UNIVERSITY OF KWAZULU-NATAL

UNIVERSITY OF
KWAZULU-NATAL

INYUVESI
YAKWAZULU-NATALI

Masters Thesis

APPLICATION OF SURVIVAL ANALYSIS METHODS TO STUDY UNDER-FIVE CHILD MORTALITY IN UGANDA

Author:
Miss. Nasejje Justine
Signature:.....
Date: .....

Supervisor:
Prof. Mwambi Henry
Signature:.....
Date: .....

Co-supervisor:

Dr. Thomas Achia

Signature:.....

Date: .....

A thesis submitted in full fulfilment of the requirements for the Masters in StatisticsUniversity of KwaZulu–Natal

in the

School of Mathematics, Statistics and Computer Science.

November 2013

UNIVERSITY OF KWAZULU-NATAL

Abstract

College of Agriculture, Engineering and Science
School of Mathematics, Statistics and Computer Science

Masters in Statistics

APPLICATION OF SURVIVAL ANALYSIS METHODS TO STUDY UNDER-FIVE CHILD MORTALITY IN UGANDA

by Miss. Nasejje Justine

Infant and child mortality rates are one of the health indicators in a given community or country. It is the fourth millennium development goal that by 2015, all the united nations member countries are expected to have reduced their infant and child mortality rates by two-thirds. Uganda is one of those countries in sub-Saharan Africa with high infant and child mortality rates and therefore has the need to find out the factors strongly associated to these high rates in order to provide alternative or maintain the existing interventions. The Uganda Demographic Health Survey (UDHS) funded by USAID, UNFPA, UNICEF, Irish Aid and the United kingdom government provides a data set which is rich in information. This information has attracted many researchers and some of it can be used to help Uganda monitor her infant and child mortality rates to achieve the fourth millennium goal. Survival analysis techniques and frailty modelling is a well developed statistical tool in analysing time to event data. These methods were adopted in this thesis to examine factors affecting under-five child mortality in Uganda using the UDHS data for 2011 using R and STATA software. Results obtained by fitting the Cox-proportional hazard model and frailty models and drawing inference using both the Frequentists and Bayesian approach showed that, Demographic factors (sex of the household head,sex of the child and number of births in the past one year) are strongly associated with high under-five child mortality rates. Heterogeneity or unobserved covariates were found to be significant at household but insignificant at community level.

Key: Cox-proportional hazard model, frailty models, Frequentists and Bayesian inference.

Acknowledgements

I thank God for giving me such a great opportunity and for giving me life to come thus far. I greatly appreciate my supervisors, Professor Mwambi Henry and Dr. Thomas Achia, for their constructive suggestions and comments on my research work. Their selfless support during my period of study at the University of Kwazulu-Natal is highly appreciated.

I thank my colleagues and friends Miss. Nancy Odhiambo, Mr. Mwangi Isaac, Mr. Idrisu Abdul Karim, Mr. Malinzi Joseph and Mr. Isaac Bisaso for their support and encouragement throughout this study period.

I am grateful for the spiritual and moral support from my mother Miss. Nabunya Ruth and my uncle Rev. Nathanael Mutyaba.

I am forever grateful to the University of Kwazulu-Natal and the African Institute for mathematical Sciences for the funding.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Infant and child mortality rates are important indicators of societal and national development as they have been described as key markers of health equity and access. Bryce et al. (2006), Ssewanyana and Younger (2008) observed that reducing infant and child mortality is one of the Millennium development goals and in fact it is the fourth millennium development goal (MDG4) which states that infant and child mortality rates are to be reduced by two-thirds between 1990-2015. Kyaddondo (2012) described the under-five child mortality rate in Sub-Saharan Africa as a high mortality rate and that of Southern Asia as moderate. The above mentioned concerns are some of the reasons why infant and child mortality has attracted many researchers in order to identify the factors strongly associated with high under-five child mortality rate and to evaluate the various government interventions.

1.1.1 Background of the study area

Uganda is one of the countries in East Africa, it is one of the Sub-Saharan African countries categorised as a low income or a developing country (World Bank Group, 2010). It is a landlocked country on an area of 236,580 Square Kilometres. The country has a total population of approximately 36.35 million people and it is expected to reach 103.2 million people in 2050 (Kyaddondo, 2012, World Bank Development Data Group, 2012) . Uganda is bordered by five countries which are, Southern Sudan to the North, Democratic republic of Congo to the West, Tanzania to the South, Rwanda to the South-west and Kenya to the East.

FIGURE 1.1: The map of the Republic of Uganda with its neighbours (Reinikka and Collier, 2001).

Okwero et al. (2010) described Uganda's economy as one of the fastest growing economies in the world. He indicated that Uganda's gross domestic product grew by an impressive 6.5% in 2007 and that Uganda's overall growth performance over the past twenty years has been among the best in Africa as well as in the whole world. Despite this growth Uganda's per capita income of US$320 in 2007, the country is still placed among the poor countries of the world.

Ssewanyana and Younger (2008) described Uganda's economy as one that has been rapidly growing and sustained for an extended period of time but the rate of increase in the incomes of the people did not match the rate of improvement in the health indicators such as infant and child health. He examined the factors affecting child mortality and also examined the likelihood that Uganda will meet the millennium development goal (MDG4) by 2015.

The results of his study (Ssewanyana and Younger, 2008) suggested that in order to achieve the fourth millennium development goal, there is need for further studies to asses the effects of contextual determinants of child survival in Uganda. Nuwaha et al. (2011a) by using the DHS data collected within the period of 1995-2000 did show that under-five mortality rate in Uganda increased from 147.3 to 151.5 deaths per 1000 live births and indicated that the reasons for the increase were not clear.

Ayiko et al. (2009) used the data collected between 1990 and 2006 from the Uganda demographic health surveys (UDHS) to assess the determinants of under-five child mortality. The results indicated that under-five child mortality remained unchanged in the period of 1991-1995 and 1996-2000 and it later declined in the period of 2001 -2005. He pointed out factors like sex of the child (males were at a higher risk of death), place of residence (northern Uganda recorded highest mortality rate), birth interval (Less than 24 months was linked to a high under-five child mortality) and mother's education (no primary education were associated with a higher risk of a child dying before the age of five). He therefore concluded that Uganda was not on track to meet the MGD4 by 2015 and he therefore suggested further studies to asses the effects of the contextual determinants of child survival in Uganda.

Croke (2012) identified political economic factors that help to explain the dramatic differences in the pace of child mortality reduction in Uganda and Tanzania from 1995-2007. He realised that in Uganda there was a negative shock to the health system driven by the President's decision to eliminate presidential term limits in 2001-2006. A presidential candidate was not only restricted to two terms but can go as many terms as he wants as long as he is re-elected by the citizens. This process reversed the previous health sector institution gains and had particularly negative effects on child health service delivery in Uganda over the 2001-2006 period . His results indicate that Uganda had an under-five mortality rate of 147 per 1000 live births in 1995 and Tanzania recorded a rate of 137 per 1000 live births. After a period of ten years, the Tanzania's under-five mortality rate had declined by 34-35% to 91 per 1000 live births and that of Uganda declined by 12-15% to 137 per 100 live births which was even more less than the average decline for sub-Saharan Africa over the same period which was 18% during that period.

1.2 Literature review

The past two decades have seen an increase in articles on the factors associated with infant and child survival. This is because infant and child mortality rate is considered to be one of the key health indicators in an economy.

A paper by Mosley et al. (1984), *An analytical framework for the study of child survival in developing countries* proposed an analytical framework for the study of determinants of child survival in developing countries. The study incorporates both Social and Biological variables and integrates research methods used by social and medical scientists to study child survival. The frame work is based on the premise that all social and economic determinants of child mortality necessarily operate through a common set of biological mechanism or proximate determinants that exert an impact on mortality.

Bailey (1988) examined the factors determining infant mortality using data from fertility and family planning survey in Uganda sponsored by the international research centre in Canada. The results of the analysis indicated that Background factors (current place of residence of the mother, religion and mother's tribe), Demographic factors (mother's age, age at first marriage, number of children ever born and duration of breast feeding), Social-economic factors (maternal and paternal education and occupation, household income) are the key determinants of infant and child mortality. Many other studies on the same research topic, in different countries, using survey data are in agreement with these results (Limin, 2003, Mondai et al., 2008, Nuwaha et al., 2011b, Uddin et al., 2009).

A study by Hobcraft et al. (1984) identified five main social-economic factors that influence infant and child survival and these include mother's education, mother's work status, husband's occupation, husband's education and type of place of residence. They used a simple tabular analysis followed by a multivariate approach in order to asses the relative importance of each of the five variables. The study used the World Fertility Survey data that was based on enquiries from 28 developing countries. In Asian countries, mother's level of education was seen to be strongly associated with mortality of the child during the first five years of their life. In America, results indicated that the husband's education was most important and in a few African countries, infant mortality was relatively strongly associated to husband's occupation and education.

Rutstein et al. (2000) showed that internally comparable data derived from survey programmes like the Demographic and Health Survey (DHS) have identified some of the key factors that help to explain child and infant mortality trends. The factors that have been repeatedly given in the DHS programme are and fall into five broad categories which include;

- Fertility behaviour like spacing births;

- The use of health services by mothers and for her children;
- Environmental health conditions like outbreak of diseases, hygiene, among others;
- Nutritional status, breastfeeding, and infant feeding;
- Social economic status.

The study also showed that there has been a remarkable decrease in mortality among infants and children in most developing countries from 1980-1990's but that the decline has reversed in some sub-Saharan African countries over the period.

Ssengonzi and Shannon (2002) examined the effect of female migration on the health and survival of the most vulnerable migrants (infants and children) in Uganda. He used the Loglogistic regression techniques to analyse the probability of a child surviving up to the age of five. Results showed that 10% of the children die before age five and within group difference in mortality exists in urban and rural children depending on their mother's migration status. Other variables like parents education, household size, household hardship, mother's age at first birth, duration of breast feeding and place of delivery were seen to be significant.

Other studies on infant and child mortality in Uganda include one by Nuwaha et al. (2011b) which focused on understanding the social-economic determinants of child mortality. The study was a retrospective study and was based on the Teso region in Eastern Uganda. For this region, it was found out that in the period of the 1959-1969, infant mortality was 94 per 1000 live births compared to the 120 per 1000 live births for the whole country by that time. The data used was from the 1959 and 1969 census (Protectorate, 1961, Uganda Ministry of Planning and Economic Development Statistics Division, 1976). They concluded that the high ownership of cattle and growing of high protein and energy foods for domestic consumption might have been responsible for better childhood survival in the Teso region compared to Uganda as a whole.

Results from the 2006 DHS survey indicated that the primary factors affecting infant mortality include; gender of the child, mother's age at first birth, birth order, and birth interval. We recommend Uganda Bureau of Statistics Kampala (2007) report for more details.

From a statistical and data analysis point of view, different models have been used by different authors to study child survival. Ssengonzi and Shannon (2002) used a Logistic regression techniques on the 1995 Ugandan Demographic Health Survey (UDHS) data set to determine the relationship between child survival and migration status. Some authors choose to use Loglinear models to do the analysis for instance Curtis and McDONALD (1991) used it to determine the effect of birth spacing on infant mortality in Brazil.

The Logistic model is the most popularly used model because it assumes that child survival is a binary response (child is dead or alive, Kazembe et al. (2012)). All the above mentioned models ignore time to event and therefore fail to include the exposure to the risk of the event overtime. Other models like the Cox proportional hazard model by Cox (1972), are widely used to deal with time to event data and their relevancy on research in survival analysis in demography and related fields has increased over the years. The Cox model has several advantages and some of them are; (a) ability to include analysis of censored and truncated data (b) ability to include analysis of time varying covariate effects and lastly (c) the extensions of the Cox regression models with the inclusion of random effects and flexible modelling through semi-parametric and non-parametric approach. There is an advantage of these models over the ordinary generalised linear models as demonstrated by Kazembe et al. (2012). The random effects allows for the modelling of the unobserved covarieties (and inherent heterogeneity) or frailty. These factors may be at a family, district, community, regional or national level and these can not be ignored because they have an effect on the outcome.

Kazembe et al. (2012) and Omariba et al. (2007) used the Weibull unobserved heterogeneity (frailty) survival model on the 1998 Kenya DHS data to analyse the determinants of infant and child mortality in Kenya. They compared the results of the standard Weibull survival model to the frailty Weibull model. They also mentioned that non-frailty models are biased due to the violation of the statistical assumption of independence.

1.3 Basic survival analysis concepts

Survival analysis is a statistical method or tool used to analyse time to events data. Survival analysis is a highly active area of research with application in many fields of

study which include engineering, physical, biological and social sciences (Cleves et al., 2008, Klein and Goel, 1992). An event is an outcome on an individual unit that is of scientific interest in different studies like sociology, biology, demography, medicine, employment among other fields. The most common event of interest in the earlier development of research in this area was death, this therefore suggests the name of the research area Survival analysis.

Other events include, disease diagnosis, wedding, falling in love, infection of disease, graduation from the university among others.

The main focus of this thesis is the classical survival analysis which focuses on time to a single event for each individual unit in the study. Examples that fall under time to event data include;

- Time from disease diagnosis to death
- Time from enrolment of a graduate degree to graduation
- Time of release from prison to re-arrest.
- Time from marriage to divorce.

As earlier on explained, the original event of interest in this research area was death and this led to the name Survival analysis and therefore time to any other event which is not death is also called Survival time (Aalen and Gjessing, 2008a).

Survival analysis is special from any other statistical methods of analysis because survival data is always incomplete. This arises due to the fact that the person doing the study has to wait until the event of interest occurs but since the study period has got a time limit (end of the study), some of the individual units may not have experienced the event within the time frame and therefore at the end of the study one will face a challenge of an incomplete dataset which makes the analysis of such data using other standard statistical methods and the ordinary linear regression methods impossible (Aalen and Gjessing, 2008a). Other causes of an incomplete data set include; losses to follow up and dropouts from the study. These result into what we call censored data points.

There are three forms of censoring and these are;

- Right censoring
- Left censoring
- Interval censoring

With censored data we cannot use the ordinary statistical methods to analyse it because we can not even calculate the simple mean with such data, we cannot find the standard deviation or even perform a regression analysis. In a essence with censored data we are faced with partially observed distribution of the time to event random variable.

1.3.1 The survival and the hazard functions

There are two basic functions that are very important in the whole theory of survival analysis. These are the survival and hazard function.

1.3.2 The survival function

Given a random variable T that denotes the survival time, the survival function denoted as $S(t)$ is defined as:

$$S(t) = P(T > t) = 1 - F(t) = 1 - \int_0^t f(u)\, du, \tag{1.1}$$

where $f(t)$ and $F(t)$ are the probability density and the cumulative density functions respectively of a given distribution. The expression in (1.1) is the probability of surviving beyond time t. Note that $S(0) = 1$, $S(t) \to 0$ as $t \to \infty$. It is a downward sloping curve and can be estimated by using the Kaplan-Meier method (to be discussed in the later section).

1.3.3 The Hazard function

Given a set containing individuals who are at a risk of experiencing a certain event denoted by $R(t)$ (risk set) or individuals who have not yet experienced the event by time t, the probability of an individual in the risk set experiencing the event in the small time interval $[t, t + \Delta t)$ is defined as $h(t)\Delta t$. Therefore the hazard rate is defined as:

$$h(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t \leq T < t + \Delta t | T \geq t). \tag{1.2}$$

Unlike the survival function which is a downward sloping curve for any type of survival data given, the hazard function takes on any shape of a non negative function and it varies depending on the type of survival data given (will be discussed in details in the later chapter). The hazard function can be alternatively represented in terms of the cumulative hazard function $H(t)$.

$$H(t) = \int_0^t h(u)\, du. \tag{1.3}$$

The name cumulative is due to the fact that the function is the accumulation of the hazard over time.

According to Aalen and Gjessing (2008b), the cumulative hazard rate can be estimated using the Nelson-Aalen estimate and the increments of the Nelson-Aalen estimate can be smoothed to provide an estimate to the hazard rate.

1.3.4 The relationship between the hazard and the survival functions

From equation (1.2),

$$h(t) = H'(t) = \frac{1}{S(t)} \lim_{\Delta t \to 0} \frac{S(t) - S(t + \Delta t)}{\Delta t} = -\frac{S'(t)}{S(t)},$$

$$H(t) = -\frac{S'(t)}{S(t)} = -\ln[S(t)], \tag{1.4}$$

where $H(t) = \int_0^t h(u)\, du$ and $H'(t)$ denotes the first derivative of the cumulative hazard function.

1.3.5 Non-parametric methods

Survival data are summarized through estimates of the hazard and survival function (Le and Le, 1997, Miller Jr, 2011). The methods used to estimate these functions are known as Non-parametric or distribution free methods.

The aim of non-parametric estimation of the survival function is to come up with graphical summaries of the survival times for a given group of individuals considered in the study. These graphical summaries are for the hazard and the survival function. After estimating the survival function, the median and other percentiles can be obtained which help to give a more detailed analysis (Cleves et al., 2008, Hanagal, 2011).

