Cox D.R., and Snell E.J.(1968): A General Definition of Residuals (with discussion), *Journal of the Royal Statistical Society,* A.

[3] Khanal Prasad, Shankhar , Sreenivas,V. and Acharya,K. Subrat (2014) : Accelerated Failure Time Models: An Application in the Survival of Acute Liver Failure Patients in India, *International Journal of Science of Research (IJSR),* ISSN: 2319-7064,Impact Factor(2012):3.358,Volume 3, Issue 6,June 2014, pp 161-166.

[4] Kleinbaum,G.David and Klein, Mitchel (1996): Survival Analysis : A Self-Learning Text , Second Edition, Springer, New York.

[5] Lawless, J. F.(1982):Statistical Models and Methods for Lifetime Data Analysis, *Wiley, New York.*

[6] Nawumbeni, Ngbandor Derek , Luguterah, Albert and Adampah, Timothy(2014):Performance of Cox Proportional Hazard and Accelerated Failure Time Models in the Analysis of HIV/TB Co-infection Survival Data,*Research on Humanities and Social Sciences,* ISSN (Paper)2224-5766 ISSN (Online)2225-0484 (Online), Vol.4, No.21, pp 94-102.

[7] Orbe Jesus, Ferreira Eva and Nunez- Antion, Vicente (2002): Comparing Proportional Hazards and Accelerated Failure Time models for Survival Analysis, *Statistics in Medicine.* 21: pp 3493-3510.

[8] Ponnuraja, C., and Venkatesan, P (2010): Survival Models for Exploring Tuberculosis Clinical Trial Data-an Empirical Comparison, Indian Journal of Science and Technology*,* 3(7): 0974- 6846.

[9] Qi,Jiezhi,(2009):Comparison of Proportional Hazards and Accelerated Failure Time Models, A Thesis Submitted to the College of Graduate Studies and Research in Partial Fulfillment of the Requirements for the Degree of Master of Science in the Department of Mathematics and Statistics University of Saskatchewan Saskatoon, Saskatchewan.

[10] Ravangard, R., Arab M., Rashidian, A.,Akbarisari, A., Zare A., Zeraat, H. (2011): Compared Cox Model and Parametric Models in the Study of Length of Stay in a Tertiary Teaching Hospital in Tehran, Iran. *Acta MedicaIranica*, 49(10): 650-658.

[11] Sayehmiri, K., Eshraghian R. M., Mohammad K., Alimoghaddam K., Foroushani R. A., Zeraati H., Golestan, A., Ghavamzadeh, A. (2008): Prognostic Factors of Survival time after Hematopoietic Stem Cell Transplant in acute Lymphoblastic Leukemia Patients in Shariati Hospital, Tehran , *Journal ofExperimental & Clinical Cancer Research,* **27**(1): 1-9.

[12] Stute W. (1993): Consistent Estimation under Random Censorship when Co variables are Present , *Journal of Multivariate Analysis,* 45: pp. 89-103.

[13] Stute W.(1996):Distributional Convergence Under random Censorship when Covariables are Present Scan dinavian „*Journal of Statistics*, 23, pp. 461- 471.

[14] Vallinayagam, V., Prathap, S., and Venkatesan, P. (2014): Parametric Regression Models in the Analysis of Breast Cancer Survival Data *, International Journal of Science and* Technology*,* 3(3) 2049-7318.

[15] Vanderhoeft ,Camille and Vanderhoeft ,Camille (1982): Accelerated Failure Time Models: An Application To Current Status Breast- Feeding Data From Pakistan,Vol.38, No.1/2(Gennaio- Giugno), pp. 135-157.

[16] Wei,L.J.,(1992): The Accelerated Failure Time Mode: A Useful Alternative To the Cox Regression Model in Survival Analysis, *Statistics in Medicine,* 11, pp.1871-1879.

