American Journal of Mathematics and Statistics 2014, 4(5): 205-213
DOI: 10.5923/j.ajms.20140405.01

Parametric Models and Future Event Prediction Base on Right Censored Data

Joseph. O. Okello[1,*], **D. Abdou Ka**[2]

[1]Pan Africa University Institute of Basic Sciences, Technology and Innovation (PAUISTI), Nairobi, Kenya
[2]Department of statistics, University of Gaston Berger, Senegal and a visiting lecturer to (PAUISTI), Nairobi, Kenya

Abstract In this paper, we developed a parametric displacement models base on the time that the Internally Displaced Persons (IDPs) took to return from IDPs Camps to their ancestral homes in Northern Uganda. The objective is to analyze the displaced proportion of the IDPs using suitable time-to-event parametric models. The accelerated failure time (AFT) models (Weibull, Exponential and log-logistic) were considered. A retrospective data of seven years study of 590 subjects is considered. Maximum likelihood method together with the Davidon-Fletcher- Powell optimization algorithm in MATLAB is used in the estimation of the parameters of the models. The estimated displaced proportions of these AFT models are used in predicting the displaced proportion of the IDPs at a time t. Weibull and exponential models provided better estimates of the displaced proportion of the IDPs due to their good convergence power to four decimal points and predicted the 2027 and 2044 respectively as the year when the displaced proportion can be approximated to be zero.

Keywords Survival Techniques, Parametric Models (Weibull, Exponential and Log-logistic), Displaced Proportion, Retrospective Data, Lord Resistance Army (LRA)

1. Introduction

The underlying foundation of most inferential statistical analysis is the concept of probability distribution. An understanding of probability distribution is critical in using quantitative methods such as hypothesis testing, regression analysis, and time-series analysis. The mathematical expression that describes the individual probabilities that a random variable will take on each of a set of specified values is known as its probability density function. In life data analysis, the practitioner attempts to make predictions about the life of all products in the population by fitting a statistical distribution to life data from a representative sample of units. The parameterized distribution for the data set can then be used to estimate important life characteristics of the product such as reliability or probability of failure at a specific time, the mean life and the failure rate.

In this paper, we present the parametric distributions of the accelerated failure time models (Weibull, exponential and Log-logistic) for the analysis of the time the internally displaced persons took to return from IDPs camps to their ancestral homes. The paper takes the form of case study in which 590 families displaced by the lord resistance army in Northern Uganda were studied. The data was previously modelled parametrically by [18]to test for the distribution fit in which Weibull regression model showed a superior fit.

There is already a very wide literature on parametric distribution (Weibull, Exponential and Log-logistic) in analyzing the time-to-event data, for instance [11], [13], [16], and [21]. The Weibull and Exponential regression model have been used in medical research in [19] to model survival data of CABG patients. On the other hand researchers such as: [1], [6], [9], [10], [11] [12], [13] and [14] provide literature on parametric regression analysis of time-to-event data. The primary advantage of Weibull analysis has been stressed out by [1] as the ability to provide reasonably accurate failure analysis and failure forecasts with extremely small samples and providing a simple and useful graphical plot of the failure data. Furthermore, [1] maintain that AFT interpretation is usually presented in coefficients where Positive coefficient means increasing that covariate extends the time until failure which is the opposite of the proportional hazard covariate coefficient interpretation where positive coefficient increases the hazard, therefore decreasing the time until failure. Exponential model is a special case of the Weibull distribution model. In spite of the wide literature provided on parametric models, we feel that there is some important attribute worth discussing. First, although there are large literatures on application of parametric regression models in estimating the time to event, most of the events of interest are always negative occurrences such as death from a certain disease, failure of a machine parts and above all too much leaning toward hard sciences. Furthermore many data sets have been consider in the study of such kind but there is no

* Corresponding author:
j.okello@ymail.com (Joseph. O. Okello)

Published online at http://journal.sapub.org/ajms

attempt involving the time-to-return of the internally displaced persons to their ancestral homes after war which is a very important social attribute. In this paper, we use the parametric regression models (Weibull, exponential and Log-logistic) to predict the time that all the internally displaced persons would have returned to their ancestral homes after being displaced by the Lord Resistance army war in Northern Uganda using a sample of 590 displaced families from seven different villages in Otuke district. The idea is to estimate the parameters of the distribution of interest and used it in the formulation of the displaced function which has the same properties as that of a survival function in medical research or reliability function in engineering research. The displaced function estimates the displaced proportions of the IDPs at a given time.

2. Analysis Techniques

According to [3], parametric, non-parametric and semi-parametric techniques are the three well known techniques used for analyzing the time-to-event data, each with its own limitation but parametric approach is thought to yield better results provided the assumption made in the analysis are correct. With Parametric models, the outcome is assumed to follow a certain known distribution. There are a number of texts that discuss comprehensively parametric time-to-event-models such as; [5], [7], [9], [13], [15], [17] and [19]. For instance [15], suggests that exponential, Weibull, lognormal and gamma distribution are the most commonly used parametric models in analyzing time-to-event data.

According to [4], Survival analysis is a phrase used to describe the analysis of data in the form of time from a well-defined time origin until the occurrence of the particular event of interest or the end point of the study. On the other hand, [2] defined survival analysis as a class of statistical techniques used for studying the occurrence and timing of events. They were originally designed for the event of death occurrence and hence name survival analysis. The techniques is extremely useful for studying many different kinds of events in both the social and natural sciences research, such as the onset of disease in Biostatistics, equipment failures in engineering, earthquakes, automobile accidents, stock market crashes, revolutions, job terminations, births, marriages, divorces, promotions in job places, retirements, contracting Lung cancer due to smoking, arrests and many other time to event data. In Biostatistics this techniques is often referred to as Clinical trials; in engineering the term is referred to Reliability or failure time analysis; in econometric it is either duration analysis or transition analysis; and in Sociology it is often referred to as event history analysis; We therefore apply the technique to analyze the time the internally displaced persons took to return to their ancestral homes using the sample of seven villages and draw inference for the bigger population. The study has a leaning toward sevent history analysis. The time origin is when the Ugandan Government declared the villages safe in 2006 after signing of the truce and formation of satellite camps.