Given survival curves of two groups, we can as well compare the two groups of individuals by plotting them on the same graph and then observe if there exists a difference between the curves (whether one group has a higher chance of survival than the other group). However, this is an informal procedure because sometimes the difference between the two groups may exist but not significant. A more formal procedure is by using the Logrank test which is the most powerful test against the alternatives that the hazard functions are proportional (Fleming and Harrington, 2011, Hanagal, 2011).

1.3.6 The Empirical survival function

Assuming that in the given sample of survival data none of the data points is censored and that also there exists no tied observations. The survival function denoted as $S(t)$ which is the probability that an individual survives beyond time t can be estimated by using the empirical survival function. The empirical function which is the estimate to the survival function in absence of censored data denoted as $\hat{S}(t)$ is given by:

$$\hat{S}(t) = \frac{\text{Number of individuals with survival time } \geq t}{\text{Number of individuals in the data set}}, \tag{1.5}$$

$\hat{S}(t) = 1$ for all values of t before the first failure and $\hat{S}(t) = 0$ after the final failure or occurrence of event.

The estimated survival function ($\hat{S}(t)$) is observed to be constant between two adjacent times and therefore its plot turns out to be a step function, this function decreases immediately after each observed event time (Collett, 2003, Hosmer Jr et al., 2011).

1.3.7 The Kaplan-Meier estimator

The Kaplan-Meier estimator also known as the product limit estimator was presented by Kaplan and Meier (1958). It gives a simple and quick estimate of the survival function in the presence of censoring. It uses the exact failure time (Collett, 2003, Hanagal, 2011).

The standard error is therefore given by:

$$s.e\left(\hat{S}(t)\right) = \hat{s}(t)\left[\sum_{i=1}^{k}\frac{d_i}{n_i(n_i-d_i)}\right]^{\frac{1}{2}}.$$

The cumulative hazard function

From equation (1.4), if $\hat{S}(t)$ is the Kaplan-meier estimate to the survival function, then:

$$\hat{H}(t) = -\sum_{i=1}^{k}\ln\left(1-\frac{d_i}{n_i}\right),$$

is an estimate to the cumulative hazard function.

From Taylor series expansion:

$$\ln\left(1-\frac{d_i}{n_i}\right) = -\frac{d_i}{n_i} - \left[\frac{d_i}{n_i}\right]^2 + \ldots \approx -\frac{d_i}{n_i},$$

by ignoring higher order terms. The estimate to the cumulative hazard function is therefore given as:

$$\hat{H}(t) = \sum_{i=1}^{k}\frac{d_i}{n_i}.$$

1.4 Study objectives

The broad objectives of the thesis is to review survival analysis techniques with frailty models and apply these methods on Uganda Demographic Health Survey data for the year 2011 to examine factors affecting under-five child mortality in Uganda and to also examine whether unobserved heterogeneity or unobserved covariates help to explain the under-five children mortality in Uganda using both the Frequentist and the Bayesian approach.

The specific objectives of this study is to apply survival analysis techniques including frailty models on the 2011 Demographic Health Survey data for Uganda;

- To identify the factors responsible for the under-five child mortality in Uganda
- To examine the effects of unobserved covariates (frailty) on under-five mortality both at family and community level.

Chapter 2

Important statistical distributions for a time to event random variable

The parametric methods used in survival analysis are based on distribution functions which allow for only positive random variables. This is because time to an event T is a positive random variable.

2.1 Parametric methods

There are quite a number of parametric models for survival data. The choice of the model is sometimes based on the physics of the failure mode (Hanagal, 2011, Hosmer Jr et al., 2011, Le and Le, 1997) but it is sometimes due to the fact that the model fits the given survival data. With parametric models one has to make a choice of the distribution to assume for the given dataset, the following distributions are some of the distributions that are most often used. These are;

- Exponential distribution;
- Weibull distribution;
- Log-normal distribution;
- Gamma distribution;
- Log-logistic distribution.

Note: All the distributions assumed for survival data have one special characteristic, that they take on only positive random variables. This is because survival times are always positive ($T \geq 0$).

2.1.1 Exponential distribution

The exponential model is the most simplest of all life distribution models with only one parameter. The exponential model has got a constant hazard rate. The probability that an individual will die within a small interval of time Δt given that the person has survived up to time t is constant for any time period.

The probability density function of an exponential distribution is given as:

$$f(t) = \lambda \exp(-\lambda t), \quad t > 0 \quad \lambda > 0. \tag{2.1}$$

The cumulative distribution function denoted as $F(t)$, the probability that an individual has survived up to time t is given as:

$$F(t) = 1 - \exp(-\lambda t), \quad 0 \leq t < \infty.$$

The survival function $S(t)$, the probability that an individual survives beyond time t can be derived from $F(t)$ as follows:

$$S(t) = 1 - F(t) = \exp(-\lambda t). \tag{2.2}$$

The hazard function denoted as $h(t)$ is given as:

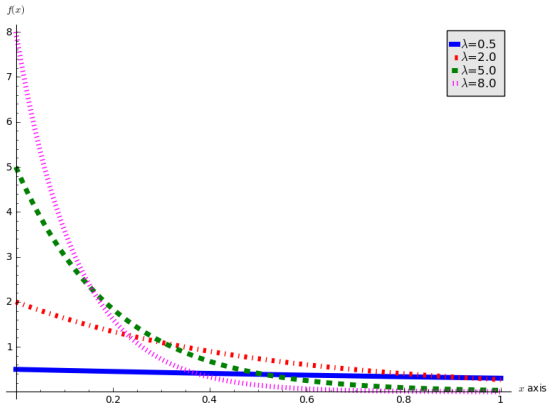
$$h(t) = \frac{f(t)}{S(t)} = \lambda. \tag{2.3}$$

And the cumulative hazard function is given as below:

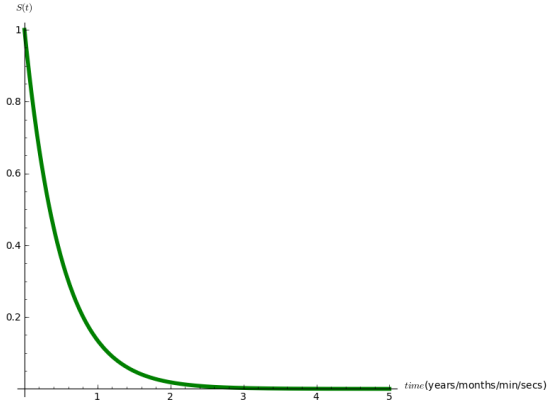
$$H(t) = \lambda t.$$

Note: The exponential distribution has a most important property called the loss of memory property because it has a constant hazard rate. The exponential model is

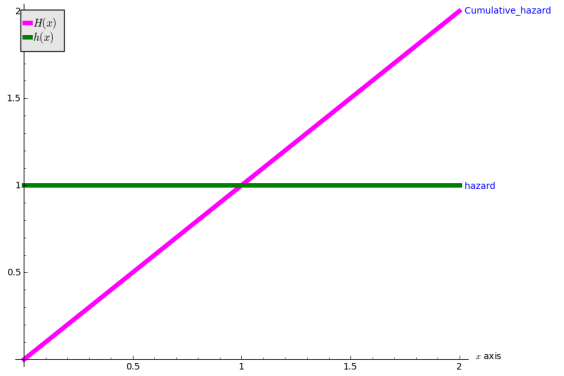
most suitable in modelling lifetime data of engineering applications. For example the probability that a light bulb will blow today given that it did not blow yesterday is the same as the probability that the bulb will blow tomorrow given that it did not blow today therefore in such a problem, the hazard function can be modelled using the exponential hazard. The exponential model is unrealistic when it comes to lifetime data which involves ageing which therefore calls for a more general model.



(A) The exponential density function for different values of the scale parameter.



(B) Survival Function.



(C) The exponential hazard and its cumulative hazard function.

FIGURE 2.1: The probability density, the survival function and the cumulative hazard functions of the Exponential distribution.

2.1.2 Weibull distribution

The exponential distribution is not applicable in some of lifetime data like Biological and social process because it assumes a constant hazard rate. The Weibull distribution is the most flexible and more general distribution, it has got two parameters; one of the parameters is called the scale parameter and the other the shape parameter.

The probability density function of the Weibull distribution is formulated in a number of similar forms but one of the commonly used version is given by:

$$f(t) = \frac{\alpha}{t}\left[\frac{t}{\lambda}\right]^{\alpha} \exp\left(-\left[\frac{t}{\lambda}\right]^{\alpha}\right), \quad 0 < t < \infty, \lambda, \alpha \ > 0. \tag{2.4}$$

where λ and α are the scale and shape parameters respectively.

The Weibull distribution is a more general distribution compared to the exponential distribution, because when $\alpha = 1$, the Weibull distribution simplifies to an Exponential distribution.

The cumulative density function denoted as $F(t)$ is given as:

$$F(t) = 1 - \exp\left(-\left[\frac{t}{\lambda}\right]^{\alpha}\right).$$

The survival function $S(t)$ is given as:

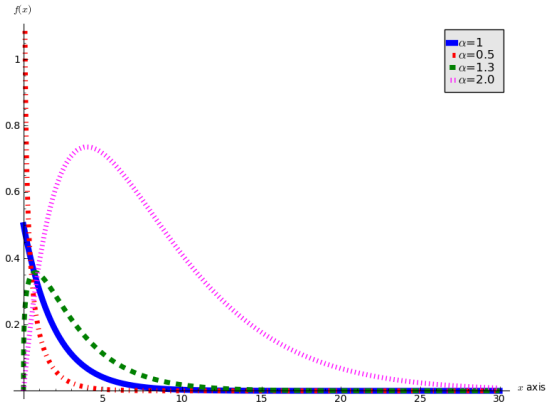
$$S(t) = \exp\left(-\left[\frac{t}{\lambda}\right]^{\alpha}\right). \tag{2.5}$$

It follows that the hazard rate is given by:

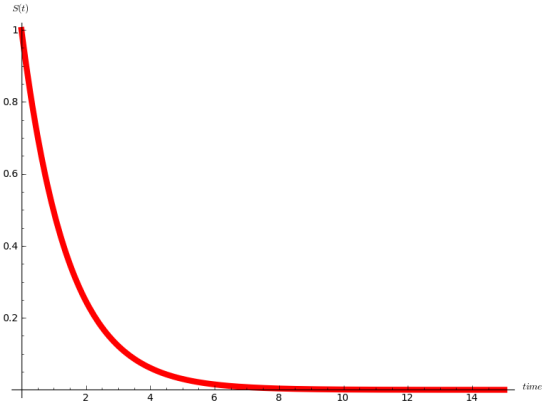
$$h(t) = \frac{\alpha}{\lambda}\left[\frac{t}{\lambda}\right]^{\alpha-1}. \tag{2.6}$$

The cumulative hazard function for the Weibull hazard distribution is given by:

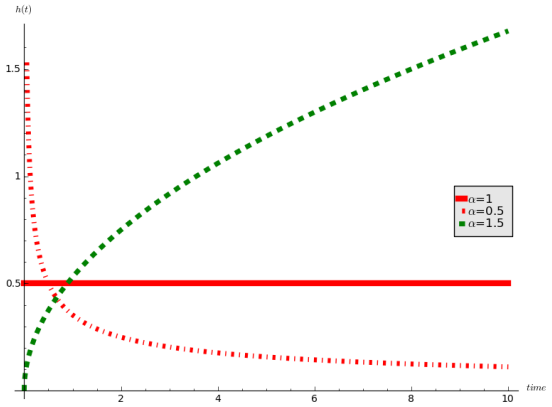
$$H(t) = \int_0^t h(u)\,du = \left[\frac{t}{\lambda}\right]^{\alpha}.$$



(A) The probability density function of the Weibull distribution for different values of the shape and scale parameter.



(B) The Weibull survival function.



(C) The shapes of the Weibull hazard functions for some selected values of the shape parameter.

FIGURE 2.2: The probability density, the Survival function and the Hazard functions of the Weibull distribution.

The Weibull model can fit a range of survival data because of its flexibility. When $\alpha = 1$ the shape parameter is set to one, the failure rate is a constant.

When $\alpha > 1$, the failure rate is increasing and when $\alpha < 1$, we have a decreasing failure rate.

Thus the Weibull probability model can be used to model survival times of populations whose hazard rate is assumed to be decreasing, increasing or constant.

2.1.3 Log-normal distribution

According to Hanagal (2011), the Log-normal distribution is a very flexible distribution just like the Weibull distribution therefore it can fit many types of failure data. The two parameter Log-normal distribution with scale parameter σ and shape parameter τ_m

(median time) has a probability density function give as below:

$$f(t) = \frac{1}{\sigma t\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\left[\ln t - \ln \tau_m\right]\right), \quad 0 < t < \infty. \tag{2.7}$$

The survival function is given by:

$$S(t) = 1 - F(t) = 1 - \Phi\left[\frac{\ln t - \ln \tau_m}{\sigma}\right],$$

where $F(t) = \int_0^t \frac{1}{\sigma x\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\left[\ln x - \ln \tau_m\right]\right) dx = \Phi\left[\frac{\ln t - \ln \tau_m}{\sigma}\right]$ and $\Phi(z)$ denotes the standard normal cumulative distribution function.

The hazard rate denoted as $h(t)$ is derived as follows:

$$h(t) = \frac{f(t)}{S(t)} = \frac{\left[\frac{1}{t\sigma\sqrt{2\pi}} \exp\left(-\frac{[\ln t - \ln \tau_m]^2}{2\sigma^2}\right)\right]}{1 - \Phi\left[\frac{\ln t - \ln \tau_m}{\sigma}\right]}.$$

The three parameter form of the Log-normal distribution is more general because it has got a third parameter θ known as the shift or location parameter.

Given that the failure times follow a Log-normal distribution, the natural log of the failure times have a normal distribution with mean $\ln \tau_m$ and standard deviation σ. This fact makes it easy to work with data assumed to follow a Log-normal distribution because you just have to take the logarithm of the failure and censor times and then analyse the data as normally distributed data. It is a flexible distribution and because of this property it can easily be transformed into a normal distribution. This makes it easy to work with this distribution mathematically and also due to the presence of many good software analysis programmes to deal with normal data.

2.1.4 Log-logistic distribution

The survival time T has a Log-logistic distribution if $\ln T$ has a Logistic distribution. The probability density function of a Log-logistic distribution is given by:

$$f(t) = \frac{\alpha\lambda\left[\alpha t\right]^{\lambda-1}}{\left[1 + \left[\alpha t\right]^{\lambda}\right]^2}, \quad \alpha > 0, \lambda > 0 \quad \text{and} \quad t > 0. \tag{2.8}$$

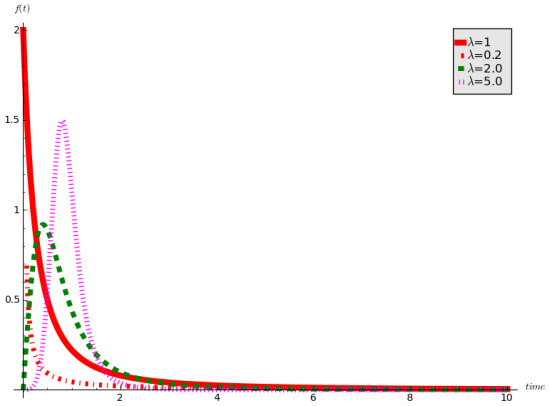
The survival function is given as:

$$S(t) = \frac{1}{1 + [\alpha t]^{\lambda}}$$

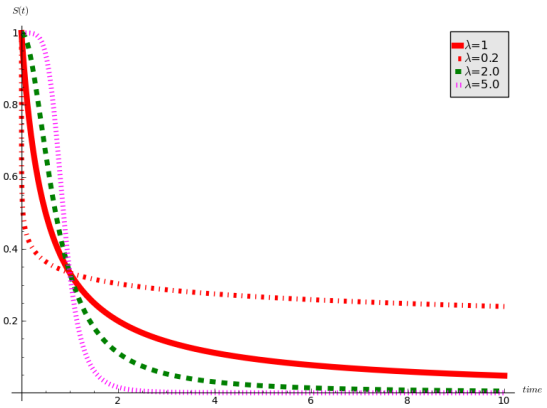
and the hazard function is given as below:

$$h(t) = \frac{\alpha\lambda\,[\alpha t]^{\lambda-1}}{1 + [\alpha t]^{\lambda}}.$$

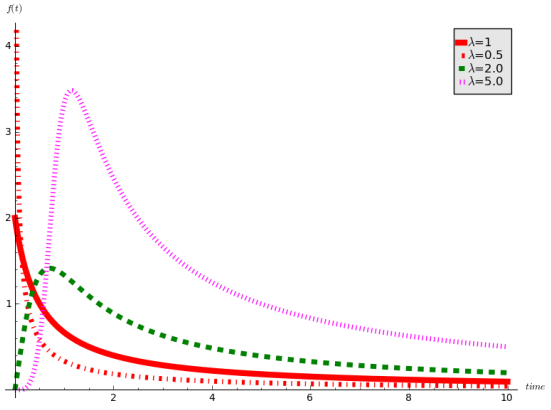
When $\lambda > 1$ (the scale parameter), the Log-logistic hazard first increases and it becomes a maximum at $t = \frac{[\lambda-1]^{\frac{1}{\lambda}}}{\alpha^{\frac{1}{\lambda}}}$ and it then declines, when $\lambda = 1$, the hazard starts at $\alpha^{\frac{1}{\lambda}}$ and it declines monotonically and when $\lambda < 1$, the hazard is very big as $t \to 0$ and it then declines towards zero as $t \to \infty$ (as time increases). The Log-logistic distribution therefore can be used to describe a hazard that first increases and later decreases or a monotonically decreasing hazard. For example, the hazard of an individual after a major surgery like a Kidney transplant increases just after surgery due to the open wounds and it later declines as the individual responds to the treatment.