APPENDIX

Table I: Demographic, Treatment and Disease profile of the esophagus cancer patients

Characteristics	Frequency(%)
Location	
Rural	84 (47.2)
Urban	94 (52.8)
Sex	
Male	120 (67.4)
Female	58 (32.6)
Age	
Less than 50	33 (18.5)
50 to70	118 (66.3)
Above 70	27 (15.2)
Cancer Directed Treatment	
Surgery & others	49 (27.5)
Other than Surgery	106 (59.5)
No treatment	23 (12.9)
Socio- economic status	
Lower	24 (13.5)
Middle	133 (74.7)
Higher	21 (11.8)
Stage	
Localized	34 (19.1)
Regional	67 (37.6)
Distant	49 (27.5)
Unknown	28 (15.7)

Table II : Goodness of fit of the Model on the basis of AIC and BIC and R^2

Models	AIC	BIC	R^2
Exponential model	1928.06	1966.24	32.83%
Weibull model	1920.67	1961.93	36.21%
Log-Normal model	1926.08	1967.44	35.74%
Gamma model	1919.49	1960.84	36.61%
Log-Logistic	1919.75	1961.12	37.06%

Table III : Results of the Gamma AFT model

Characteristics	Hazard Ratio(HR)	95% confidence Interval
Location		
Rural	Reference	Reference
Urban	1.22	.91-1.63
Sex		
Male	Reference	Reference
Female	.97	.73-1.33
Age		
Less than 50	Reference	Reference
50 to70	.97	.66-1.42
Above 70	1.49	.92-2.39
Cancer Directed Treatment		
Surgery & others	Reference	Reference
Other than Surgery	1.49	1.03-2.14
No treatment	2.66	1.59-4.45

Socio- economic status		
Lower	Reference	Reference
Middle	.54	.36-.81
Higher	.42	.23-.76
Stage		
Localized	Reference	Reference
Regional	1.73	1.11-2.71
Distant	3.57	2.23-5.71
Unknown	2.39	1.43-3.98

Fig :Cox Snell Residuals forvarious AFT models

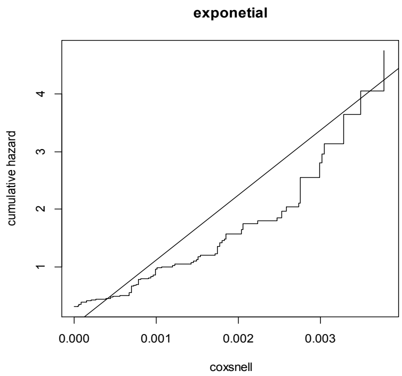


Fig1. Exponential AFT model

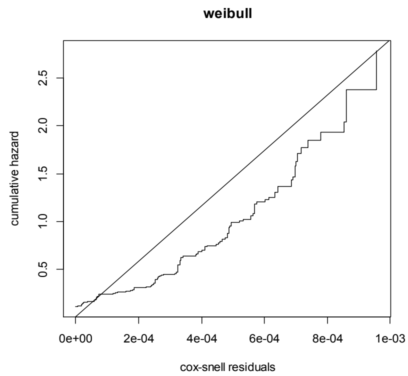
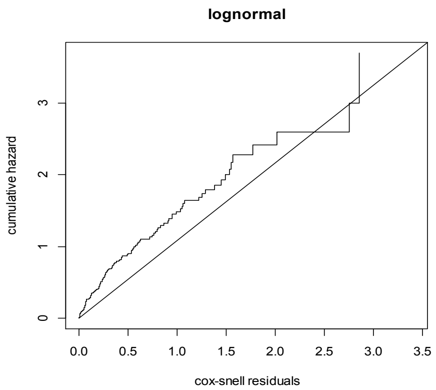


Fig2. Weibull AFT model



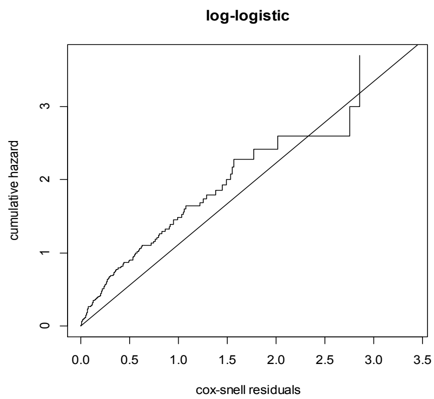


Fig3. Lognormal AFT modelFig4: Log-logistic AFT model

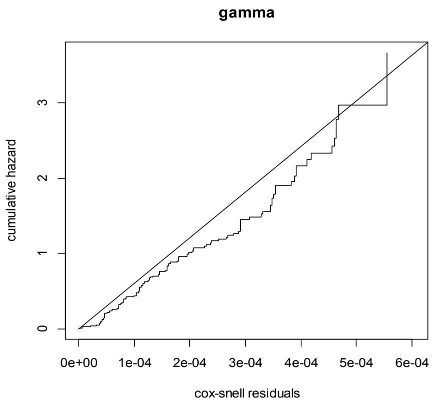


Fig5: Gamma AFT model