According to [20], time-to-event analysis is frequently used with *retrospective* data in which subjects are asked to recall the dates when the events of interest happened to them. This was the case employed in this study where subjects were asked to recall the year when they returned to their ancestral homes and the censored subjects' information were extracted from the record kept by the Local Council Chairpersons of the seven villages. Our study therefore considered a retrospective data of 590 subjects that were previously studied by Okello, Odongo and AbdouKa in [18]. The study period was between the years 2007 to 2013. The uncensored subjects were those whose return times were known and the censored subjects were those whose return time is unknown may be because they had not yet returned to their ancestral homes by the end of 2013 or had died within the study time. This generated a right censored data set.

Several researches have been conducted using the technique of time-to-event analysis for many case studies. Although much of the work in this paper pays much attention to internally displaced persons return time and prediction of the return event, the explored methods of parametric model are much more general. They can be applied to any study of time-to-event analysis.

3. Methodology

3.1. Introduction

In this paper, the parameters of Weibull, Exponential and Log-logistic distribution are estimated based on censored data of the IDPs return time to their ancestral homes. The uncensored observation under this study were the subjects who have resumed their ancestral homes within the predetermine study period and the censored subjects are those whose time of return are not known. The status of the subjects was defined as:

$$\delta_i = \begin{cases} 1 & if\ the\ i^{th} person\ \ retuned\ time\ is\ known \\ 0 & if\ the\ i^{th} person\ \ retuned\ time\ is\ unknown \end{cases}$$

The status contribution to the likelihood function for the subject who have returned to their ancestral home would be $f(t_i;\underline{\theta})$ and for those who have not returned to their ancestral home yet would be $D(t_i;\underline{\theta})$

Lawless in [13] proposed the form of likelihood function for the survival model in the presence of censored data. The maximum likelihood method works by developing a likelihood function based on the available data and finding the estimates of parameters of a probability distribution that maximizes the likelihood function. The likelihood function for all observed and censored Subjects is of the form:

$$L(t_i,\underline{\theta}) = \prod_{i\in u}[f(t_i,\underline{\theta})] \times \prod_{i\in c} D(t_i;\ \underline{\theta})$$

$$L(t_i,\underline{\theta}) = \prod_{i=1}^{n}[f(t_i,\underline{\theta})]^{f_{t_i}} \prod_{i=1}^{n}[D(t_i;\ \underline{\theta})]^{c_{t_i}}$$

where f_{t_i} are the number of observed subjects until the event of interest has happened in the interval i and c_{t_i} are the number of censored individuals in the interval i each of length t, $f(t_i,\underline{\theta})$ is probability density function (pdf) in a parametric model with displaced function, $D(t_i,\underline{\theta})$ and the hazard function, $h(t_i,\underline{\theta})$ with the vector parameter $\underline{\theta}$ of the model. To obtain maximum likelihood estimates of parameters of a distribution of interest, we take the negative natural logarithm of the Likelihood function. i.e. the log-likelihood function $l(t_i,\underline{\theta}) = -lnL(t_i,\underline{\theta})$ result into:

$$l(t_i,\underline{\theta}) = -\sum_{i=1}^{n} f_{t_i} ln\,[f(t_i,\underline{\theta})] - \sum_{i=1}^{n} c_{t_i} ln\,[D(t_i,\underline{\theta})]$$

Since $f(t_i,\underline{\theta}) = h(t_i,\underline{\theta}) \times D(t_i,\underline{\theta})$, then

$$l(t_i,\underline{\theta}) = -\sum_{i=1}^{n} f_{t_i} ln\,[h(t_i,\underline{\theta})] - \sum_{i=1}^{n} (f_{t_i} + c_{t_i}) ln\,[D(t_i,\underline{\theta})]$$

Where, the first summation is for failure and the second summation is for all censored individuals.

Letting $N_{t_i} = (f_{t_i} + c_{t_i})$, the total number of failed and censored subjects at time t_i, of the i^{th} interval then

$$l(t_i,\underline{\theta}) = -\sum_{i=1}^{n} f_{t_i} ln\,[h(t_i,\underline{\theta})] - \sum_{i=1}^{n} N_{t_i} ln\,[D(t_i,\underline{\theta})]$$

In this study time is partitioned into intervals, which are of unit length t starting from zero. Moreover, failures and censoring of the subjects occurs in each interval I of equal length of time t, i=1,2, …, n

For the maximum likelihood estimation of the parameters of a distribution based on censored data, there is need to find out the hazard function and the survival function (Displaced function) to be substituted in the log likelihood function and hence apply suitable iteration techniques to come out with the parameter estimates.

3.2. Survival (Displaced) Function

This is the probability that the event of interest has not occurred on a subject by time t. for our case, it is the probability that an individual displaced has not return to his/her ancestral home by time t. mathematically, for the parametric regression model, the displaced function is defined by

$$D(t;\underline{\theta}) = \int_{t}^{\infty} f(x;\underline{\theta})dx$$

Where; $\underline{\theta}$ is the vector of the parameters of the distribution and $f(x;\ \underline{\theta})$ is the probability density function of the distribution under consideration which for the case of this study will be Weibull, exponential and log-logistic distribution function.

3.3. Hazard Function

This is also called the force of mortality in Biostatistics and epidemiology especially in clinical trials. It is the instantaneous failure rate. For the case of this study it is the instantaneous return rate. Mathematically it is defined by

$$h(t_i;\underline{\theta}) = \frac{f(t_i;\underline{\theta})}{1 - F(t_i;\underline{\theta})} = \frac{f(t_i;\underline{\theta})}{D(t_i;\underline{\theta})}$$

$P($Eexperiencing the event of interest in the interval $(t, t+\delta_t)|$ survived past time, $t)$

3.4. Maximum Likelihood Method

(MLE) is used in the estimation of the parameters of the distribution of interest. The contribution of the subject status into the likelihood function is defined by

$$L(t_i,\delta_i) = \begin{cases} f(t_i;\underline{\theta})\ if\ \delta_i = 1(uncensored) \\ D(t_i;\ \underline{\theta})\ if\ \delta_i = 0(censored) \end{cases}$$

The contribution of individual I into the likelihood function is defined by

$$L(t_i,\delta_i) = \ [f(t_i;\ \underline{\theta})]^{\delta_i} \times [S(t_i;\ \underline{\theta})]^{1-\delta_i}$$

For the full sample in the entire period of study

$$L(t_1,\dots,t_n;\delta_1,\dots,\delta_n) = \prod_{i=1}^{n} L(t_i,\delta_i)$$

$$L(t_i,\delta_i) = \prod_{i=1}^{n}[f(t_i;\underline{\theta})]^{\delta_i} \times [D(t_i;\underline{\theta})]^{1-\delta_i}$$

$$L(t_i;\delta_i) = \prod_{i\in u}[f(t_i,\underline{\theta})] \times \prod_{i\in c} D(t_i;\ \underline{\theta})$$

Where $\prod_{i\in u}$, denote the product over the uncensored observation and $\prod_{i\in c}$, the product over the censored observation.