(A) The probability density function of the Log-logistic distribution for different values of the shape and scale parameter.



(B) The Loglogistic survival function.



(C) The shapes of the Log-logistic hazard functions for some selected values of the shape parameter.

FIGURE 2.3: The probability density, the Survival function and the Hazard functions of the Log-logistic distribution.

Chapter 3

Regression models in survival analysis

3.1 Introduction

In most cases, the survival time data is presented together with the other factors that help to explain the survival time. These factors may include; age, sex, health status among others . With this additional information one may decide to model the survival time including such information in the model, the resulting model after considering the available information (factors/ covariates) is known as a regression model. A regression model can be parametric or semi-parametric. Consider a vector of covariates $\boldsymbol{X}$, in this case the regression model linking the covariates to the hazard is given by:

$$h\left(t|\boldsymbol{X}\right) = h_0\left(t\right)\eta\left(\boldsymbol{X}^T\boldsymbol{\beta}\right),$$

where $h_0\left(t\right)$ is the baseline hazard, $\boldsymbol{\beta}$ is the vector of the regression parameters and $\eta\left(\boldsymbol{X}^T\boldsymbol{\beta}\right)$ is the link function. The link function has to be a positive function since the hazard function is always positive. The best choice for the link function is the exponential

$$\eta\left(X\right) = \exp\left(X\right) > 0, \quad \forall X. \tag{3.1}$$

From equation (3.1),

$$h\left(t|\boldsymbol{X}\right) = h_0\left(t\right)\exp\left(\boldsymbol{X}^T\boldsymbol{\beta}\right).$$

When we specify the distribution of the baseline hazard function, the resulting model is known as a parametric regression model and such a model takes on the name of the distribution assumed. For example;

- The Exponential regression model;
- The Weibull regression model and
- others.

3.1.1 The Exponential and Weibull regression models

The exponential model has a constant hazard

$$h_0(t) = \lambda.$$

Give a vector of covariates $\boldsymbol{X}$, the regression model for the hazard function is given by:

$$h(t, \boldsymbol{X}) = \lambda \exp\left(\boldsymbol{X}^T \boldsymbol{\beta}\right).$$

In a similar way the Weibull regression model for the hazard function can be derived. In this case,

$$h_0(t) = \alpha \lambda^{\alpha} t^{\alpha - 1}.$$

The Weibull regression model for the hazard as a function of covariates is given by:

$$h(t, \boldsymbol{X}) = \alpha \lambda^{\alpha} t^{\alpha - 1} \exp\left(\boldsymbol{X}^T \boldsymbol{\beta}\right),$$

where $\boldsymbol{\beta}$ represents a vector of regression coefficients.

3.1.2 Semi-parametric models

3.1.2.1 The Cox-proportional hazard model

The most commonly used regression model is the Cox-proportional hazard model. With this model the distribution for the baseline hazard function is not specified and that is why it is called a semi-parametric model. The Cox-proportional hazard model is a more general model in modelling the hazard and survival function because it does not place distributional assumptions on the baseline hazard. The Cox model was introduced by Cox (1972). It has the from:

$$h\left(t|\boldsymbol{X}\right) = h_0\left(t\right)\exp\left(\boldsymbol{X}^T\boldsymbol{\beta}\right).$$

The measure of the effect of the given covariates on survival time is given by the hazard ratio denoted as HR. Consider a categorical variable with two levels say $X = 1$ and $X = 0$, then the hazard ratio for the two groups is defined as:

$$HR = \frac{h\left(t|X=1\right)}{h\left(t|X=0\right)} = \exp\left(\beta\right).$$

When $HR = 1$, it implies that the individuals in the two categories are at the same risk of getting the event, when $HR > 1$, it implies that the individuals in the first category $(X = 1)$ are at a high risk of getting the event and if $HR < 1$, the individuals in the second category $(X = 0)$ are at a high risk of getting the event.

The Cox-proportional hazard model assumes a proportional hazard this therefore implies that the model can not be used in the situation where the assumption is violated.

3.2 Estimation of unknown parameters in both parametric and semi-parametric regression models

To demonstrate how to determine the estimates of the unknown parameters in both parametric and semi-parametric regression models we used the most common model in survival analysis, the Cox-PH model.

The main objective in fitting the Cox proportional hazard model is to come up with estimates of the regression parameters ($\boldsymbol{\beta}'$s). Assuming that there are no tied event times,

Cox (1972) described how to estimate the regression parameters ($\boldsymbol{\beta}'$s) by maximising the partial likelihood given by:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left[\frac{\exp\left(\boldsymbol{X}_i^T \boldsymbol{\beta}\right)}{\sum_{j \in R(t_i)} \exp\left(\boldsymbol{X}_j^T \boldsymbol{\beta}\right)} \right]^{\delta_i}; \quad i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m, \tag{3.2}$$

where $R(t_i) = \{j : t_j \geq t_i\}$, denotes the risk set at time t_i, n represents the number of individuals in the data set and m the observed survival times. Only event times contribute their factor to the numerator but both the censored and uncensored observations are included in the denominator where the sum over the risk set includes all individuals who are still at risk just before time t_i. It is easy to work with the partial log-likelihood which is given by;

$$l(\boldsymbol{\beta}) = \ln\left(L(\boldsymbol{\beta})\right) = \sum_{i=1}^{n} \boldsymbol{\delta}_i \left(\boldsymbol{X}_i^T \boldsymbol{\beta}\right) - \sum_{i=1}^{n} \boldsymbol{\delta}_i \ln \left[\sum_{j \in R(t_i)} \exp\left(\boldsymbol{X}_j^T \boldsymbol{\beta}\right) \right]. \tag{3.3}$$

Let $\hat{\boldsymbol{\beta}}$, denote the maximum partial likelihood estimate for $\boldsymbol{\beta}$ obtained by maximising the partial log-likelihood function (3.3), the first derivative of $l(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ is called a vector of efficient scores and is given by:

$$U(\boldsymbol{\beta}) = \frac{dl}{d\beta} = \boldsymbol{X}^T \boldsymbol{\delta} - \sum_{i=1}^{n} \boldsymbol{\delta}_i \frac{\sum_{j \in R(t_i)} \exp\left(\boldsymbol{X}_j^T \boldsymbol{\beta}\right) \boldsymbol{X}_j}{\sum_{j \in R(t_i)} \exp\left(\boldsymbol{X}_j^T \boldsymbol{\beta}\right)}, \tag{3.4}$$

where $\boldsymbol{\delta} = [\delta_1, \ldots, \delta_n]^T$ denotes the vector of censoring indicators and $\boldsymbol{X}$ is the $n \times p$ matrix of covariate values with the j^{th} row containing covariates of the j^{th} individual $(\boldsymbol{X}_{(j)})$. To calculate the maximum likelihood estimates $\hat{\boldsymbol{\beta}}$, we solve a nonlinear system $U(\boldsymbol{\beta}) = 0$ and we use the Newton-Raphson algorithm (Jennrich and Robinson, 1969).

The information matrix $I(\boldsymbol{\beta})$ is given by the negative of the second derivative of $l(\boldsymbol{\beta})$.

$$I(\boldsymbol{\beta}) = -\frac{d^2 l}{d\beta^2}. \tag{3.5}$$

For large samples, the maximum likelihood estimate $\hat{\boldsymbol{\beta}}$ is known to follow asymptotic p-variate normal distribution:

$$I(\boldsymbol{\beta})^{\frac{1}{2}} \left\{\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\} \rightarrow_{n \rightarrow \infty} N\left[\mathbf{0}, I_p\right]. \tag{3.6}$$

The inverse of the information matrix $I^{-1}\left(\hat{\boldsymbol{\beta}}\right)$ is a consistent estimate of the covariance matrix of $\hat{\boldsymbol{\beta}}$. It is used to construct confidence intervals for the components of $\boldsymbol{\beta}$.

3.3 Drawing inference on the hazard ratios and the regression parameters in regression models.

Aalen and Gjessing (2008a) argues that to test for a simple null hypothesis, one may use the likelihood based tests. These tests include; The Likelihood ratio test, Wald's test and the Score test. These tests are asymptotically equivalent and they all follow a chi-square distribution with p degrees of freedom. Where p is the dimension of the vector of the regression parameters.

Assume that $\boldsymbol{\beta}$ is the vector consisting of regression parameters, we want to test the hypothesis that $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ against $H_a : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$. Where $\boldsymbol{\beta}_0$ is known.

- The Likelihood ratio test is given by:

$$T_{LR} = 2\left[l\left(\hat{\boldsymbol{\beta}}\right) - l\left(\hat{\boldsymbol{\beta}_0}\right)\right].$$

  Under the null hypothesis, the asymptotic distribution of T_{LR} is χ^2_p. We calculate the $p-$value at tail probability of the χ^2_p distribution as $P\left(\chi^2_p \geq T_{LR}\right)$.

- The Wald test

  The Wald test statistic has a quadratic form:

$$W = \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)' I\left(\hat{\boldsymbol{\beta}}\right)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right).$$

  The test statistic has a chi-square distribution with p degrees of freedom, where p is the dimension of the information matrix $I\left(\boldsymbol{\beta}\right)$. In order to conduct a one-sided test, we compare W to $\chi^2_{1-\alpha}$ at the α-level of significance. Asymptotically, $\hat{\boldsymbol{\beta}}$ follows a a standard normal distribution that is.

$$\hat{\boldsymbol{\beta}} \sim N\left[\boldsymbol{\beta}, I^{-1}\left(\hat{\boldsymbol{\beta}}\right)\right].$$

  Taking the square root of W, we get a standard normal statistic,

$$Z = \frac{\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}}{s.e\left(\hat{\boldsymbol{\beta}}\right)} \quad \sim N\left(0, 1\right),$$

with

$$s.e\left(\hat{\boldsymbol{\beta}}\right) = \frac{1}{\sqrt{I\left(\hat{\boldsymbol{\beta}}\right)}}.$$

Using this statistic together with the standard error, we can conduct a two-sided test of the hypothesis about a single parameter $\boldsymbol{\beta}_i$ and construct its confidence intervals at the $\alpha-$level of significance.

According to Dobson (2002) the statistic

$$Z_i^* = \frac{\hat{\boldsymbol{\beta}}_i}{s.e\left(\hat{\boldsymbol{\beta}}_i\right)},$$

is called a Wald test statistic with a null hypothesis

$$H_0 : \boldsymbol{\beta}_i = \mathbf{0}.$$

Given that $\hat{\boldsymbol{\beta}}_i$ is the estimate of the regression parameter $\boldsymbol{\beta}_i$, then $\hat{h_i} = \exp\left(\boldsymbol{\beta}_i\right)$ is the hazard ratio estimate of a covariate X_i. By using the Delta method described in Appendix A, the variance of $\hat{h_i}$ is given by:

$$Var\left(\hat{h_i}\right) = \hat{h_i}^2 Var\left(\hat{\boldsymbol{\beta}}_i\right).$$

The $100\left(1-\alpha\right)$ for $\hat{h}$ is given by:

$$\hat{h} \pm s.e\left(\hat{h}\right) Z_{1-\frac{\alpha}{2}}.$$

If 1 lies within the confidence interval of $\hat{h}$, then the variable is non-significant or has no effect on the response variable.

3.4 Variables and best fitting model selection

Schoenfeld Residuals

In order to fit the standard cox-proportional hazard model, one has to be aware of one of its main assumptions. The model assumes that the hazard of the different strata formed by the levels of the covariates are proportional (Abeysekera and Sooriyarachchi, 2009, Collett, 1994, Cox, 1972). One can use the Kaplan-Meier plots to test for this assumption but these graphical techniques may be inadequate in cases where the violation of the proportional hazard assumption is marginal.

Kleinbaum et al. (2002) presents the Goodness of fit (GOF) testing approach. This approach gives a test statistic and a p-value for assessing the proportional hazard assumption. This test enables a researcher to make an objective decision than when using the graphical method. A number of tests have been presented in literature but for this research we used the one discussed by Schoenfeld (1982). The schoenfeld residuals are further discussed by Grambsch and Therneau (1994). The idea behind this statistical test is that if the PH assumption holds for a particular covariate, then the schoenfeld residuals for that covariate will not be related to the survival time.

Akaike information criteria

According to Akaike (1987), the empirical principle of parsimony in a statistical model building dictates that the increase in the number of parameters should be stopped as soon as it is seen that further increase does not produce a significant improvement of the fit of the model to the data. He suggested the Akaike Information Criteria (AIC) given by:

$$AIC = -2\ln(L) + 2\text{number of parameters}$$

as the measure of GOF of a model defined with parameters estimated by the maximum likelihood method. The value of the AIC will always increase if an unnecessary variable is included in the model. This therefore implies that the smaller the AIC the better the model.

Chapter 4

Application to UDHS 2011

4.1 Data description and exploratory analysis

The data for this study is from the 2011 Uganda Demographic Healthy Survey (UDHS) which was collected from may 2011 through December 2011 (DHS, 2012). This is the fifth comprehensive survey conducted in Uganda as part of the world wide demographic and health survey, the first one having been done in 1988-1989 (Macro, 2004). The surveys were funded by USAID, UK, department of international development (DFID), The president's Emergency plan for AIDS Relief, the Government of Uganda, the Health Partnership Fund, the United Nations Children's Fund (UNICEF), the United Nations Population Fund (UNFPA) and the government of Japan with the technical assistance of ORC Macro, Calverton, Maryland, USA through its project MEASURE DHS.

A representative sample of 10,086 for the 2011 households was randomly selected to participate in the survey, and these households contained 9,247 women aged between 15-49 years of age (DHS, 2012). These women were interviewed by distributing questionnaires and information on their birth history was recorded, the information on the birth history includes;

- Child survival status (alive or dead);
- Male or female;
- Age at the day of interview if the child is alive and if dead, the age at death;
- Year of birth.

Information on other characteristic of the woman recorded in the survey were the education level, type of occupation, place of residence, partner's education, household's income and many other characteristics.

Out of the 9,247 women interviewed, it should be noted that only 6,692 women were considered for this research because we excluded all births in the year 2011. 62 of the children born in the period of 2006-2010 were recorded as still births.

The survival time for the children who were dead and those still alive was the age of the child in months and it was calculated as:

$$\text{survival time for children still alive} = \frac{(\text{Date of interview } - \text{Date of birth})}{12},$$

$$\text{survival time for children who are dead} = \frac{\text{Age at death (in years)}}{12},$$

where date of interview, date of birth and age at death are all provided in the data set.

The factors affecting under-five child survival from the data set were selected based on the paper by Mosley et al. (1984). These factors were categorised into the following four categories; social demographic, social economic, environmental and proximate or biological.

The social demographic factors include;

- Mother's age at first birth;
- Sex of household head;
- Religion;
- Mother's age group.

The social economic factors include;

- Social economic status of the family;
- Mother's education;
- Partner's education.

The environmental factors include;

- Source of drinking water;
- Type of place of residence.

The proximate and biological determinants include;

- Type of birth (multiple or single birth);
- Sex of the child;
- Birth order;
- Previous birth interval.

The table below shows the distribution of deaths of the children under the age of five for each factor level included in the analysis;

TABLE 4.1: Distribution of births and deaths by survival determinants.