To estimate the parameters of interest, we take negative logarithm of the likelihood function above and by denoting $l(t_i;\underline{\theta}) = -\ln\ (L(t;\underline{\theta}))$, then

$$l(t_i;\underline{\theta}) = -\ln\left\{\prod_{i=1}^{n}[f(t_i;\underline{\theta})]^{\delta_i} \times [S(t_i;\underline{\theta})]^{1-\delta_i}\right\}$$

But $f(t_i;\underline{\theta}) = h(t_i;\underline{\theta})S(t_i;\underline{\theta})$. Substituting for $f(t_i;\underline{\theta})$ in the likelihood function, we get

$$l(t_i;\underline{\theta}) = -\ln\left\{\prod_{i=1}^{n}[h(t_i;\underline{\theta})S(t_i;\underline{\theta})]^{\delta_i} \times [S(t_i;\underline{\theta})]^{1-\delta_i}\right\}$$

$$l(t_i;\underline{\theta}) = -\sum_{i=1}^{n} \delta_i ln\ h(t_i;\underline{\theta}) - \sum_{i=1}^{n} \delta_i ln\ S(t_i;\underline{\theta})$$

$$-\sum_{i=1}^{n}(1-\delta_i) ln\, D(t_i;\underline{\theta})$$

If the total number of the internally displaced persons returning to their ancestral homes at time interval t_i is f_{t_i} and the total censored individuals at time interval t_i is

c_{t_i}, then

$$l\left(t_i;\underline{\theta}\right) = -\sum_{i=1}^{k} f_{t_i} ln\ h\left(t_i;\underline{\theta}\right) - \sum_{i=1}^{k} f_{t_i} ln\, D\left(t_i;\underline{\theta}\right) - \sum_{i=1}^{k} c_{t_i} ln\, D\left(t_i;\underline{\theta}\right)$$

$$l\left(t_i;\underline{\theta}\right) = -\sum_{i=1}^{k} f_{t_i} ln\ h\left(t_i;\underline{\theta}\right) - \sum_{i=1}^{k} \left(f_{t_i} + c_{t_i}\right)\ ln\, D\left(t_i;\underline{\theta}\right)$$

$$l\left(t_i,\underline{\theta}\right) = -\sum_{i=1}^{n} f_{t_i} ln\left[h\left(t_i,\underline{\theta}\right)\right] - \sum_{i=1}^{n} N_{t_i} ln\left[D\left(t_i,\underline{\theta}\right)\right] (1)$$

3.5. Application

3.5.1. Weibull Distribution Model

The probability distribution function (*pdf*) of a Weibull distribution defined by:

$$f\left(t;\ \underline{\theta}\right) = \left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1} e^{-\left(\frac{t}{\alpha}\right)^{\beta}}$$

$$t \geq 0 \quad \alpha > 0\,, \qquad and\,\beta > 0$$

Where $\underline{\theta}$ is the vector of $\alpha$the scale parameter, and β the shape parameter also known as Weibull slope.

The displaced function (the probability that an IDP has not returned to their ancestral homes by time t) for a Weibull distribution model is given by

$$D\left(t_i;\underline{\theta}\right) = e^{-\left(\frac{t_i}{\alpha}\right)^{\beta}}$$

The hazard function of a Weibull distribution model is given by:

$$h\left(t_i;\underline{\theta}\right) = \left(\frac{\beta}{\alpha}\right)\left(\frac{t_i}{\alpha}\right)^{\beta-1}$$

Replacing values of the displaced and hazard functions of Weibull distribution model in equation (1), we get

$$l\left(t;\underline{\theta}\right) = -(F)ln\left(\frac{\beta}{\alpha}\right) - (\beta - 1)\sum_{i=1}^{k} f_{t_i} ln\left(\frac{t_i}{\alpha}\right) + \sum_{i=1}^{k} N_{t_i}\left(\frac{t_i}{\alpha}\right)^{\beta} \qquad (2)$$

where, $F = \sum_{i=1}^{k} f_{t_i}$ is the total number of returned subjects in a given time, k is the maximum time interval and $N_{t_i} = \left(f_{t_i} + c_{t_i}\right)$ is the total number of failure and censored subjects in a given time interval.

Differentiating (2) with respect to $\alpha\, and\, \beta$ and simplifying we get

$$\frac{\partial l\left(t_i;\underline{\theta}\right)}{\partial\alpha} = F\left(\frac{\beta}{\alpha}\right) - \left(\frac{\beta}{\alpha}\right)\sum_{i=1}^{n} N_{t_i}\left(\frac{t_i}{\alpha}\right)^{\beta} \qquad (3)$$

$$\frac{\partial l\left(t_i;\underline{\theta}\right)}{\partial\beta} = -\left(\frac{F}{\beta}\right) - \sum_{i=1}^{n} f_{t_i} ln\left(\frac{t_i}{\alpha}\right) + \sum_{i=1}^{n} N_{t_i}\left(\frac{t_i}{\alpha}\right)^{\beta} ln\left(\frac{t_i}{\alpha}\right) \quad (4)$$

By using (2), (3) and (4) in the DFP optimization method, we obtain the parameter estimates for which value of the likelihood function of the Weibull distribution is maximum. MATLAB program for the parameters estimation of the Weibull models is developed.