	% of deaths	Child is dead	N(Total)		% of deaths	Child is dead	N(Total)
Mother's education level				**Mother's occupation**			
Illiterate Mothers	7.7	344	4493	Not-working	6.9	93	1353
Mother completed primary	6.4	119	1868	Sales and Services	6.5	110	1699
Secondary and higher	4.2	14	331	Agriculture	7.5	274	3640
Partner's level of education				**Births in past 5 years**			
Illiterate Father	7.7	266	3446	1-Birth	4.5	93	2075
Father completed primary	6.9	170	2457	2-Birth	6.5	227	3515
Secondary and higher	5.2	41	789	3-Births	13.6	140	1027
Birth status				4-Births	22.7	17	75
Singleton births	6.7	431	6479	**Births in past 1 year**			
Multiple births (Twins)	21.5	46	213	No-births	6.8	309	4521
Sex of the child				1-Birth	7.6	163	2134
Males	7.8	258	3325	2-Births	13.5	5	37
Females	6.3	212	3367	**Children Under 5 in Household**			
Type of place of residence				No-child	34.9	101	289
Urban	5.8	81	1389	1-Child	10.5	178	1689
Rural	7.5	396	5303	2-Children	4.9	146	2977
Wealth index				3-Children	2.5	35	1384
Poorest	7.5	131	1754	4-Children	4.8	17	353
Poorer	8.5	112	1317	**Mother's age group**			
Middle	7.2	86	1195	Less than 20 years	8.9	29	325
Richer	6.9	72	1041	20-29 years	6.5	235	3611
Richest	5.5	76	1385	30-39 years	7.4	164	2218
Children ever born				40 years^{+}	7.9	49	538
One child	3.3	20	601	**Birth order number**			
Two children	7.1	81	1146	First child	7.6	95	1249
Three children	6.6	67	1020	Second to Third child	5.6	117	2091
Four and more	7.9	309	3925	4^{th}-6^{th} child	7.1	149	2098
Birth order number				th+child	9.1	116	1254
First child	7.6	95	1249	**Sex of household head**			
Second to Third child	5.6	117	2091	Male	6.7	341	5112
4^{th}-6^{th} child	7.1	149	2098	Female	8.6	136	1580
7^{th}+child	9.2	116	1254	**Source of drinking water**			
Religion				Piped water	5.9	76	1280
Catholics	7.4	217	2939	Borehole	7.3	216	2947
Muslims	7.5	69	921	Well	6.7	93	1354
Other Christians	6.8	187	2758	Surface/Rain/Pond/Lake/tank	8.5	70	826
Others	5.4	4	74	Other	7.7	22	285
Type of toilet facility				**Age at first birth**			
Flush toilet	4.1	5	121	Less than 20 years	7.5	347	4638
Pitlatrine	6.9	376	5407	20-29 years	6.2	127	2026
No-facility	8.2	96	1164	30-39 years	12.0	3	25

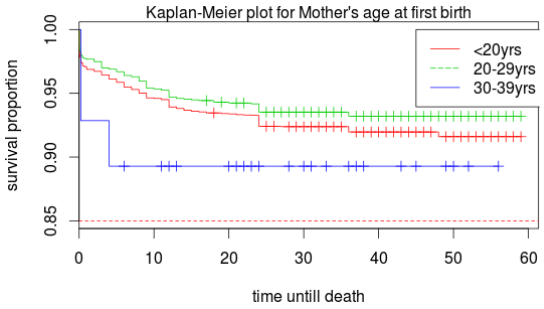
In this study most of the variables were categorical. The other variables which were not categorical, their categorizations were adopted from previous research and the paper by (Bailey, 1988, Bolstad and Manda, 2001, Croke, 2012, Mbonye et al., 2012, Mosley et al., 1984).

Table (4.1) shows that among the illiterate mothers, out of the 4493 children born 7.7% died before celebrating their fifth birthday which was the highest death proportion compared to other education levels of mothers. These were followed by mothers who had completed primary with 6.7% of the deaths. The category of mothers with the least percentage of deaths was those who had acquired secondary and higher education with 4.2% of deaths. The results in the table also show that the number of men who have achieved secondary and higher education is more than double that of women. Out of the 789 children born in families where the father is illiterate 7.7% of the children died before reaching the age of five. The families where the father had achieved secondary and higher education, 6.2% of the children had died before reaching the age of five which was the least death proportion compared to the other levels of father's education. Children born as a result of multiple births recorded the highest percentage of death compared to those as a result of a singleton birth. Out of the 213 born 21.5% had died before the age of five and this was the highest proportion compared those born out of a singleton birth which recorded 6.7% of the death. Most of the children in this sample were born to women who resided in rural areas. In this sample 5303 children were born to mothers who resided in the rural areas and they also recorded the highest percentage of deaths of 7.5% compared to those born of parents who resided in urban areas which was recorded to be 5.8% of the deaths. Families with no decent toilet facility recorded the highest percentage of children under the age of five deaths of about 8.2% compared those families with decent toilet facilities like a pit-latrine and flash toilets at 6.9% and 4.1% of the deaths respectively. The table also summarises the distribution of deaths and births of children in all the other factors included in this study.

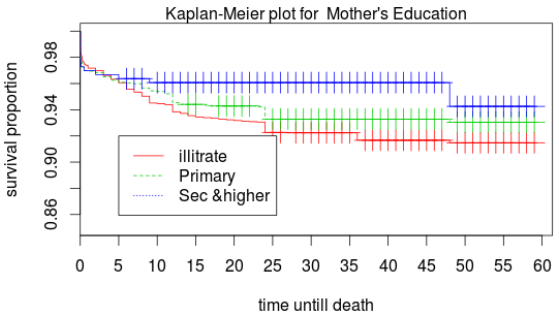
Non-parametric exploratory analysis methods

Given that t_i represents the time to death of a child under-five years of age (age of the child) in the Uganda DHS data set 2011, non-parametric methods which are mainly graphical were used to describe how the risk of death for the children under-five is distributed across the strata of some of the chosen covariates. Note that what we meant by time to death is actually years of life until death for the child under the age of five years. This time is recorded in months for our analysis. In other words it is not a typical

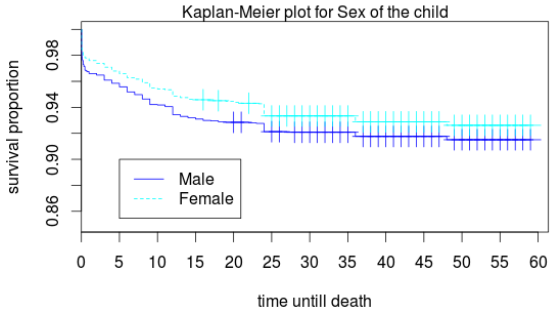
follow up study. The graphs produced below are Kaplan-Meier and Cumulative hazard plots for a few selected factors affecting under-five child survival in Uganda from the UDHS 2011 data set.



(A) Mother's age at first birth.

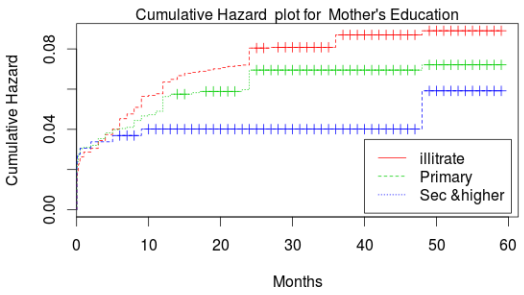


(B) Mother's education level.

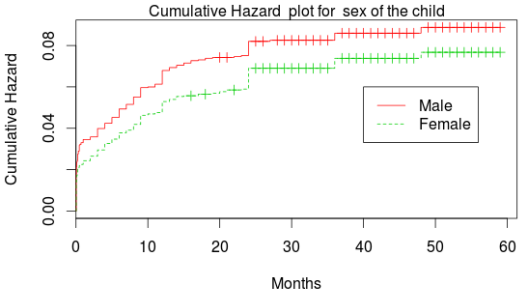


(C) Sex of the child.

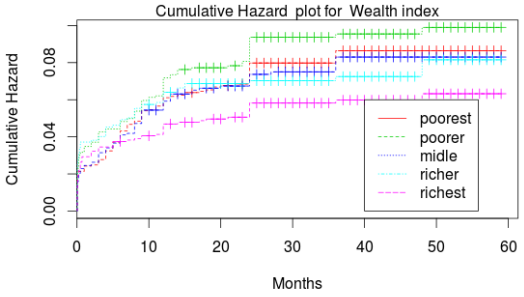
Figure (4.1b) shows that children born of mothers with secondary and higher education are at a higher chance of surviving to the age of five years than children born of mothers with primary or no education at all. The survival curve for children with mothers who have acquired secondary and higher education is above the survival curve of those children born of mothers who have primary or illiterate. From Figure (4.1c) one sees that a female child is at a higher chance of survival than a male child. From Figure (4.1b) one sees that mothers whose age at first birth birth was below 20 years were at a higher probability of having their children dead before reaching the age of five years, this is because the survival curve of this category of children was below the one for children whose mother's age at first birth was between the age of $20 - 29$years. This is probably due to the ill preparedness of these mothers on parenting plus birth complications. The probability of surviving becomes less for children born of mothers whose age at first birth is between $30 - 39$years, this is probably due to the birth complications involved with mothers of this age group.



(D) Mother's education.



(E) Sex of the child.



(F) Family wealth index.

Figure (4.1d) shows the a cumulative hazard curve for the children mother's level of education. The cumulative hazard curve for children whose mothers have secondary and higher education is below the rest of the other children whose mothers have a lower education level which indicates a lower probability of death for children from such mothers and an increased probability of death for the rest of the other children whose mothers had primary or no education at all. The male child had an increased probability of death, and this can be seen from the cumulative hazard curve being higher than that of the female child under the age of five see Figure (4.1e). Figure (4.1f) indicated that children born from the richest families were exposed to a lower probability of death before reaching the age of five compared to the rest of the children from richer, middle, poorer and poor families. Figure (4.1f) shows that the curve for the cumulative hazard of death for children in rich families is below all the other curves.

4.1.1 Fitting a standard Cox-proportional hazard model on the 2011 UDHS data

Suppose that $t = (t_1, t_2, \ldots, t_n)'$ are n independent identically distributed survival time.

The Proportional Hazard Model which stems from the work of the Cox (1972), assumes that for individuals with a vector of covariate $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$, the hazard rate at time t is given by:

$$h\left(t|\boldsymbol{X}\right) = h_0\left(t\right) \exp\left(\boldsymbol{X}^T\boldsymbol{\beta}\right).$$

Assuming that the survival times or time to death of the children under the age of five in the 2011 UDHS data are identically and independently distributed, the CPH model with covariates such as; mother's education, father's education, total number of children ever born, sex of the child, type of place of residence, wealth index, birth order, age at first birth, previous birth interval, number of births in the past one year, number of births in the past five years, mother's age, sex of the household head, source of drinking water, mothers occupation and religion was fitted on the 2011 UDHS data. A total of 6692 survival times for the children under the age of five were considered in this study ($n = 6692$) together with the survival indicator which helps us to identify whether the survival time was censored or observed. The results are presented in the table below.

Results

The Figures (4.1a), (4.1b) and (4.1c) present Kaplan-Meier curves that show how the risk of death of children under-five years of age is distributed across the categories of a given covariates. However, it is not possible to include all the Kaplan-Meier plots in this thesis for all the covariates included in the study but by fitting a univariate Cox-proportional hazard model, the hazard ratios have enough information to give about the distribution of the hazard on a given covariate.

Table (4.2) presents the unadjusted hazard ratios, the p-values and the 95% confidence intervals from fitting the univariate cox-proportional hazard model. The variables that were found significant by using the likelihood ratio test at 0.05 level of significance include; social economic factors (mothers education, father's education and wealth index), demographic factors (type of place of residence, sex of the household head sex of the child, total children ever born, number of births in the last one year and number of births in the past five years) and the biological factors (type of birth, birth order and the previous birth interval) are presented in the table.

Children born of mothers with secondary and higher education had a low risk (HR=0.56, p-value 0.03) of dying before reaching the age of five years compared to those children

TABLE 4.2: The univariate standard Cox-ph model.

Variable	Unadjusted HR[95%CI]	p-value	Variable	Unadjusted HR [95%CI]	p-value
Mother's education			4yrs +	0.62[0.45, 0.86]	<0.01
			Number of births in the past one year		
Illiterate	1	...			
Primary	0.83 [0.68, 1.03]	0.09			
secondary	0.56 [0.33, 0.95]	0.03			
and higher			No birth	1	...
Father's education			1 birth	1.18[0.98, 1.43]	0.08
			2	2.34[0.97, 5.67]	0.06
			Number of births in the last five years		
Illiterate	1	...			
Primary	0.89 [0.74, 1.08]	0.25			
Secondary	0.67[0.48, 0.92]	0.015			
and higher					
Sex of the child			1 births	1	...
			2 births	1.45 [1.14, 1.85]	<0.01
Male	1	...	3 births	3.11[2.39, 4.05]	<0.01
Female	0.83[0.69, 0.99]	0.04	4^+	5.70[3.40, 9.56]	<0.01
Total number of children ever born			**Mother's age**		
			< 20	1	...
			$20-29$yrs	0.66[0.45, 0.98]	0.04
			$30-39$yrs	0.74[0.50, 1.10]	0.14
			40+	0.90 [0.57, 1.43]	0.67
1.child	1	...	**Sex of household head**		
2.	2.04[1.25, 3.32]	<0.01			
3.	1.84 [1.12, 3.04]	0.02			
4+.	2.22[1.41, 3.48]	<0.01			
Type of place of residence			Male	1	...
			Female	1.30 [1.07, 1.59]	0.01
			Source of drinking water		
Rural	1	...			
Urban	1.29[1.01, 1.63]	0.04			
Wealth index			Piped water	1	...
Poorest	1	...	Borehole	1.24[0.96, 1.62]	0.1
Poorer	1.147[0.89, 1.48]	0.29	Well water	1.17 [0.86, 1.58]	0.32
Middle	0.96[0.73, 1.26]	0.78	Surface/pond/		
Richer	0.93[0.70, 1.24]	0.64	lake/Rain/etd	1.44 [1.04, 1.98]	0.03
Richest	0.73[0.55, 0.97]	0.03	Others	1.32 [0.82, 2.13]	0.25
Birth order			**Mother's occupation**		
1St	1	...			
2nd	0.72[0.55, 0.95]	0.02	Not working	1	...
3rd	0.92[0.71, 1.19]	0.53			
$4^{th}+$	1.22[0.93, 1.59]	0.15	Sales and Services	0.93[0.71, 1.23]	0.61
Age at first birth			Agriculture	1.08 [0.86, 1.37]	0.50
			Type of birth		
< 20	1	...			
20-29	0.84[0.68, 1.02]	0.08	single	1	...
30^+	1.52[0.49, 4.73]	0.47	birth		
Previous birth interval			Multiple births	3.66 [2.69, 4.96]	< 0.01
			Religion		
< 2yrs	1	...	Catholic	1	...
2yrs	0.69 [0.54, 0.87]	<0.01	Muslim	1.01[0.77, 1.33]	0.94
3yrs	0.66[0.49, 0.90]	0.01	Other Christians	0.91 [0.75, 1.11]	0.37
			Others	0.717[0.27, 1.93]	0.51

born of mothers who had not even achieved the basic primary education. The results about the mother's education are in agreement with the father's education, the risk of death of a child under the age of five born in a family with a father who has attained

secondary and higher education was at a lower risk of dying (0.67, p-value 0.015) compared to those with fathers who have not acquired even the basic primary education. A Male child under the age of five years had a higher risk of death compared to that of a female child. The hazard of a male child dying before reaching age five was 20% (HR=1.20, p-value 0.04) higher than that of a female child under age five.

However, not all the variables presented in Table (4.2) satisfy the proportionality hazard assumption. In order to fit the multiple covariate cox-proportional hazard model, we need to first test for the proportionality hazard assumption. To test for the proportionality hazard assumption, we based the test on the scaled Schoenfeld residuals and a command *cox.zph* in the R software is used.

TABLE 4.3: Testing the proportional hazard assumption using the scaled Schoenfeld residuals by Grambsch and Therneau (1994) see appendix 5.

Variable (Determinant)	χ^2**(df)**	**p-value**	**Variable (Determinant)**	χ^2**(df)**	**p-value**
Mother's education			3yrs	0.97	0.32
Illiterate	1	...	4yrs $^+$	2.53	0.11
Primary	4.83	0.03	GLOBAL	8.69	0.03
secondary and higher	7.52	<0.01	**Number of births in the past one year**		
GLOBAL	11.25	< 0.01	No birth	1	
Father's education			1 birth	0.7	0.40
Illiterate	1	...	2	1.24	0.27
Primary	0.51	0.48	GLOBAL	1.81	0.40
Secondary and higher	0.86	0.35	**Number of births in the last five years**		
GLOBAL	1.12	0.57	1 births	1	...
Sex of the child			2 births	0.11	0.75
Male	1	...	3 births	0.03	0.86
Female	1.99	0.16	4^+	5.00	0.03
Total number of children ever born			GLOBAL	5.85	0.12
1.child	1	...	**Mother's age**		
2.	5.39	0.02	< 20	1	...
3.	0.44	0.51	20 − 29yrs	0.16	0.69
4+.	0.26	0.61	30 − 39yrs	0.63	0.43
GLOBAL	14.61	< 0.01	40+	0.08	0.78
Type of place of residence			GLOBAL	5.58	0.13
Rural	1	...	**Sex of household head**		
Urban	8.43	< 0.01	Male	1	...
Wealth index		...	Female	0.07	0.79
Poorest	1	...	**Source of drinking water**		
Poorer	0.17	0.7	Piped water	1	...
Middle	0.00	0.98	Borehole	0.17	0.68
Richer	6.94	< 0.01	Well water	0.12	0.73
Richest	2.26	0.13	Surface/pond/		
GLOBAL	9.29	0.05	lake/Rain/etc	2.58	0.11
Birth order		...	Others	1.82	0.18
1St	1	...	GLOBAL	6.55	0.16
2nd	0.28	0.59	**Mother's occupation**		
3rd	6.69	< 0.01	Not working	1	...
4^{th}+	2.64	0.10	Sales and Services	0.202	0.65
GLOBAL	8.46	0.04	Agriculture	6.88	< 0.01
Age at first birth			GLOBAL	14.41	< 0.01
< 20	1	...	**Type of birth**		
20-29	0.10	0.75	single birth	1	...
30^+	0.41	0.52	Multiple births	13	< 0.01
GLOBAL	0.54	0.76	**Religion**		
Previous birth interval	1	...	Catholic	1	
< 2yrs	Ref		Muslim	0.009	0.92
2yrs	1.83	0.18	Other Christians	0.73	0.39
			Others	1.59	0.21
			GLOBAL	2.21	0.53

Table (4.3) shows the results of testing for the proportional hazard assumption and these results show that mother's education, total children ever born, type of place of residence,

Wealth index, birth order, mother's occupation and type of birth do not satisfy the proportionality hazard assumption because their p-values are less than 0.05 significance level from the scaled Schoenfeld residuals test. Based on the scoenfeld residuals, these factors can not be fitted in the final cox-proportional hazard model to get the adjusted hazard ratios. Other variables like the sex of the child satisfy this assumption and can therefore be fitted in the final cox-proportional hazard model.