3.5.2. Exponential Model

The probability distribution function of an exponentially distributed random variable, t with mean μ is defined by

$$f(t_i;\ \mu) = \left(\frac{1}{\mu}\right) e^{-\left(\frac{t_i}{\mu}\right)}\ t > 0\ \&\ \mu > 0$$

With the displaced and the hazard functions defined by

$$D(t_i;\mu) = e^{-\left(\frac{t_i}{\mu}\right)}$$

And

$$h(t_i;\mu) = \left(\frac{1}{\mu}\right)$$

Replacing values of the displaced and hazard functions of Exponential distribution model in equation (1), we get

$$l(t_i;\mu) = -\sum_{i=1}^{k} f_{t_i} ln\left(\frac{1}{\mu}\right) - \sum_{i=1}^{k}\left(f_{t_i} + c_{t_i}\right)\ ln\left[e^{-\left(\frac{t_i}{\mu}\right)}\right]$$

Simplifying the equation we get

$$l(t_i;\mu) = -(F)ln\left(\frac{1}{\mu}\right) + \left(\frac{1}{\mu}\right)\sum_{i=1}^{k} N_{t_i} \times t_i$$

Differentiating with respect to μ, we get

$$\frac{\partial l\left(t_i;\mu\right)}{\partial\mu} = \frac{F}{\mu} - \left(\frac{1}{\mu^2}\right)\sum_{i=1}^{k} N_{t_i} \times t_i \qquad (5)$$

Making μ the subject by equating (5) to zero we get

$$\hat{\mu} = \frac{\sum_{i=1}^{k} N_{t_i} \times t_i}{\sum_{i=1}^{k} f_{t_i}} \qquad (6)$$

Equation (6) is solved analytically to find the value of the parameter estimate for which value of the likelihood function of the exponential distribution is maximum.

3.5.3. Log-logistic Distribution

A random variable t has a log-logistic distribution if the logarithm of t has a logistic distribution.

The probability distribution function (*pdf*) of a log-logistic distribution model is defined by:

$$f(t;\ \gamma,\sigma) = \frac{\left(\frac{\gamma}{\sigma}\right)\left(\frac{t}{\sigma}\right)^{\gamma-1}}{\left[1 + \left(\frac{t}{\sigma}\right)^{\gamma}\right]^2};\quad t > 0, \gamma > 0\ and\ \sigma > 0$$

Here, γ is the shape parameter and σ is the scale parameter.

The Displaced and the hazard functions of the log-logistic distribution function is defined by

$$D(t_i;\gamma,\sigma) = \frac{1}{1 + \left(\frac{t_i}{\sigma}\right)^{\gamma}}$$

And

$$h(t_i;\gamma,\sigma)=\frac{\left(\frac{\gamma}{\sigma}\right)\left(\frac{t}{\sigma}\right)^{\gamma-1}}{1+\left(\frac{t_i}{\sigma}\right)^{\gamma}}$$

Replacing values of the displaced and hazard functions of log-logistic distribution model in equation (1), we get

$$l(t_i;\gamma,\sigma)=-\sum_{i=1}^{k}f_{t_i}ln\ [\frac{\left(\frac{\gamma}{\sigma}\right)\left(\frac{t_i}{\sigma}\right)^{\gamma-1}}{1+\left(\frac{t_i}{\sigma}\right)^{\gamma}}]-\sum_{i=1}^{k}N_{t_i}ln\left[\frac{1}{1+\left(\frac{t_i}{\sigma}\right)^{\gamma}}\right]$$

Simplifying the expression

$$l(t_i;\gamma,\sigma)=-(F)ln\left(\frac{\gamma}{\sigma}\right)-(\gamma-1)\sum_{i=1}^{k}f_{t_i}ln\left(\frac{t}{\sigma}\right)$$

$$+\sum_{i=1}^{k}f_{t_i}ln\left[1+\left(\frac{t}{\sigma}\right)^{\gamma}\right]+\sum_{i=1}^{k}N_{t_i}ln\left[1+\left(\frac{t}{\sigma}\right)^{\gamma}\right] \quad (8)$$

Differentiating (8) with respect to γ and σ we get

$$\frac{\partial l(t_i;\gamma,\sigma)}{\partial\sigma}=\left(F\left(\frac{\gamma}{\sigma}\right)\right)-\left(\frac{\gamma}{\sigma}\right)\sum_{i=1}^{k}f_{t_i}\left[\frac{\left(\frac{t_i}{\sigma}\right)^{\gamma}}{1+\left(\frac{t_i}{\sigma}\right)^{\gamma}}\right]$$

$$-\left(\frac{\gamma}{\sigma}\right)\sum_{i=1}^{k}N_{t_i}\left[\frac{\left(\frac{t_i}{\sigma}\right)^{\gamma}}{1+\left(\frac{t_i}{\sigma}\right)^{\gamma}}\right] \quad (9)$$

$$\frac{\partial l(t_i;\gamma,\sigma)}{\partial\gamma}=-\left(\frac{F}{\gamma}\right)-\sum_{i=1}^{k}f_{t_i}ln\left(\frac{t_i}{\sigma}\right)+\sum_{i=1}^{k}f_{t_i}\left[\frac{\left(\frac{t_i}{\sigma}\right)^{\gamma}ln\left(\frac{\gamma}{\sigma}\right)}{1+\left(\frac{t_i}{\sigma}\right)^{\gamma}}\right]$$

$$+\sum_{i=1}^{k}N_{t_i}\left[\frac{\left(\frac{t_i}{\sigma}\right)^{\gamma}ln\left(\frac{\gamma}{\sigma}\right)}{1+\left(\frac{t_i}{\sigma}\right)^{\gamma}}\right] \quad (10)$$

By using (8), (9) and (10) in the DFP optimization method, we find the parameters estimates for which value of the likelihood function of the log-logistic models is maximum. MATLAB program for the parameters estimation of the log-logistic distribution model is developed.

3.6. Variance-Covariance Estimates

The asymptotic variance-covariance matrix is obtained by inverting the information matrix by elements that are negatives of the expected values of the second order derivatives of the log-likelihood function. In this paper we used the negative log-likelihood instead of the log-likelihood and as such we only get the expectation of the second derivatives and invert to get the asymptotic variance covariance matrix. We approximate the expected values by their respective maximum likelihood estimates.