TABLE 4.4: The adjusted hazard ratios from fitting the Cox-proportional hazard model for only the variables that satisfy the proportionality hazard assumption.

Variable	Unadjusted HR[95%CI]	Adjusted HR[95%CI]	p-values	Variable	Unadjusted HR[95%CI]	Adjusted HR [95%CI]	p-value
Mother's education				**Number of births in the past one year**			
Illiterate	1	1	...	No birth	1	1	
Primary	0.83 [0.68, 1.03]	...		1 birth	1.18[0.98, 1.43]	1.22 [1.01, 1.48]	0.04
secondary and higher	0.56[0.33, 0.95]	...		2	2.34[0.97, 5.67]	2.57[1.06, 6.25]	0.04
Father's education				**Number of births in the last five years**			
Illiterate	1	1	...	1 births	1		...
Primary	0.89 [0.74, 1.08]	0.92 [0.76, 1.12]	0.43	2 births	1.45[1.14, 1.85]	...	
Secondary and higher	0.67[0.48, 0.92]	0.72[0.51, 1.01]	0.06	3 births	3.11[2.39, 4.05]	...	
Sex of the child				4^{+}	5.70[3.40, 9.56]		...
Male	1			**Mother's age**			
Female	0.83 [0.69, 0.99]	0.83 [0.69, 0.99]	0.04	< 20	1	1	
Total number of children ever born				$20 - 29$yrs	0.66[0.45, 0.98]	0.71[0.48, 1.05]	0.08
				$30 - 39$yrs	0.74[0.50, 1.10]	0.79 [0.53, 1.19]	0.27
1.child	1			40+	0.90[0.57, 1.43]	0.99[0.62, 1.59]	0.98
2.	2.04 [1.25, 3.32]	...		**Source of drinking water**			
3.	1.84 [1.12, 3.04]						
4+.	2.22[1.41, 3.48]	...		Piped water	1	1	
Type of place of residence				Borehole	1.24[0.96, 1.62]	1.12 [0.86, 1.48]	0.39
				Well water	1.17[0.86, 1.58]	1.06[0.78, 1.45]	0.69
Rural	1		...	Surface/pond/			
Urban	1.29 [1.01, 1.63]	...		lake/Rain/etc	1.44 [1.04, 1.98]	1.28[0.91, 1.79]	0.15
Wealth index			...	Others	1.32 [0.82, 2.13]	1.21[0.75, 1.94]	0.44
Poorest	1	...		**Mother's occupation**			
Poorer	1.147 [0.89, 1.48]	...					
Middle	0.96[0.73, 1.26]			Not working	1		...
Richer	0.93 [0.70, 1.24]	...		Sales and Services	0.93 [0.71, 1.23]	...	
Richest	0.73[0.55, 0.97]	...		Agriculture	1.08 [0.86, 1.37]	...	
Birth order			...	**Type of birth**			
1St	1		...	single birth	1		...
2nd	0.72 [0.55, 0.95]	...		Multiple births	3.66[2.69, 4.96]	...	
3rd	0.92[0.71, 1.19]	...		**Religion**			
$4^{th}+$	1.22[0.93, 1.59]	...		Catholic	1	1	
Age at first birth				Muslim	1.01[0.77, 1.33]	1.02[0.77, 1.34]	0.91
				Other Christians	0.91 [0.75, 1.11]	0.94[0.77, 1.14]	0.51
< 20	1	1		Others	0.717 [0.27, 1.93]	0.67[0.25, 1.81]	0.43
20-29	0.84[0.68, 1.02]	0.86[0.69, 1.06]	0.16				
30^{+}	1.52[0.49, 4.73]	1.59[0.51, 5.02]	0.42				
Previous birth interval							
< 2yrs	1			...			
2yrs	0.69[0.54, 0.87]		...				
3yrs	0.66[0.49, 0.90]	...					
4yrs $^{+}$	0.62 [0.45, 0.86]	...					
Sex of household head							
Male	1	1					
Female	1.30 [1.07, 1.59]	1.33[1.09, 1.63]	0.01				

Table (4.4) show that sex of the child and number of births in the past one year were strongly associated with high under five child mortality rates. There was not enough evidence to conclude that father's eduction, age of the mother at first birth, source of drinking water, mother's age group and religion are strongly associated with high under-five children mortality rates.

TABLE 4.5: The best fitting model selected by the AIC by Akaike (1973) of the standard cox proportional hazard model.

Variable	$\beta(s.e)$	**HR** $[95\%CI]$	**pvalues**
Father's education			
Illiterate	1		...
Primary	-0.09 (0.09)	0.90 [0.75, 1.09]	0.31
Secondary and Higher	-0.41 (0.17)	0.66 [0.47, 0.92]	0.014
Sex of the child			
Male	1		...
Female	-0.18 (0.09)	0.83 [0.69, 0.99]	0.04
Number of births in the past one year			
No.birth	1		...
1.birth	0.20 (0.09)	1.22 [1.01, 1.48]	0.04
2.births	0.922(0.45)	2.51[1.04, 6.09]	0.04
Household head			
Male	1		...
Female	0.28 (0.10)	1.33[1.09, 1.62]	0.01
Mother's age group			
less than 20yrs	1		...
20-29	-0.38 (0.19)	0.68 [0.46, 1.01]	0.05
30-39	-0.27 (0.20)	0.77 [0.51, 1.14]	0.17
40+	-0.05 (0.24)	0.95 [0.59, 1.51]	0.83

$**$*Best fitting model* $**$
$**$ < 0.05, *level of significance* $**$.

Using Akaike's Information Criterion (AIC) by Akaike (1973), the best fitting standard cox proportional hazard model consisted of five variables which include father's education, sex of the child, number of births in the past one year, mother's age group and the sex of the household head, Table (4.5).

From Table (4.5), there is a enough evidence to conclude that sex of household head, number of births in the past one year and sex of the child affect under-five children survival in Uganda. Households headed by women or regarded as broken families showed a high risk of death for children under the age of five compared to those that are headed my males. A female child had a low risk of death compared to a male child and this is attributed to the value most communities in Uganda attach to a female child. A female child is seen as source of bride price or wealth they are therefore given more care which may explain why they have an edge over the males.

Chapter 5

Frailty modelling

This chapter seeks to analyse the factors affecting under-five child mortality in Uganda taking into account any extra heterogeneity present in the data using the so called frailty models. The frailty term takes into consideration the situation where some of the children may be exposed to the hazard of death before the age of five more than the others. This brings about the possibility that due to the unobserved or unmeasured causes some children are more likely to experience the hazard than others. The frailty term captures total effects of all factors that influence the child's risk of death that are not included in the standard Cox-proportional hazard model presented in chapter 3. The model presented in this chapter accounts for both observed and unobserved effects. Nonetheless the ensuing model is still based on the standard proportional hazard model. Frailty terms enter the Cox proportional hazard model as random effects. The estimated variance of the frailty effects is used to test if the frailty term is significant. This can be used to test the variability between individuals with the same cluster for example a household or community. A zero variance of the frailty term indicates that observations from the same household or community are independent. A large variance indicates a large heterogeneity or difference across a given household or community and therefore a greater correlation among individuals from the same household or community.

In this chapter we therefore present preliminary concepts of the model, the development and the statistical estimation algorithms which are supported in STATA and R software. The tables presented at the end of this chapter show the results from the analysis.

5.1 History about frailty models

In recent years, research papers and books have been written to show how one can extend the proportional hazard models to analyse more complex survival data (Hanagal, 2011,

Vaupel and Stallard, 1979, Wienke, 2009). Among them are the research papers on frailty models. These models help to explain the relationship between individuals in a given cluster or the difference among individuals in different clusters. The concept of frailty was introduced by Vaupel and Stallard (1979) showing that some individuals are more frail or susceptible or at risk than others although they may appear to be similar while considering the observable or measurable attributes like sex, age and weight.

A frailty model is a hazard model with a multiplicative frailty factor. The major assumption of a frailty model is that the information about the hidden internal or external factors is contained in the shape and structure of the hazard function and in the form of the frailty distribution (Hanagal, 2011).

5.1.1 Model development

In the presence of covariates with a covariate vector represented as $\mathbf{X}$, the Cox-proportional hazard model is:

$$h\left(t, \boldsymbol{X}\right) = h_0\left(t\right) \exp\left(\boldsymbol{X}^T \boldsymbol{\beta}\right), \tag{5.1}$$

where $\boldsymbol{X} = [X_1, X_2, \ldots, X_n]$ and $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_n]$ are the covariate and regression parameter vectors respectively.

In the presence of the unobserved components represented by a vector denoted as $\boldsymbol{U}$, (5.1) is modified as follows:

$$h\left(t, \boldsymbol{X}, \boldsymbol{U}\right) = h_0\left(t\right) \exp\left(\boldsymbol{X}^T \boldsymbol{\beta} + \boldsymbol{U}\right) = h_0\left(t\right) \exp\left(\boldsymbol{U}\right) \exp\left(\boldsymbol{X}^T \boldsymbol{\beta}\right), \tag{5.2}$$

let $\boldsymbol{Z} = \exp\left(\boldsymbol{U}\right)$ then:

$$h\left(t, \boldsymbol{X} | \boldsymbol{Z}\right) = \boldsymbol{Z} h_0\left(t\right) \exp\left(\boldsymbol{X}^T \boldsymbol{\beta}\right). \tag{5.3}$$

Equation (5.3) is what is known as a frailty model and $\boldsymbol{Z}$ is a variable representing the frailty term.

The frailty models are in two different forms;

1. The univariate frailty model;
2. The multivariate frailty model.

5.1.2 The univariate frailty model

Given that an individual child under-five years of age i, $i = 1, 2, \ldots, n$ has a survival time denoted as t_i, the covariate vector $\boldsymbol{X}_i$, with a frailty term denoted as z_i; Vaupel et al (1979) stated that the hazard function of the individual i is given as:

$$h_i\left(t_i, \boldsymbol{X}_i | z_i\right) = z_i h_0\left(t_i\right) \exp\left(\boldsymbol{X}_i^T \boldsymbol{\beta}\right). \tag{5.4}$$

When $z_i > 1$ (frailty is greater than one), the individual i is said to be at an increased risk of failure and therefore more frail than an average individual in that given cluster or group. On the other hand if $z_i < 1$, the individual is less frail and therefore tends to survive longer. This implies that a more frail individual experiences failure earlier on than the less frail individual.

The survival function of individual i conditional on frailty is given by:

$$S_i\left(t_i, \boldsymbol{X}_i | z_i\right) = \exp\left(-z_i e^{\boldsymbol{X}_i^T \boldsymbol{\beta}} \int_0^{t_i} h_0\left(s, \boldsymbol{X}_i | z_i\right) ds\right) = \exp\left(-z_i H_0\left(t_i\right) \exp\left(\boldsymbol{X}_i^T \boldsymbol{\beta}\right)\right), \tag{5.5}$$

$H_0\left(t_i\right) = \int_0^{t_i} h_0\left(s\right) ds$ is the cumulative baseline hazard function. The survival function given in (5.5) is at an individual level and is therefore unobservable (Wienke, 2009).

The unconditional survival function of an individual i at the population level is given as the mean of the survival function conditional on frailty with respect to the frailty distribution:

$$S_i\left(t_i, \boldsymbol{X}_i\right) = E\left[S_i\left(t_i, \boldsymbol{X}_i | z_i\right)\right] = E\left[\exp\left(-z_i \exp\left(\boldsymbol{X}_i^T\right) H_0\left(t_i\right)\right)\right].$$

5.1.3 The distributions of frailty

In the previous section we have noted that to determine the unconditional survival function of an individual one needs to know the frailty distribution and this is the same for the unconditional hazard.

Given the conditional hazard of an individual i, the unconditional hazard (Average hazard)of an individual i is given by:

$$\bar{h_i}\left(t_i, \boldsymbol{X}_i\right) = \int_0^\infty z_i h_0\left(t_i\right) \exp\left(\boldsymbol{X}_i^T \boldsymbol{\beta}\right) f_z\left(z_i\right) dz_i,$$

assume that $h_0\left(t_i\right)$ is a constant, then:

$$\begin{aligned}\bar{h_i}\left(t_i, \boldsymbol{X}_i\right) &= h_0\left(t_i\right) \exp\left(\boldsymbol{X}_i^T \boldsymbol{\beta}\right) \int_0^\infty z_i f_z\left(z_i\right) dz_i, \\ &= h_0 \exp\left(\boldsymbol{X}_i^T \boldsymbol{\beta}\right) \bar{z_i},\end{aligned}$$

where $\bar{z_i} = \int_0^\infty z_i f_z\left(z_i\right) dz_i$ is the mean frailty.

Assuming one covariate $\boldsymbol{X}_i$ representing a variable gender which is $1 = Female, 0 = Male$:

$$\bar{h_i}\left(t_i, \boldsymbol{X}_i\right) = h_0 \exp\left(\boldsymbol{\beta}\right) \bar{z_i}. \tag{5.6}$$

Equation (5.6) shows that the average hazard or the unconditional hazard depends on the frailty distribution.

Most of the researchers assume a Gamma Frailty but frailty can be assumed to follow other distributions which include;

- The Log-normal distribution;
- The Positive Stable frailty model;
- The Inverse Gaussian frailty model;
- The compound Poisson frailty model.

The Gamma frailty model is the most frequently used frailty model due to the following reasons;

- The Gamma distribution takes on positive random variables. Since the frailty term is a positive random variable, it makes the Gamma distribution the most suitable choice for the frailty term;
- The Gamma frailty is conjugate to the hazard model thus simplifying computation (tractability);

- The Gamma distribution is flexible. That is to say, the pdf of a two parameter Gamma distribution is given by;

$$f(z,\alpha,\beta)=\frac{\beta^{\alpha}z^{\alpha-1}\exp(-\beta z)}{\Gamma(\alpha)},\quad \alpha>0,\quad \beta>0\quad \text{and}\quad z>0.$$

  The mean and variance are given as $E(Z)=\frac{\alpha}{\beta}$ and $V(Z)=\frac{\alpha}{\beta^2}$ respectively. Where α and β are the shape and scale parameters respectively. When the shape parameter is equal to 1 ($\alpha=1$), the distribution becomes exponential with parameter β and for large values of α, the distribution assumes a bell shape that is identical to that of the a normal distribution.

3. It is analytically tractable and easy to compute because of its simple Laplace transform.

$$\begin{aligned}L\{f(z)\}(s)&=\int_0^{\infty}f(z)\exp(-zs)\,dz\\&=\int_0^{\infty}\frac{\beta^{\alpha}z^{\alpha-1}e^{-\alpha z}}{\Gamma(\alpha)}\exp(-zs)\,dz\\&=\frac{\beta^{\alpha}}{(s+\beta)},\end{aligned}$$

where the parameter s is a complex number:

$$s=a+ib$$

with a and b real numbers.