4. Results, Discussion and Conclusions

4.1. Result of Estimations

Equation (2), (3) and (4) are used in the MATLAB DFP program to obtain the parameters estimates of the Weibull distribution models. Equation (6) can be solved analytically to obtain the parameter estimates of the exponential distribution model. Equation (8), (9) and (10) are used in the MATLAB DFP program to obtain the parameters estimates of the log-logistic distribution model. The displaced proportions $\hat{y}_E\hat{y}_W$ and $\hat{y}_L$ of IDP of exponential, weibull and log-logistics models respectively are obtained for parameter estimates. $\hat{y}$, the Kaplan-Meier displaced proportions of the IDPs return time. The optimal estimates of parameters obtained by maximizing the log-likelihood function of the AFT models are given below in **tables 1, 2** and **3**.

Equation (3) and (4) are differentiated with respect to α and β to obtain the information necessary in formulating the information matrix which is in turn used in estimating the asymptotic variance covariance estimates of the Weibull distribution model. Equation (5) is differentiated with respect to μ to get the second order derivative of the exponential negative log-likelihood function necessary in computing the variance estimate of the parameter estimate. Similarly differentiating equation (9) and (10) with respect to γ and σ we obtain a system of second order derivatives of the negative log-likelihood function of the log-logistic distribution model that is used in the estimation of the asymptotic variance-covariance matrix of the log-logistic parameter estimates.

The estimated values of scale parameter and shape parameter for *IDP* are given in the tables 1, 2 and 3 along with their t-ratios. In case of Weibull distribution the estimated value of the shape parameter is greater than 1 which indicates increasing failure rate with time (increasing return rate). The positive value of co-variance for the Weibull distribution model indicates that the movements of α and β parameters are in the same directions. The negative value of co-variance for log-logistic distribution models of the IDPs indicates that the movements γ and σ are in the opposite direction.

The estimated values of scale parameter and shape parameter for *IDPs* are all more than zero and are given in the tables 1, 2 & 3 along with their *t*-ratios, indicating that the estimates of scale and shape parameters are significant at 5% level of significance.

Table 1. Estimates of Parameters of Weibull Distribution

Parameters	Estimates	t-ratios	Gradients
Scale, α	3.8628	29.5094	5.6843×10^{-14}
Shape, β	1.3420	27.8037	8.2414×10^{-5}
Log-Likelihood	1.1190×10^{3}		
Varance-covariance matrix	$\begin{bmatrix}1.7135\times10^{-2} & 9.2087\times10^{-4}\\ 9.2087\times10^{-4} & 2.3297\times10^{-3}\end{bmatrix}$		

Table 2. Estimates of Parameters of Exponential Models

Parameters	Estimates	t-ratios	Gradient
Mean, μ	3.7591	4.4883	2.8422×10^{-14}
Log-Likelihood	1.1481×10^{3}		
variance	2.8605×10^{-2}		

Table 3. Estimates of Parameters of log-logistic Distribution

Parameters	Estimates	t-ratios	Gradients
Shape, γ	2.01766	68.7976	-5.6843×10^{-14}
Scale , σ	2.57182	36.8677	2.1304×10^{-5}
Log-Likelihood	8.4151×10^{2}		
varance-covariance matrix	$\begin{bmatrix} 8.6010 \times 10^{-4} & -3.7941 \times 10^{-3} \\ -3.7941 \times 10^{-3} & 4.8662 \times 10^{-3} \end{bmatrix}$		

Table 4. Displaced proportion for Kaplan-Meier, Weibull, exponential and Log-logistic regression models

Time (t_i)	Kaplan-Meier $(\hat{y})$	Exponential $(\hat{y}_E)$	Weibull $(\hat{y}_W)$	Log-logistic $(\hat{y}_L)$
0	1.0000	1.0000	1.0000	1.0000
1	0.7169	0.7664	0.8495	0.8706
2	0.4974	0.5874	0.6614	0.6242
3	0.3385	0.4502	0.4905	0.4229
4	0.2757	0.3450	0.3507	0.2909
5	0.2430	0.2644	0.2432	0.2073
6	0.1867	0.2027	0.1644	0.1533
7	0.1369	0.1553	0.1085	0.1171

The values of estimated displaced proportions of internally displaced person from the Weibull, exponential, and log-logistic regression models are given belowin **table 4** together with the Kaplan-Meier displaced proportion and the corresponding graphs (Displaced proportion curves) in **Fig. 1, 2, 3** and **4**. The table shows the displaced proportion column based on the lower, point and upper interval. Since there is no close form of the Displaced function (survival) for a Gamma distribution then it displaced proportion and graphs required more sophiscated program which was lacking.

4.2. Discussion

Accelerated Failure Time model interpretation is usually presented in coefficients: Positive coefficient means increasing that covariate extends the time until failure which is the opposite of the proportional hazard covariate coefficient interpretation where positive coefficient increases the hazard, therefore decreasing the time until failure.

The estimated value of the shape parameter of the Weibull regression model shown in table.1 is more than one which implies an increasing rate of return of the IDPs. On the other hand, the estimated value of the shape parameter of a log-logistic distribution shown in table.3 is more than one implying increasing and decreasing return rate of the IDPs. While the estimated value of the mean parameter of the exponential regression model shown in table.2 implies a constant return rate of the IDPs of 0.26602.

Using the estimated parameters in table.1, then the displaced function of the Weibull distribution model is defined by