Note: The mean and variance of the Gamma distribution can be obtained by using the first and second derivatives of the Laplace transform respectively.

$$\begin{aligned}L^{(1)}(s)&=\frac{-\alpha\beta^{\alpha}}{(s+\beta)^{\alpha+1}}\\L^{(2)}(s)&=\frac{\alpha(\alpha+1)\beta^{\alpha}}{(s+\beta)^{\alpha+2}},\end{aligned}$$

on evaluating the derivatives at $s=0$,

$$\begin{aligned}E(Z)&=-L^{(1)}(0)\\&=\frac{\alpha}{\beta},\end{aligned}$$

and

$$\begin{aligned} Var\left(Z\right) &= L^{(2)}\left(0\right) - \left(-L^{(1)}\left(0\right)\right)^2 \\ &= \frac{\alpha}{\beta^2}. \end{aligned}$$

5.1.4 The univariate semi-parametric Gamma frailty model

For simplicity, we restrict our selves to a one parameter Gamma distribution ($\alpha = \beta$), with mean $E\left(Z\right) = 1$ and variance $V\left(Z\right) = \frac{1}{\alpha}$. Assuming that the variance of Z, the frailty term is denoted as θ, then $V\left(Z\right) = \frac{1}{\alpha} = \theta$. The probability density function of a one parameter Gamma distribution $f\left(z\right)$:

$$f\left(z\right) = \frac{\alpha^{\alpha} z^{\alpha-1} \exp\left(-\alpha z\right)}{\Gamma\left(\alpha\right)}.$$

By using the fact that the variance $Var\left(Z\right) = \theta = \frac{1}{\alpha}$,

$$f\left(z\right) = \frac{z^{\frac{1}{\theta}-1} e^{\frac{-z}{\theta}}}{\Gamma\left(\frac{1}{\theta}\right) \theta^{\frac{1}{\theta}}}.$$

Let T denote the random variable representing the survival times and Z denote frailty which is distributed as a Gamma. The conditional survival function is given by:

$$S_i\left(t|z\right) = \exp\left(-zH_0\left(t\right)\right),$$

the unconditional survival function is derived by integrating z out from the conditional survival function;

$$\begin{aligned} S_i\left(t\right) &= E\left[S\left(t|z\right)\right], \\ &= \int_0^{\infty} e^{-zH_0 \exp\left(\boldsymbol{X}_i^T \boldsymbol{\beta}\right)(t)} f\left(z\right) dz = L\left(H\left(t\right)\right), \end{aligned}$$

where L denotes the Laplace transform.

The Laplace transform of a one parameter Gamma distribution is given by:

$$L(s) = [1 + \theta s]^{\frac{-1}{\theta}},$$

thus the unconditional survival function can be written as:

$$S_i(t) = \left[1 + \theta H_0(t) \exp\left(\boldsymbol{X}_i^T \boldsymbol{\beta}\right)\right]^{\frac{-1}{\theta}}.$$

The likelihood denoted as $L(t, \boldsymbol{X}_i, \boldsymbol{\beta}, \theta)$ is calculated as:

$$L(t, \boldsymbol{X}_i, \boldsymbol{\beta}, \theta) = \prod_{i=0}^{G} \prod_{j=1}^{n_i} z_i^{\delta_i} h_0(t_{ij})^{\delta_i} \exp\left(\delta_i \boldsymbol{X}_{ij}^T \boldsymbol{\beta}\right), [S(t_{ij})]^{1-\delta_i}, \tag{5.7}$$

where G denotes the total number of clusters in the data set, and n_i the total number of individuals in cluster i.

5.1.5 The multivariate semi-parametric frailty model

In analysing time to event data, we always assume that the time to event of the individuals considered in the study are independent. However this may not always be true because there is a possibility of the survival times of individuals in the same cluster for example in a family or community to be correlated. The correlation among the survival time violates the independence assumption and such data can not be analysed using the univariate semi-parametric model. The data with correlated survival times is known as the multivariate survival data (Phipson and Mwambi, 2010). The models which were developed to analyse such data include the Shared frailty model.

5.1.6 The shared frailty model

This model was introduced by Clayton (1978) and Vaupel and Stallard (1979) (Wienke and Yashin, 2003). Hanagal (2011), Shih and Louis (1995) observed that individuals in the same cluster are assumed to share the same frailty and this is the reason why it is called the shared frailty model. Frailty is assumed to be independent across the groups or clusters while the survival times of individuals within the same group are conditionally dependent.

A univariate frailty model on the other hand assumes a frailty term for each individual and this frailty term represents the individual's unmeasured or hidden covariates after considering the measured covariates.

Let N denote the number of individuals in a given cohort with each individual in the cohort assigned to a group. Let the total number of groups be denoted by G such that, given the i^{th} group that consists of n_i individuals, then;

$$\sum_{i=1}^{G} n_i = N.$$

Given that $\delta = \begin{cases} 1, & \text{Uncensored or event occurred} \\ 0, & \text{Censored or event did not occur at last observation.} \end{cases}$

is the censoring indicator. Uncensored or event occurred implies that the survival time of the individual in the data set was observed and therefore recorded whereas Censored or event did not occur implies that the event time was not observed and in this case it is believed that the event time is more than the recorded time (right censored). The hazard function of the j^{th} individual of the i^{th} group is given as:

$$h_{ij}(t) = h_0(t) \exp\left(\boldsymbol{X}_{ij}^T \boldsymbol{\beta} + \boldsymbol{u}_i\right),$$

$\boldsymbol{X}_{ij}$ is a vector of covariates for the individual j in the i^{th} group, u_i the unobserved covariates and $h_0(t)$ the baseline hazard function.

As already stated, $z_i = \exp(u_i)$ is the frailty term. The hazard function can therefore be written as:

$$\begin{aligned} h_{ij}(t) &= h_0(t) \exp\left(\boldsymbol{X}_{ij}^T \boldsymbol{\beta} + \boldsymbol{u}_i\right), \\ &= z_i h_0(t) \exp\left(\boldsymbol{X}_{ij}^T \boldsymbol{\beta}\right). \end{aligned}$$

Note: The $z_i's$ are independent with an identical probability density function denoted as $f(z)$.

The full likelihood of the shared frailty model is given as:

$$L(t_{ij}, \boldsymbol{X}_{ij}, \theta) = \prod_{i=1}^{G} \prod_{j=1}^{n_i} (z_i)^{\delta_i} h_0(t_{ij})^{\delta_i} \exp\left(\delta_i \boldsymbol{X}_{ij}^T \boldsymbol{\beta}\right) [S(t_{ij})]^{1-\delta_i} \prod_{i=1}^{G} f(z_i).$$

5.2 Estimating parameters in a semi-parametric frailty model

Given that an individual $i's$ survival and hazard functions are given as $S_i(t_i)$ and $h_i(t_i)$ respectively and that the probability density function of the frailty term is given by $f(z)$ then the likelihood contribution by such an individual is denoted as $L_i(z_i, t_i, \boldsymbol{X}_i, \theta)$ is given by:

$$L_{fulli}(z_i, t_i, \boldsymbol{X}_i, \theta) = f(z) [S_i(t_i) h_i(t_i)]^{\delta_i} [S_i(t_i)]^{1-\delta_i}.$$

The estimates of the parameters are derived by differentiating the likelihood function with respect to the parameters but it is always easier to work with the Log-likelihood. Let us denote the log-likelihood as $l_{fulli}(z_i, t_i, \boldsymbol{X}_i, \theta)$;

$$\begin{aligned} l_{fulli}(z_i, t_i, \boldsymbol{X}_i, \theta) &= \ln(L_i(z_i, t_i, \boldsymbol{X}_i, \theta)), \\ &= \ln\left(f(z) [S_i(t_i) h_i(t_i)]^{\delta_i} [S_i(t_i)]^{1-\delta_i}\right). \end{aligned}$$

Since the frailty term is unobserved, we therefore need to find the observed likelihood by integrating the frailty term out with respect to its distribution.

$$l_{obsi}(z_i, t_i, \boldsymbol{X}_i, \theta) = \int_0^\infty f(z_i) l_i(z_i, t_i, \boldsymbol{X}_i, \theta) dz_i. \tag{5.8}$$

The parameter estimates are derived by differentiating equation (5.8) with respect to all the parameters in the model and the resulting equations solved simultaneously. With frailty models, it is not usually possible to solve the equations simultaneously due to the presence of latent variables (Arthur et al., 1977). If the latent variables are known, the solution becomes easy to find by using the maximum likelihood method as described above. The presence of the latent variables in addition to the unknown parameters requires us to use a more advanced method, these methods include;

- The Expectation-Maximisation Algorithm (EM-Algorithm);
- The Markov Chain Monte Carlo (MCMC) methods;
- The Monte Carlo EM (MCEM) approach;
- The penalised partial likelihood (PPL).

5.2.1 The Expectation-Maximisation Algorithm (EM-Algorithm)

The EM algorithm is a numerical method used to find the maximum likelihood estimates for the parameters of the statistical model which has got latent variables in addition to the unknown parameters given the observed data. With such models, the equation derived by differentiating the log-likelihood with respect to the unknown parameters and the variance of the distribution of the latent variable can not be solved directly (Arthur et al., 1977). Hanagal (2011), mentioned that the EM algorithm differs from the maximum likelihood estimation method (MLE) because it maximises the full likelihood instead of the likelihood in equation (5.7). In the full likelihood the latent variables are treated as additional parameters.

The EM-algorithm consists of two steps;

- The E-step;
- The M-step.

Let $\boldsymbol{Y} = (\boldsymbol{X}, \boldsymbol{Z})$ be a full data set for a given statistical problem, suppose that only $\boldsymbol{X}$ is observed and $\boldsymbol{Z}$ is unobserved or missing data and also suppose that $\boldsymbol{\Phi}$ denotes the vector of the unknown parameters. The complete data-log-likelihood function of $\boldsymbol{Y}$ is denoted as $L_{full}(\boldsymbol{\Phi}|\boldsymbol{Y})$ and the missing information (Frailty) is $\boldsymbol{Z}$ whose distribution can generally be taken to be Gamma. The EM algorithm consists first of the E-step, which is the expectation of the complete data likelihood with respect to the missing information $\boldsymbol{Z}$ given the observed data $\boldsymbol{X}$ and the initial parameter estimates (Small and Wang, 2003).

5.2.1.1 The E-step

Let $\boldsymbol{\Phi}^0$ denote some initial parameter estimates then on the first iteration, the E-step requires the calculation of:

$$Q\left(\boldsymbol{\Phi}, \boldsymbol{\Phi}^0\right) = E_{\boldsymbol{\Phi}^0}\left[\ln\left(L\left(\boldsymbol{\Phi}|z\right)\right)|x, \boldsymbol{\Phi}^0\right].$$

The complete-data log-likelihood is unobserved so we replace it by a conditional expectation given the observed data using the current fit for $\boldsymbol{\Phi}$.

5.2.1.2 The M-step

In M-step, the EM-algorithm maximises the expectation of the computed expectation in E-step.

$$\boldsymbol{\Phi}^1 = argmax_{\boldsymbol{\Phi}} Q\left(\boldsymbol{\Phi}, \boldsymbol{\Phi}^0\right).$$

We iterate between the Expectation step and the Maximisation step until convergence is realised.

At the k^{th} step, the E-step and the M-step are stated below:

$$Q\left(\boldsymbol{\Phi}, \boldsymbol{\Phi}^{k-1}\right) = E_{\boldsymbol{\Phi}^{k-1}}\left[\ln\left(L_{full}\left(\boldsymbol{\Phi}|z\right)\right)|x, \boldsymbol{\Phi}^{k-1}\right],$$
$$\boldsymbol{\Phi}^{k+1} = argmax_{\boldsymbol{\Phi}} Q\left(\boldsymbol{\Phi}, \boldsymbol{\Phi}^{k-1}\right).$$

Arthur et al. (1977) suggested a modified form of the EM algorithm called the Generalised Expectation Maximisation Algorithm (GEM) in cases where the solution to the M-step does not exist in closed form (Small and Wang (2003)). With the GEM algorithm, we choose $\boldsymbol{\Phi}^{k+1}$ such that $Q\left(\boldsymbol{\Phi}^{k+1}, \boldsymbol{\Phi}^k\right) \geq Q\left(\boldsymbol{\Phi}^k, \boldsymbol{\Phi}^{k-1}\right)$. Small and Wang (2003) described the M-step as one that chooses $\boldsymbol{\Phi}^{k+1}$ to increase the function $Q\left(\boldsymbol{\Phi}, \boldsymbol{\Phi}^k\right)$ over its value at $\boldsymbol{\Phi} = \boldsymbol{\Phi}^k$ other than maximising it over the parameter space ω. Details about the EM algorithm and its proofs of convergence we recommend a paper by Arthur et al. (1977).

5.2.2 Application of the EM algorithm on a shared Gamma frailty.

Let the full likelihood be denoted as $L_{full}\left(t_i, h_0, \boldsymbol{\beta}, \theta\right)$. The full likelihood of a shared frailty model of a cohort consisting of N individuals, each individual assigned to a group and with a total number of groups equal to a number denoted as G with each group consisting of n_i number of individuals is given by:

$$L_{full}\left(t_i, h_0, \boldsymbol{\beta}, \theta\right) = \prod_{i=1}^{G}\prod_{j=1}^{n_i}\left(z_i\right)^{\delta_i} h_0\left(t_{ij}\right)^{\delta_i} \exp\left(\delta_i \boldsymbol{X}_{ij}^T \boldsymbol{\beta}\right) \exp\left(-z_i H_0\left(t_{ij}\right)\right) f\left(z_i\right).$$

The frailties are assumed to follow a Gamma distribution, the full likelihood is therefore given by:

$$L_{full}\left(t_i, h_0, \boldsymbol{\beta}, \theta\right) = \prod_{i=1}^{G}\prod_{j=1}^{n_i}\left(z_i\right)^{\delta_i} h_0\left(t_{ij}\right)^{\delta_i} \exp\left(\delta_i \boldsymbol{X}_{ij}^T \boldsymbol{\beta}\right) \exp\left(-z_i H_0\left(t_{ij}\right)\right) \prod_{i=1}^{G} \frac{z_i^{\frac{1}{\theta}+D_i-1} \exp\left(\frac{-z_i}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right)\theta^{\frac{1}{\theta}}},$$

where D_i is the total number of events in the cluster i. With the EM algorithm one needs to find initial estimates for $\boldsymbol{\beta}, H_0(t_{ij})$ and θ ($\hat{\boldsymbol{\beta}}$, $\hat{H}_0(t_{ij})$ and $\hat{\theta}$ respectively). To get the initial estimates $\hat{\boldsymbol{\beta}}$ and $\hat{H}_0(t_{ij})$ we use the model which has no frailties, this is equivalent to letting $\theta = 0$ (Phipson and Mwambi, 2010). Use the estimates $\hat{\boldsymbol{\beta}}$ and $\hat{H}_0(t_{ij})$ obtained together with $\hat{\boldsymbol{\beta}} = \mathbf{0}$ in the expectation step (E-step) to get the expected values of the frailty terms (z_i^s).

Hanagal (2011), suggests that the expectation for the frailties given the observed data as was calculated by (Parner 1997) and it is given by the general formula:

$$E(z_i) = -\frac{L^{D_i+1}\left(\sum_j \hat{H}_0(t_i j)\exp\left(-\boldsymbol{X}_{ij}^T\hat{\boldsymbol{\beta}}\right)\right)}{L^{D_i}\left(\sum_j \hat{H}_0(t_i j)\exp\left(-\boldsymbol{X}_{ij}^T\hat{\boldsymbol{\beta}}\right)\right)}, \quad i = 1, 2, \ldots, G,$$

where L^{D_i+1} denotes the $\{D_i+1\}^{th}$ derivative of the Laplace transform of $L^{D_i+1}\left(\sum_j \hat{H}_0(t_i j)\, e^{-\boldsymbol{X}_{ij}^T\hat{\boldsymbol{\beta}}}\right)$.

For the Gamma frailty model there exists a short cut to derive $E(z_i)$ and $E(\ln(z_i))$ based on the full likelihood (Hanagal, 2011, Small and Wang, 2003). They argue that the distribution of the frailty terms z_i is a Gamma with the shape and scale parameters $\widehat{\alpha} = \frac{1}{\theta} + D_i$ and $\widehat{\theta} = \frac{1}{\theta} + \sum_j H_0(t_{ij})\exp\left(\boldsymbol{X}_{ij}^T\boldsymbol{\beta}\right)$. The expected value of the frailties is therefore given by:

$$E(z_i) = \frac{\widehat{\alpha}_i}{\widehat{\theta}_i},$$
$$E(\ln(z_i)) = \psi(\widehat{\alpha}_i) - \ln\left(\widehat{\theta}_i\right).$$

$\psi(.)$ represents a Di-gamma function given by:

$$\psi = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}.$$

At the M-step, we plug in the expected values of the frailty terms into the modified partial likelihood to obtain the estimates of $\boldsymbol{\beta}$ and h_0. The modified partial likelihood is given by:

$$L(\boldsymbol{\beta}) = \prod_{k=1}^{M} \frac{\exp(\hat{z}\boldsymbol{\beta}s_k)}{\left(\sum_{l\in R(t_k)} z_l \exp\left(\boldsymbol{X}_l^T\boldsymbol{\beta}\right)\right)^{d_k}}.$$

t_k is the smallest failure time, d_k is the number of failures at time t_k, D_k is the set of all individuals who fail at time t_k and $s_k = \sum_{j\in D_k} x_j$. The maximum likelihood estimate

for the baseline hazard function is obtained from the expression below:

$$\hat{h_{0k}} = \frac{d_k}{\sum_{l \in R(t_k)} \hat{z}_l \exp\left(\hat{\boldsymbol{X}_l^T \boldsymbol{\beta}}\right)}, \quad k = 1, 2, \ldots, M.$$

Finally we Plug in the estimates $\hat{\boldsymbol{\beta}}, \hat{h_0}$ and $\hat{z}_i$ in the full likelihood to obtain the estimates of θ.