$$D(t; 3.8628, 1.3420) = e^{-\left(\frac{t}{3.8628}\right)^{1.3420}}$$

As $t \to 21, D(t; 3.8628, 1.3420) \to 4.2699 \times 10^{-5}$

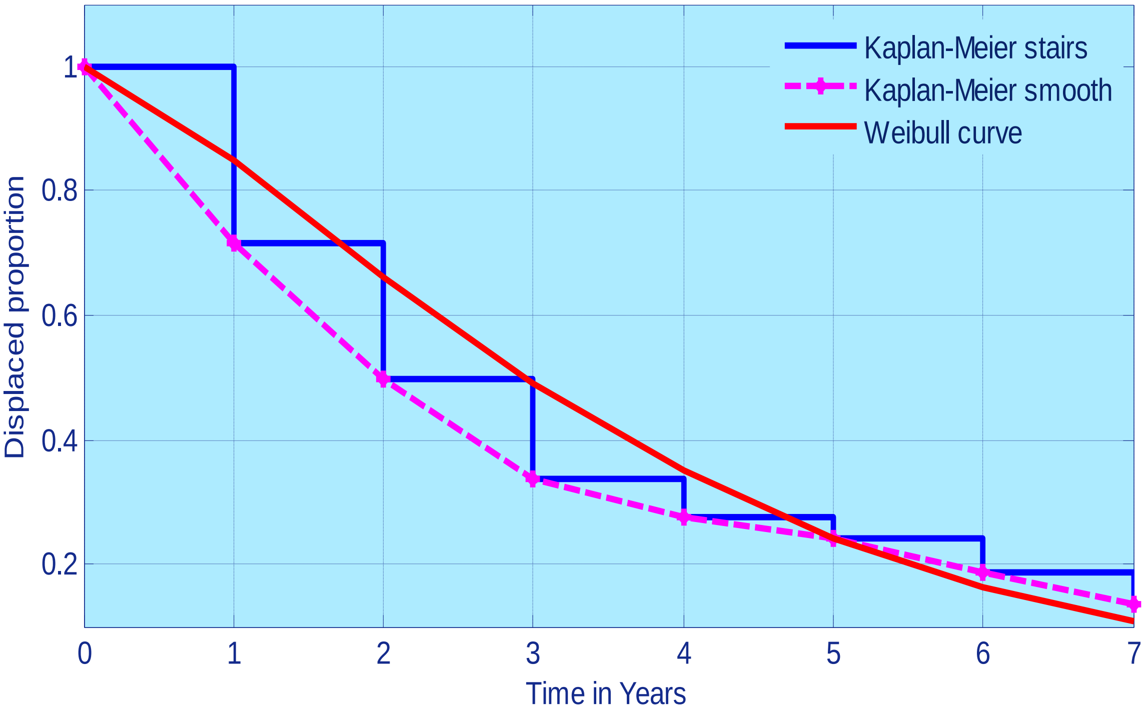

Figure 1. Weibull regression models on Kaplan-Meier plots

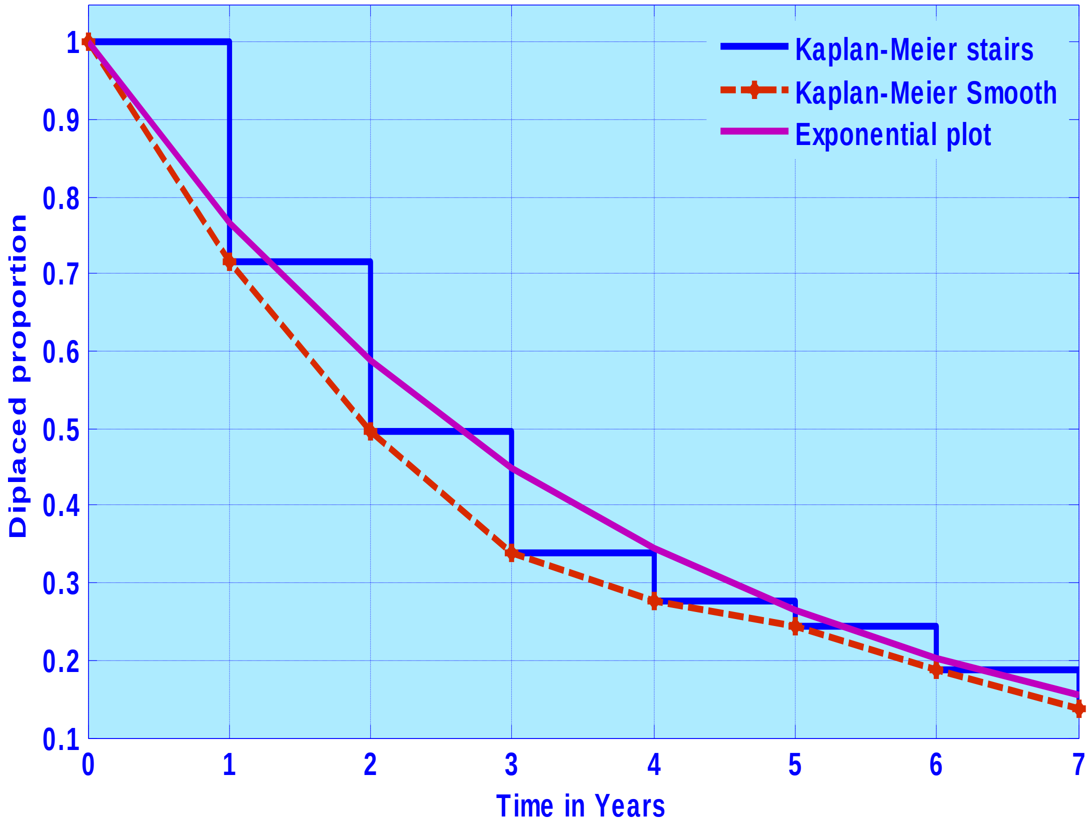

Figure 2. Exponential model imposed on Kaplan-Meier plots

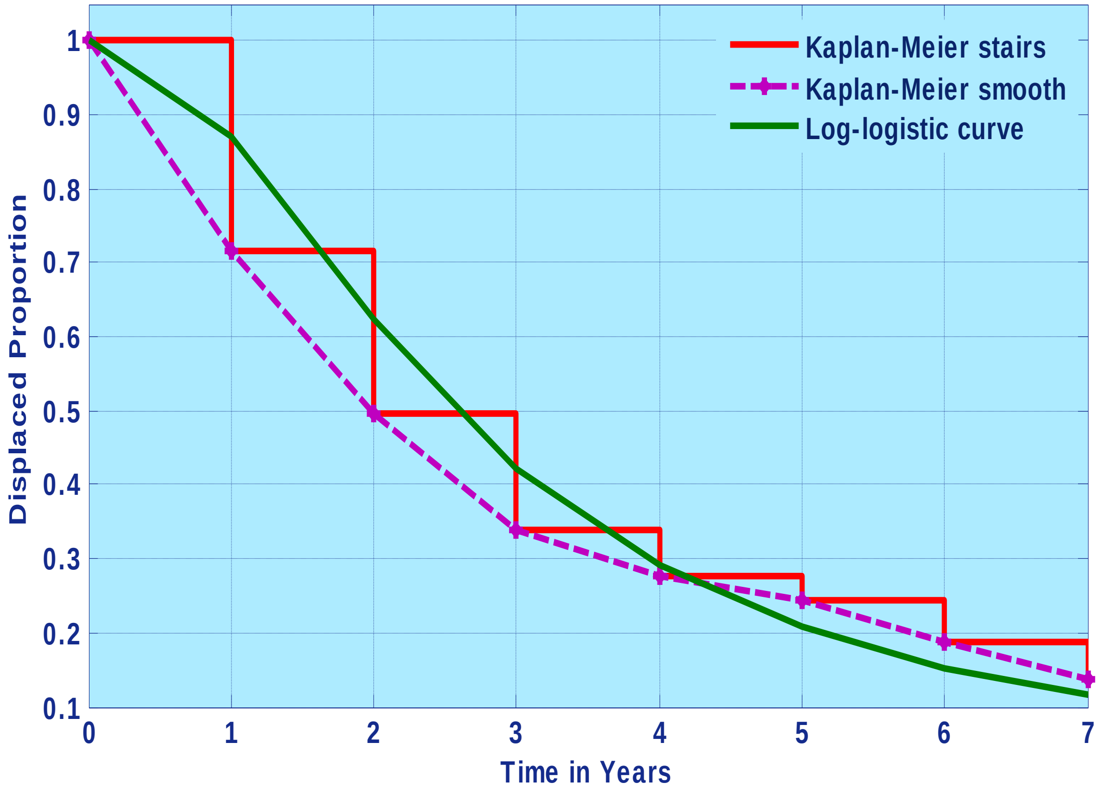

Figure 3. Log-logistic curve imposed on the Kaplan-Meier plots

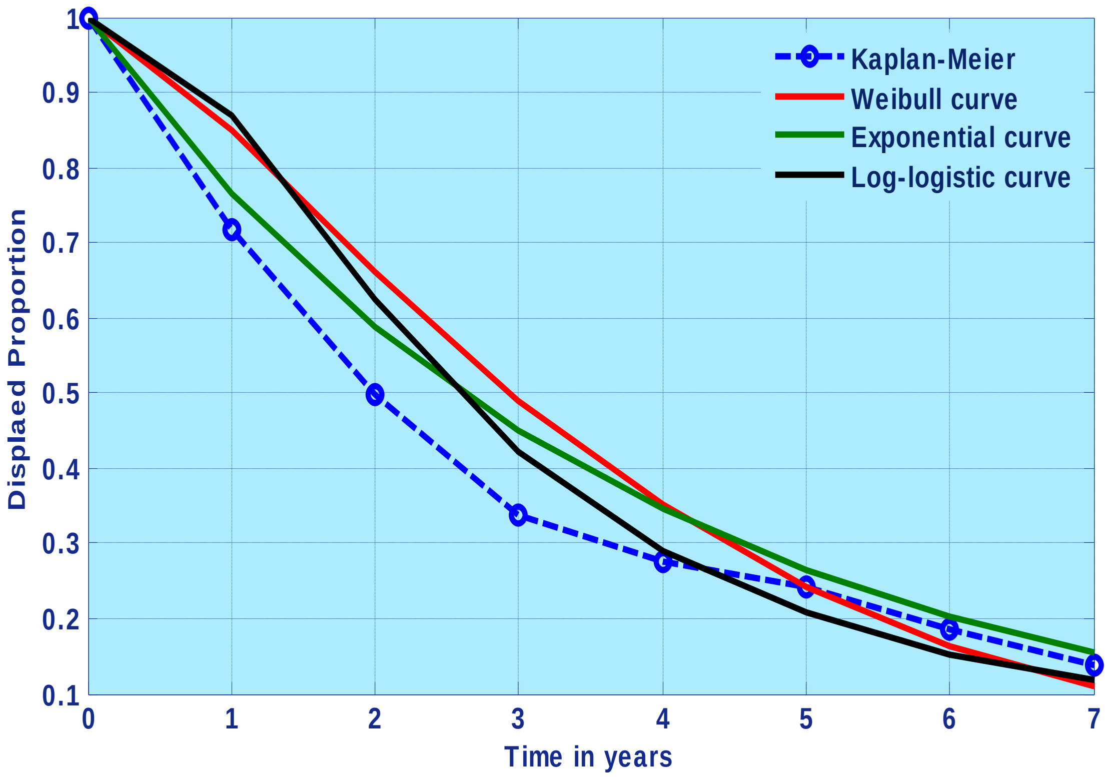

Figure 4. The three models' graphical comparision

This proportion constitute $4.2699 \times 10^{-3}\%$ of the study population which approximately zero. This implies that we can predict the year 2027 as the year when every displaced person would have returned to their ancestral homes. This is possible when using the Weibull model and assuming other factors constant.

Using the estimated parameter in table.2, then the displaced function of the exponential distribution model is defined by

$$D(t; 3.7591) = e^{-\left(\frac{t}{3.7591}\right)}$$

As $t \to 38, D(t; 3.7591) \to 4.072 \times 10^{-5}$

This proportion constitute $4.072 \times 10^{-3}\%$ of the study population which approximately zero. This implies that we can predict the year 2044 as the year when every displaced person would have returned to their ancestral homes. This is possible when using the exponential model and assuming other factors remain constant.

Using the estimated parameter in table.3, then the displaced function of the log-logistic distribution model is defined by

$$D(t; \gamma, \sigma) = D(t; 2.01766, 2.5718) = \frac{1}{1 + \left(\frac{t}{2.5718}\right)^{2.01766}}$$

This equation however has a weak convergence power and hence cannot be a good parametric prediction model for this data set. This affirms the result obtained by [18] when the same data was tested for the distribution fit of the data. It however provides good statistical inferences such as the mean return time, variance and median return time whenever required.

Fig. 1, 2, 3 and **4** show graphs of the displaced proportion of the three Accelerated failure time models imposed on the Kaplan-Meier curves of both stairs and smooth plot.