In order to determine the level of precision (Standard errors) of the estimates from the true parameters we have to derive the information matrix, I. The covariance matrix is the inverse of the information matrix.

5.2.3 The penalised partial likelihood approach

The EM-algorithm is slow and in order to get the variance estimates, it requires further computations. The penalised partial likelihood (PPL) is an alternative approach. In this approach, the frailty terms are treated as additional regression coefficients which are constrained by a penalty function added to the log-likelihood. The PPL and the EM give the same parameter estimates for a Gamma frailty model (Duchateau and Janssen, 2007, Hanagal, 2011, Phipson and Mwambi, 2010).

The penalised partial likelihood approach uses the random effects u_i instead of the frailty terms $z_i = \exp(u_i)$.

Assuming a univariate frailty model, the hazard function and the survival function of an individual is given by:

$$\begin{aligned} h_i(t_i) &= h_0(t_i) \exp\left(\boldsymbol{X}_i^T \boldsymbol{\beta} + \boldsymbol{u}_i\right), \\ S_i(t_i) &= \exp\left(-H_0(t_i) \exp \boldsymbol{X}_i^T \boldsymbol{\beta} + \boldsymbol{u}_i\right). \end{aligned}$$

The full likelihood is given as:

$$L_{ifull}(\boldsymbol{u}_i, \theta, \boldsymbol{\beta}) = f_i(u_i) \left[h_i(t_i)\, s_i(t_i)\right]^{\delta_i} \left[s_i(t_i)\right]^{1-\delta_i},$$

where θ is the variance of the $\boldsymbol{u}_i^{'}s$.

The log of the full likelihood is given by:

$$l_{ifull}(\boldsymbol{u}_i, \theta, \boldsymbol{\beta}) = \ln f(u_i) + \delta_i \ln h_0(t_i) + \delta_i \boldsymbol{X}_i^T \boldsymbol{\beta} + \delta_i u_i - H_0(t_i) \exp \boldsymbol{X}_i^T \boldsymbol{\beta} + \boldsymbol{u}_i. \quad (5.9)$$

The full likelihood can be written in two parts (Duchateau and Janssen, 2008). It is argued that the first part consists of the conditional likelihood of the data given the frailties and the second part corresponds to the distribution of the frailties.

$$l_{ifull}\left(\boldsymbol{u}_i, \theta, \boldsymbol{\beta}\right) = l_{part1i}\left(\boldsymbol{\beta}, \boldsymbol{u}_i\right) + l_{part2i}\left(\theta, \boldsymbol{u}_i\right). \tag{5.10}$$

From equation (5.10), $l_{part2i}\left(\theta, u_i\right)$ is called the penalty term. Therefore from the full likelihood in equation (5.9),

$$\begin{aligned} l_{part1i}\left(\boldsymbol{\beta}, \boldsymbol{u}_i\right) &= \delta_i \ln h_0\left(t_i\right) + \boldsymbol{X}_i^T\boldsymbol{\beta} + u_i - H_0\left(t_i\right)\exp\left(\boldsymbol{X}_i^T\boldsymbol{\beta} + \boldsymbol{u}_i\right), \\ l_{part2i}\left(\theta, \boldsymbol{u}_i\right) &= \ln f\left(u_i\right). \end{aligned}$$

When the actual value of the random effect is far away from its mean, the logarithm of the absolute value of the probability density function at that value of the random effect ($\ln\left|f\left(u_i\right)\right|$) takes on a large value and this implies that the penalty term has a large negative contribution to the likelihood.

The penalised partial likelihood is given as:

$$l_{ppl}\left(\boldsymbol{\beta}, \theta, \boldsymbol{u}_i\right) = l_{part}\left(\boldsymbol{\beta}, u_i\right) - l_{pen}\left(\theta, \boldsymbol{u}_i\right).$$

$$\begin{aligned} l_{peni}\left(\theta\right) &= -\ln f\left(u_i\right), \\ l_{pen}\left(\theta\right) &= -\sum_{i=1}^{n} \ln f\left(u_i\right). \end{aligned}$$

If we let $\boldsymbol{\eta}_i = \boldsymbol{X}_i^T + \boldsymbol{u}_i$, then

$$l_{part}\left(\boldsymbol{\beta}, \boldsymbol{u}_i\right) = \sum_{i=1}^{n} \delta_i \left[\boldsymbol{\eta}_i - \ln\left(\sum_{qw \in R(t_i)} \exp\left(\boldsymbol{\eta}_{qw}\right)\right)\right],$$

where $R\left(t_i\right)$ is the risk set at time t_i. Thus with all the contributing terms defined l_{ppl} gives the penalised partial likelihood used to draw inference on the parameters θ and $\boldsymbol{\beta}$.

5.2.4 Implementation of the penalised partial likelihood on a Normal random effects density

The random effects are assumed to be normally distributed $u_i \sim N(0, \theta)$. This implies that the frailties denoted as $z_i = \exp(u_i)$ is log-normally distributed.

The probability density function of a normal distribution whose random variable is the random effects u_i is given by:

$$f(u_i, \theta) = \frac{\exp\left(\frac{-u_i^2}{\theta}\right)}{\theta^{\frac{1}{2}}\sqrt{2\pi}}, \theta > 0,$$

where θ is the variance of the random effects.

$$\begin{aligned} l_{pen} &= -\frac{1}{2}\sum_{i=1}^{n} \ln f(u_i, \theta) \\ &= \frac{1}{2}\sum_{i=1}^{n}\left(\frac{u_i^2}{\theta} + \ln 2\pi\theta\right). \end{aligned}$$

The maximisation of the penalised partial likelihood consists of an outer loop and an inner loop. The Newton-Raphson method is used in the inner loop to maximise the penalised partial likelihood ($l_{ppl}(\boldsymbol{\beta}, \theta, \boldsymbol{u})$) to obtain the best linear unbiased predictors (BLUPS) of $\boldsymbol{\beta}$ and $\boldsymbol{u}$ given a provisional value of θ. In the outer loop: After fixing $\boldsymbol{\beta}$ and $\boldsymbol{u}$ at the best linear unbiased predictors (BLUPS), the estimate of θ is obtained by maximising $l(\boldsymbol{\beta}, \boldsymbol{u})$ which is similar to the log-likelihood maximised in the EM. The procedure is repeated until convergence is realised (Duchateau and Janssen, 2008, Phipson and Mwambi, 2010).

Procedure:

Let s denote the outer loop and k the inner loop index. Let θ^s denote the estimate of θ at the s^{th} iteration in the outer loop.

Suppose $\beta^{(s,k)}$ and $u^{(s,k)}$ are the estimates and predictors for $\boldsymbol{\beta}$ and u at the k^{th} step in the inner loop given θ^s. Starting with the initial values $\boldsymbol{u}^{(1,0)}, \boldsymbol{\beta}^{(1,0)}, \theta^{(0)}$ and θ^1, the k^{th} iterative step for the Newton-Raphson method given $\theta^{(1)}$ is given by:

$$\begin{bmatrix} \boldsymbol{\beta}^{(s,k)} \\ \boldsymbol{u}^{(s,k)}, \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}^{(s,k-1)} \\ \boldsymbol{u}^{(s,k-1)} \end{bmatrix} - V \begin{bmatrix} 0 \\ \left(\theta^{(s)}\right)^{-1} \boldsymbol{u}^{(s,k-1)} \end{bmatrix} + \boldsymbol{V}\left[\boldsymbol{XZ}\right] \frac{dl_{part}(\boldsymbol{\beta}, \boldsymbol{u})}{d\eta},$$

where $\boldsymbol{V} = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}$ is the inverse of a square $(p+a)$ dimensional matrix $\boldsymbol{B}$, with $\boldsymbol{B}$ given by:

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^T \\ \boldsymbol{Z}^T \end{bmatrix} \left(-\frac{\partial^2 l_{apart}\left(\boldsymbol{\beta}, u\right)}{\partial \eta \partial \eta} \right) [\boldsymbol{X}, \boldsymbol{Z}] + \begin{bmatrix} 0 & 0 \\ 0 & \left(\theta^{(s)}\right)^{-1} I_G. \end{bmatrix}$$

Once the Newton-Raphson procedure has converged for the current value of $\theta^{(s)}$, a restricted maximum likelihood (REML) estimate for θ is given by:

$$\theta^{(s+1)} = \frac{\sum \left(u_i^{(s,k)}\right)^2}{n-r},$$

where $r = \frac{trace(v_{22})}{\theta^{(s)}}$.

The outer loop is iterated until the absolute difference between two sequential values for θ;

$$|\theta^{(s)} - \theta^{(s-1)}| \quad \text{is sufficiently small.}$$

The asymptotic variance for $\hat{\boldsymbol{\beta}}$ is given by v_{11}. The asymptotic variance for $\hat{\theta}$ is generally given by:

$$\frac{2\theta^2}{n - 2r + \theta^{-2} trace\left(v_{22}^2\right)}.$$

Detailed information about the Penalised partial likelihood and further understanding of the penalised partial likelihood approach we recommend a book by Duchateau and Janssen (2008).

5.2.5 The markov chain monte carlo methods

The above mentioned methods of estimating parameters (The EM algorithm and the Penalised Partial likelihood) are mainly used when the survival data is right censored. With complicated forms of censoring like interval and left censoring, more advanced

methods have to be used to estimate the parameters of the model. Markov Chain Monte Carlo methods (MCMC) are the alternative methods that can be used to estimate parameters of parametric frailty models in circumstances where there exists left and interval censored data points in the data set.

The MCMC methods due to Clayton (1991) are statistical techniques (Jones, 2004). These methods are statistical simulation techniques. Under these methods a process is directly simulated given the probability density functions that describe it, instead of writing down complex system of equations (such as a system of differential equations) that describe the behaviour of the system which may increase the complexity. Given the probability density functions (p.d.f's) together with its initial defining parameters, the simulation process begins by iteratively re sampling each parameter given the current values of the other parameters.

They are known as Markov Chain Monte Carlo Methods because one uses the previous sample values to generate randomly the next sample values which results into a Markov chain.

The Markov chain is a sequence of random variables $\{X_0, X_1, \ldots\}$ with a property that; Given the present state, the future and past sates are independent. Stated mathematically this means

$$P\left(X_{n+1} = x | X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\right) = P\left(X_{n+1} = x | X_n = x_n\right).$$

This is known as the Markov property.

Suppose we have a sequence of random variables $\{X_0, X_1, \ldots\}$with time as the index set which are generated in such a way that at each time $t \geq 0$, the next state X_{t+1} depends only on the current state of the chain. That is to say X_{t+1} is sampled from $P\left(X_{t+1}|X_t\right)$ and the state X_{t+1} does not depend on the rest of the chain $\{X_0, X_1, \ldots, X_{t-1}\}$, the random variables generated in such a sequence result in to what is known as a Markov Chain (Gilks et al., 1996).

Jones (2004) argues that it is important to visualise each parameter as a node in a given network, where by one visits the node and resamples its value by picking randomly a value from its conditional distribution given the current values of the other parameters. This process results into a Markov chain which when it converges after a sufficiently long burn-in period of say k iterations, gives a sample from the full posterior distribution of any parameter interest (Gilks et al. (1996)). Jones (2004) mentions that from a theoretical point of view, the MCMC implies the Bayesian approach. However other

schemes based on MCMC like the Gibbs sampling and the Metropolis-Hastings can be used to fit frailty models on clustered failure data.

Clayton (1991), Jones (2004), used Gibbs sampling to fit frailty models on clustered failure time data with right censored observations, by sampling iteratively from the full conditional distribution of the parameters in the model.

5.3 The results from fitting the shared gamma frailty model on 2011 Uganda DHS data correcting for both community and household effects

The software used for fitting the Shared Gamma Frailty model to determine the parameter estimates using a semi-parametric Penalized Likelihood on the hazard function was R, using the library called *frailtypack* with a command "*frailtyPenal*". There are 404 communities and 4285 households presented in the data set.

Two models were fitted, one was the CPH model with community effects as the frailty term and a CPH model with household effects as the frailty term. The fixed effects (covariates) used in both models were those considered to obey the proportional hazard assumption which were discussed in chapter (5.2). The results of these two models are presented in the tables below;

[1]

TABLE 5.1: The results from fitting a shared frailty (community effect and household effects) model on 2011 Uganda DHS data to determine factors affecting under-five children survival in Uganda.

Correcting for household effects				**Correcting for Community effects**			
variable	**HR (SE $\hat{\beta}$)**	$[95\%CI]$	**pvalues**	**variable**	**HR (SE $\hat{\beta}$)**	$[95\%CI]$	**pvalues**
Sex of the child				**Sex of the child**			
Male	**1**	...		**Male**	**1**	...	
Female	**0.83 (0.09)**	[0.68, 1.00]	**0.05**	**Female**	**0.84 (0.09)**	[0.70, 1.00]	**0.06**
Father's education				**Father's education**			
Illiterate	**1**	...	...	**Illiterate**	**1**	...	...
Primary	**0.95 (0.11)**	[0.77, 1.18]	**0.65**	**Primary**	**0.95 (0.10)**	[0.78, 1.16]	**0.60**
Secondary and Higher	**0.73 (0.19)**	[0.50, 1.06]	**0.09**	**Secondary and Higher**	**0.74(0.18)**	[0.53, 1.05]	**0.09**
Age at first birth				**Age at first birth**			
Below 20	**1**	...		**Below 20**	**1**	...	
20-29	**0.84 (0.12)**	[0.67, 1.07]	**0.15**	**20-29**	**.86 (0.11)**	[0.70, 1.06]	**0.17**
30-39	**1.89 (0.69)**	[0.49, 7.32]	**0.35**	**30-39**	**1.66 (0.58)**	[0.52, 5.28]	**0.39**
Number of births in past 1year				**Number of births in past 1year**			
No birth	**1**	**-**		**No birth**	**1**	**-**	
One birth	**1.25 ((0.11)**	[1.01, 1.55]	**0.04**	**One birth**	**1.25 (0.11)**	[1.03, 1.52]	**0.03**
Two births	**4.57 (0.62)**	[1.36, 15.32]	**0.014**	**Two births**	**2.76 (0.46)**	[1.11, 6.85]	**0.03**
Sex of household head				**Sex of household head**			
Male	**1**	...		**Male**	**1**	...	
Female	**1.39 (0.11)**	[1.11, 1.74]	**0.00**	**Female**	**1.36 (0.10)**	[1.11, 1.66]	**0.00**
Mother's age group				**Mother's age group**			
Below 20 years	**1**	...		**Below 20 years**	**1**	...	
20-29 years	**0.89 (0.2)**	[0.59, 1.33]	**0.56**	**20-29 years**	**0.84 (0.19)**	[0.57, 1.22]	**0.36**
30-39	**1.00 (0.22)**	[0.65, 1.52]	**0.98**	**30-39**	**0.94 (0.12)**	[0.63, 1.39]	**0.74**
40years +	**1.28 (0.26)**	[0.77, 2.12]	**0.34**	**40years +**	**1.18 (0.23)**	[0.74, 1.88]	**0.50**
Source of drinking water				**Source of drinking water**			
Piped water	**1**	...		**Piped water**	**1**	...	
Borehole	**1.23 (0.15)**	[0.92, 1.65]	**0.16**	**Borehole**	**1.21 (0.14)**	[0.92, 1.61]	**0.16**
Well	**1.16 (0.17)**	[0.82, 1.62]	**0.39**	**Well**	**1.15 (0.16)**	[0.84, 1.59]	**0.37**
Surface/Rain/Lake	**1.38 (0.19)**	[0.95, 2.00]	**0.09**	**Surface/Rain/Lake**	**1.36 (0.17)**	[0.96, 1.92]	**0.08**
Others	**1.34 (0.27)**	[0.79, 2.28]	**0.28**	**Others**	**1.30 (0.17)**	[0.80, 2.11]	**0.28**
Religion				**Religion**			
Catholic	**1**	...		**Catholic**	**1**	...	
Muslim	**1.05 (0.16)**	[0.78, 1.43]	**0.74**	**Muslim**	**1.05 (0.14)**	[0.79, 1.39]	**0.75**
Other Christians	**0.95 (0.11)**	[0.76, 1.18]	**0.61**	**Other Christians**	**0.96 (0.10)**	[0.79, 1.18]	**0.71**
Others	**0.64 (0.54)**	[0.22, 1.86]	**0.41**	**Others**	**0.69 (0.51)**	[0.25, 1.89]	**0.45**
Household Frailty parameter (Variance)	**θ = 1.78 (0.43)**			**Community Frailty parameter (Variance)**	**θ = 0.12(0.07)**		
Penalised Marginal loglikelihood	$= -3025.98$			**Penalised marginal loglikelihood**	$= -3042.18$		

[1]Level of significance $**$ < 0.05, The Likelihood-ratio test of $\theta = 0$, $\chi = 2.63$ and p-value of $= 0.052$ from the STATA software for the community frailty term.