Fig. 1 shows the estimated displaced proportion of the Weibull model plotted together with the Kaplan-Meier stairs and smoothed curve. The Weibull model overestimated the displaced proportion for the first five years and thereafter underestimates. At five years, the estimate is slightly accurate.

Fig.2 shows the estimated displaced proportion of the exponential model plotted together with the Kaplan-Meier stairs and smoothed curve. The exponential model overestimated the displaced proportion throughout the study but the deviation decreases as with time.

Fig.3 shows the estimated displaced proportion of the log logistic model plotted together with the Kaplan-Meier stairs and smoothed curve. The log-logistic model overestimated the displaced proportion for the first four years and thereafter underestimates. At four years, the estimate is slightly accurate.

Fig. 4 shows the estimated displaced proportion of all the three models plotted together with the Kaplan-Meier smoothed curve. The first two and half years the estimated displaced proportion of the exponential model provides better estimates then the Weibull and the log-logistic models. Between the $2\frac{1}{2}$ and the fourth year, the estimated displaced proportion of the log-logistic model provides better estimates. Thereafter, the Weibull and exponential models became better in estimating the proportion of the displaced person. With time, the Weibull and the exponential models seem to provide better estimates of the displaced proportion and therefore we can use them to predict fairly the time when the displaced proportion is approximately zero.

4.3. Conclusions

The values of estimated parameters and their t-ratio are shown in **Table 1, 2** and **3**with the results indicating that the values of the parameter estimates for all the three models are statistically significant at 5% level of significance. The exponential and the Weibull regression models provide better estimates of the displaced proportion and can be used in the prediction of the displaced proportion at a given time. For the IDPs, it can be predicted by the Weibull and exponential models that by 2027 and 2044 respectively, 0.00% of the IDPs would be displaced when computed to four decimal points. The log-logistic model has a weak convergence rate and cannot be used to predict the maximum time that the IDPs take in displacement. This result affirms to that in [18] when the same data was tested for the distribution fit. The estimated shape parameter of a log-logistic model shows that the return rate increases then decreases. This reflects the huge return at the beginning when the camps were dissolved and slow returns for those who were displaced to relatives and friends and not to the camps who could return at will when there is sure peace. The estimated shape parameter of the Weibull model shows increasing return rate. This reflects the increased return after confirming sure peace i.e. the more there is sign of improved peace, the more IDPs return.

This result depicts the true picture of the IDPs return time to their ancestral homes and can be induced for the entire IDPs population. However, concerning the parametric model prediction of how long the entire IDPs take to return to their ancestral homes there are many factors that can be brought into play. Much as the predictions may hold for the 590 subjects, factors like education level of the family head, the family status (single mother, single father or complete family) may need consideration in future study for better inferences.

REFERENCES

[1] Abernathy, R. B. (1998). *The New Weibull Handbook.* 3rd ed. SAE Publications, Warren dale. PA.

[2] Allison, P.D. (1995). *Survival Analysis using SAS; A practical Guide* Cary, NC: SAS Institute.

[3] Buis, M.L., (2006), *An introduction to Survival Analysis,* Department of Social research Methodology, Vrije Universiteit Amsterdam.

[4] Collett D. (2003), *Modelling survival data in medical research*, second edition, Chapman and Hall/CRC.

[5] Cox, D.R. and Oakes, D. (1984*). Analysis of Survival Data*. London: Chapman and Hall, New York.

[6] Crow, L.H. (1982), "Confidence Interval Procedures for the Weibull Process With Applications to Reliability Growth," *Technometrics,* 24(1):67-72.

[7] Crowder, M.J., Kimber, A.C., Smith, R.L. and Sweeting, T.J. (1991). *Statistical Analysis of Reliability Data.* Chapman Hall, London, U.K.

[8] Ibrahim J.G., Chen M.H., and Sinha D. (2005). *Bayesian Survival Analysis*. Springer series in statistics ISBN 978-1-4757-3447-8.

[9] Gross A.J. and Clark V.A. (1975). *Survival Distribution: Reliability Applications in the Biomedical Sciences* Wiley.

[10] Khan K.H, Saleem M and Mahmud. Z. (2011). Survival Proportions of CABG Patients: A New Approch. Volume 3, Number 3.

[11] Klein. P.J and Moeschberger. L.M (1997, 2003). *Survival Analysis Techniques for Censored and Truncated Data*.

[12] Kleinbaum, D.G. and Klein, M. (2005*). Survival analysis: a self-learning text*, Second Edition, Springer-Verlag Publishers, New York, Chapters 4–7, 11.

[13] Lawless, J.F. (2003), *Statistical models and methods for lifetime data*, second edition, Wiley-Inter-Science, A John Wiley & Sons, Inc., Publication, Hoboken, New Jersey.

[14] Leemis, L.M., (1995). Reliability Probabilistic Model and Statistical Methods.

[15] Lee E.T and Wang J.W. (2003), *Statistical methods for survival data analysis,* 3rd edition, John Wiley & sons, Inc., Hoboken, New Jersey.

[16] Mann, N.R. (1984). Statistical estimation of the parameters of the Weibull and Frechet distributions*. In statistical extremes and applications, (J.Tiago de Oliveira, ed.)* 81-89 Dordrecht: Reidel.

[17] Nelson, W. (1982). *Applied Life Data Analysis*. Newyork: Wiley.

[18] Okello, J.O. Abdou KA, D. and Odongo, L.O., (2014), Modelling Internally Displaced Persons'(IDPs) time to resuming their ancestral homes after IDPs' camps in Northern Uganda using parametric methods, *international Journal of Science and Research,* 3(5), 125-131.

[19] Saleem M., Mahmud Z., and Khan K. H. (2012) Survival Analysis of CABG Patients by Parametric Estimations In Modifiable Risk Factors - Hypertension and Diabetes. *American Journal of Mathematics and Statistics*, 2(5): 120-128.

[20] Selvin, S. (2008). *Survival analysis for epidemiologic and medical research*. A practical guide, Cambridge University Press, New York, USA.

[21] Smith, L.R. and Naylor, J.C., (1987), A comparison of Maximum Likelihood and Bayesian Estimators for a three-parameter Weibull distribution function. *Appl. Statist*. 1987, 36 (3), 358-369.