Results

By using the likelihood ratio test with a null hypothesis that the variance of the frailty term is zero ($\theta = 0$), the chi-square test statistic ($\chi = 2.63$) with a p-value of 0.052. At 0.05 level of significance, it implies that there is not enough evidence to show the existence of unobserved heterogeneity at community level. In this case therefore one can use the standard cox PH model because the results suggest that there is no difference on the conclusions that would be drawn about the data set.

The factors that were strongly associated with high under-five child mortality rates were identified by looking at their confidence intervals and the p-values. The factors whose confidence intervals contained a 1 (HR=1), implied that these factors are not significant and this results is confirmed p-values greater than 0.05 level of significance. Number of births in the past one year and the sex of the household head were found to be strongly associated to high mortality rates. The children whose mothers had more than one birth in the past one year were at a higher risk of death before reaching the age of five than those whose mothers had no birth at all. The children born in households headed by women were at a high risk of death than those born in households where the man is the head.

On comparing the estimates of the hazard ratios of the standard cox-proportional hazard model and those of the community frailty model, there was an observed increase in the estimates after correcting for community frailty as shown in table (5.2).

In the case of household or family frailty, there were 4285 households in the sample considered for analysis. The variance of the frailty term (household frailty) $\theta = 1.78$, which is different from zero, this gives evidence of the existence of the unobserved heterogeneity at family or household level. The results suggests that some households were associated to a higher risk of children dying before reaching the age of five than the others.

Summary

We observe that although the confidence intervals are wider, they are close to those under the CPH when accounting for community effects. When accounting for the household effects, the confidence intervals are even wider which implies that there is much more

TABLE 5.2: Comparing the results from the standard cox-proportional hazard model and the results of the model after correcting for community effects.

Fitting the standard Cox-ph model				**Correcting for Community effects**			
Variable	**Adjusted HR**	**HR**$[95\%CI]$	**p-values**	**variable**	**HR (SE $\hat{\beta}$)**	$[95\%CI]$	**p-values**
Sex of the child				**Sex of the child**			
Male	**1**	...		**Male**	**1**	...	
Female	**0.83**	$[0.69, 0.99]$	**0.04**	**Female**	**0.84 (0.09)**	$[0.70, 1.00]$	**0.06**
Father's education				**Father's education**			
Illiterate	**1**	...		**Illiterate**	**1**	...	...
Primary	**0.92**	$[0.76, 1.12]$	**0.43**	**Primary**	**0.95 (0.10)**	$[0.78, 1.16]$	**0.60**
Secondary and higher	**0.72**	$[0.51, 1.01]$	**0.06**	**Secondary and Higher**	**0.74(0.18)**	$[0.53, 1.05]$	**0.09**
Age at first birth				**Age at first birth**			
< 20	**1**	...		**Below 20**	**1**	...	
20-29	**0.86**	$[0.69, 1.06]$	**0.16**	**20-29**	**.86 (0.11)**	$[0.70, 1.06]$	**0.17**
30^{+}	**1.59**	$[0.51, 5.02]$	**0.42**	**30-39**	**1.66 (0.58)**	$[0.52, 5.28]$	**0.39**
Number of births in the past one year				**Number of births in past 1year**			
No birth	**1**	...		**No birth**	**1**	**-**	
1 birth	**1.22**	$[1.01, 1.48]$	**0.04**	**One birth**	**1.25 (0.11)**	$[1.03, 1.52]$	**0.03**
2	**2.57**	$[1.06, 6.25]$	**0.04**	**Two births**	**2.76 (0.46)**	$[1.11, 6.85]$	**0.03**
Sex of household head				**Sex of household head**			
Male	**1**	...		**Male**	**1**	...	
Female	**1.33**	$[1.09, 1.63]$	**0.01**	**Female**	**1.36 (0.10)**	$[1.11, 1.66]$	**0.00**
Mother's age group				**Mother's age group**			
< 20	**1**	...		**Below 20 years**	**1**	...	
$20 - 29$**yrs**	**0.71**	$[0.48, 1.05]$	**0.08**	**20-29 years**	**0.84 (0.19)**	$[0.57, 1.22]$	**0.36**
$30 - 39$**yrs**	**0.79**	$[0.53, 1.19]$	**0.27**	**30-39**	**0.94 (0.12)**	$[0.63, 1.39]$	**0.74**
40+	**0.99**	$[0.62, 1.59]$	**0.98**	**40years +**	**1.18 (0.23)**	$[0.74, 1.88]$	**0.50**
Source of drinking water				**Source of drinking water**			
Piped water	**1**	...		**Piped water**	**1**	...	
Borehole	**1.12**	$[0.86, 1.48]$	**0.39**	**Borehole**	**1.21 (0.14)**	$[0.92, 1.61]$	**0.16**
Well water	**1.06**	$[0.78, 1.45]$	**0.69**	**Well**	**1.15 (0.16)**	$[0.84, 1.59]$	**0.37**
Surface/pond/ lake/Rain/etc	**1.28**	$[0.91, 1.79]$	**0.15**	**Surface/Rain /Lake**	**1.36 (0.17)**	$[0.96, 1.92]$	**0.08**
Others	**1.21**	$[0.75, 1.94]$	**0.44**	**Others**	**1.30 (0.17)**	$[0.80, 2.11]$	**0.28**
Religion				**Religion**			
Catholic	**1**	...		**Catholic**	**1**	...	
Muslim	**1.02**	$[0.77, 1.34]$	**0.91**	**Muslim**	**1.05 (0.14)**	$[0.79, 1.39]$	**0.75**
Other Christians	**0.94**	$[0.77, 1.14]$	**0.51**	**Other Christians**	**0.96 (0.10)**	$[0.79, 1.18]$	**0.71**
Others	**0.67**	$[0.25, 1.81]$	**0.43**	**Others**	**0.69 (0.51)**	$[0.25, 1.89]$	**0.45**
Log-Likelihood	=	**-4128.68**		**Frailty parameter**	θ **= 0.12(0.07)**		
				Penalised marginal log-likelihood	$= -3042.18$		

heterogeneity at the household level than at the community level. This is something expected. Overall at community level, one might expect homogeneity but higher variability between households.

The estimate of the hazard ratios of the sex of the child remained unchanged after correcting for household effects. Number of births in the past one year and the sex of the household head were strongly associated to under-five infant mortality in Uganda even after correcting for household and community effects. Other factors included in the analysis like sex of the child, father's education, age at first birth, mother's age group, source of drinking, and religion have an effect on the survival time of the children but there was not enough evidence from the data to confirm their strong association to this survival time.

TABLE 5.3: Comparison of results from the standard cox PH model and the results after correcting for household effects.

Fitting the standard Cox-ph model				**Correcting for household effects**			
Variable	**Adjusted HR**	**HR**$[95\%CI]$	**p-values**	**variable**	**HR (SE $\hat{\beta}$)**	$[95\%CI]$	**p-values**
Sex of the child				**Sex of the child**			
Male	**1**	...		**Male**	**1**	...	
Female	**0.83**	$[0.69, 0.99]$	**0.04**	**Female**	**0.83 (0.09)**	$[0.68, 1.00]$	**0.05**
Father's education				**Father's education**			
Illiterate	**1**	...		**Illiterate**	**1**	...	...
Primary	**0.92**	$[0.76, 1.12]$	**0.43**	**Primary**	**0.95 (0.11)**	$[0.77, 1.18]$	**0.65**
Secondary and higher	**0.72**	$[0.51, 1.01]$	**0.06**	**Secondary and Higher**	**0.73 (0.19)**	$[0.50, 1.06]$	**0.09**
Age at first birth				**Age at first birth**			
< 20	**1**	...		**Below 20**	**1**	...	
20-29	**0.86**	$[0.69, 1.06]$	**0.16**	**20-29**	**0.84 (0.12)**	$[0.67, 1.07]$	**0.15**
30^+	**1.59**	$[0.51, 5.02]$	**0.42**	**30-39**	**1.89 (0.69)**	$[0.49, 7.32]$	**0.35**
Number of births in the past one year				**Number of births in past 1year**			
No birth	**1**	...		**No birth**	**1**	**-**	
1 birth	**1.22**	$[1.01, 1.48]$	**0.04**	**One birth**	**1.25 ((0.11)**	$[1.01, 1.55]$	**0.04**
2	**2.57**	$[1.06, 6.25]$	**0.04**	**Two births**	**4.57 (0.62)**	$[1.36, 15.32]$	**0.014**
Sex of household head				**Sex of household head**			
Male	**1**	...		**Male**	**1**	...	
Female	**1.33**	$[1.09, 1.63]$	**0.01**	**Female**	**1.39 (0.11)**	$[1.11, 1.74]$	**0.00**
Mother's age group				**Mother's age group**			
< 20	**1**	...		**Below 20 years**	**1**	...	
$20 - 29$**yrs**	**0.71**	$[0.48, 1.05]$	**0.08**	**20-29 years**	**0.89 (0.2)**	$[0.59, 1.33]$	**0.56**
$30 - 39$**yrs**	**0.79**	$[0.53, 1.19]$	**0.27**	**30-39**	**1.00 (0.22)**	$[0.65, 1.52]$	**0.98**
40+	**0.99**	$[0.62, 1.59]$	**0.98**	**40years +**	**1.28 (0.26)**	$[0.77, 2.12]$	**0.34**
Source of drinking water				**Source of drinking water**			
Piped water	**1**	...		**Piped water**	**1**	...	
Borehole	**1.12**	$[0.86, 1.48]$	**0.39**	**Borehole**	**1.23 (0.15)**	$[0.92, 1.65]$	**0.16**
Well water	**1.06**	$[0.78, 1.45]$	**0.69**	**Well**	**1.16 (0.17)**	$[0.82, 1.62]$	**0.39**
Surface/pond/ lake/Rain/etc	**1.28**	$[0.91, 1.79]$	**0.15**	**Surface/Rain/ Lake**	**1.38 (0.19)**	$[0.95, 2.00]$	**0.09**
Others	**1.21**	$[0.75, 1.94]$	**0.44**	**Others**	**1.34 (0.27)**	$[0.79, 2.28]$	**0.28**
Religion				**Religion**			
Catholic	**1**	...		**Catholic**	**1**	...	
Muslim	**1.02**	$[0.77, 1.34]$	**0.91**	**Muslim**	**1.05 (0.16)**	$[0.78, 1.43]$	**0.74**
Other Christians	**0.94**	$[0.77, 1.14]$	**0.51**	**Other Christians**	**0.95 (0.11)**	$[0.76, 1.18]$	**0.61**
Others	**0.67**	$[0.25, 1.81]$	**0.43**	**Others**	**0.64 (0.54)**	$[0.22, 1.86]$	**0.41**
				Frailty parameter	θ **= 1.78 (0.43)**		
Log-Likelihood	=	**-4128.68**		**Penalised Marginal log-likelihood**	$= -3025.98$		

Chapter 6

The Bayesian approach

Bayesian inference provides a unified approach to statistical modelling and inference. The Markov chain Monte Carlo (MCMC) sampling makes the Bayesian inference possible by enabling parameter estimation and quantifying of their uncertainties (Akerkar et al., 2010). As stated by Martin and Quinn (2002) Bayesian approaches have become increasingly used in science research since the early 1990's .

The highly specialised tools for doing Bayesian analysis of survival models is WinBUGS. WinBuGS is a software package for Bayesian analysis of complex statistical models using MCMC methods. Any other Markov scheme can be used instead of WinBUGS and the results will be the same. The simulations based inference are possible but with significant computing time constraints such as, updating schemes may produce slow converging samplers which can underestimate the variability of the random variables being estimated. The computations can also become too expensive because of the Monte Carlo error. There is a new tool that allows the user to easily perform approximate Bayesian inference using the integrated nested Laplace approximations introduced by Håvard and Nicolas (2009). This tool provides a deterministic alternative to MCMC which is a standard tool for inference in Latent Gaussian Models. The tool is called INLA, it computes posterior marginals for each component in the model from which posterior expectations and standard deviations can easily be found (Akerkar et al., 2010). In this chapter we present the basic concepts and origin of Bayesian analysis and present results obtained using the INLA package in R. A Weibull model (assuming that the time to death of the children under-five follows a Weibull distribution) was fitted with and without family and community effects to draw conclusions on factors affecting under-five child mortality rate in Uganda using the Uganda demographic health survey data for 2011.

6.1 The Bayesian inference

Bayesian inference is a method of inference in which Baye's theorem is used. Given that η represents the vector of the unknown parameters and X represents the data,

$$P\left(\boldsymbol{\eta}|\boldsymbol{X}\right) = \frac{P\left(\boldsymbol{\eta},\boldsymbol{X}\right)}{P\left(\boldsymbol{X}\right)}, \tag{6.1}$$

where $\boldsymbol{\eta} = \left(\boldsymbol{h}^T, \theta, \boldsymbol{\beta}^T, \boldsymbol{z}^T\right)^T$, $P\left(\boldsymbol{\eta}|\boldsymbol{X}\right)$ denotes the posterior density, $P\left(\boldsymbol{\eta},\boldsymbol{X}\right)$ the joint probability density of the data and the unknown parameters and $P\left(\boldsymbol{X}\right)$ denotes the marginal probability density. $P\left(\boldsymbol{X}\right) = \sum_{\boldsymbol{\eta}} P\left(\boldsymbol{\eta}\right) P\left(\boldsymbol{X}|\boldsymbol{\eta}\right)$ for a discrete case and $P\left(\boldsymbol{X}\right) = \int_{\boldsymbol{\eta}} P\left(\boldsymbol{\eta}\right) P\left(\boldsymbol{X}|\boldsymbol{\eta}\right) d\eta$ in a continuous case.

The joint probability density $P\left(\boldsymbol{\eta},\boldsymbol{X}\right)$ can be written as a product of the prior and the sampling distribution. The posterior distribution therefore becomes:

$$P\left(\boldsymbol{\eta}|\boldsymbol{X}\right) = \frac{P\left(\boldsymbol{\eta}\right) P\left(\boldsymbol{X}|\boldsymbol{\eta}\right)}{P\left(\boldsymbol{X}\right)}.$$

The Bayesian approach is based on specifying the probability model of the observed data $\boldsymbol{X}$ given the vector of the unknown parameter $\boldsymbol{\eta}$ leading to the likelihood function $L\left(\boldsymbol{\eta}|\boldsymbol{X}\right)$. With the Bayesian approach, $\boldsymbol{\eta}$ is assumed to be random and has got a prior distribution denoted as $f\left(\boldsymbol{\eta}\right)$. To draw inference on $\boldsymbol{\eta}$, we base analysis on the posterior distribution denoted as $f\left(\boldsymbol{\eta}|\boldsymbol{X}\right)$ (Chen et al., 2000).

From equation (6.1), the posterior distribution can be obtained as follows:

$$f\left(\eta|X\right) = \frac{L\left(\boldsymbol{\eta}|\boldsymbol{X}\right) f\left(\eta\right)}{\int_{\boldsymbol{\Psi}} L\left(\boldsymbol{\eta}|\boldsymbol{X}\right) f\left(\eta\right) d\eta}, \tag{6.2}$$

where $\boldsymbol{\Psi}$ is denotes the parameter space. From equation (6.2) we can conclude that posterior distribution $f\left(\eta|X\right)$ is proportional to the product of the likelihood and the prior distribution of the parameter. That is

$$f\left(\eta|X\right) \propto L\left(\boldsymbol{\eta}|\boldsymbol{X}\right) f\left(\eta\right). \tag{6.3}$$

Let $M(X) = \int_{\boldsymbol{\Psi}} L(\boldsymbol{\eta}|\boldsymbol{X}) f(\eta) d\eta$; $M(X)$ is called the normalising constant and it is generally known as the marginal distribution of the data.

Equation (6.3) implies that the posterior distribution has got a contribution from the observed data through the likelihood and a contribution from the prior information about the parameter.

According to Chen et al. (2000), the normalising constant does not have an analytical closed form which implies that the posterior distribution has no closed form. They suggested that in-order to sample from a multivariate distribution (the posterior distribution $f(\eta|X)$), one needs to use computational methods. The most commonly used computational method for sampling from the posterior distribution is called the Gibbs Sampler.

In this chapter we fitted the Weibull model with and without frailty effects and the Cox-proportional hazard model on the UDHS 2011 data set. Variables that were selected for analysis in the Cox-proportional hazard model are the same variables considered in the analyses using Bayesian inference.

6.1.1 The Weibull model

The Weibull distribution is the most commonly used distribution to model time event data because of its flexibility in specifying the hazard function. The Weibull model therefore is the most widely used parametric model. The Weibull distribution described in chapter two with parameters $\alpha > 0$ and $\lambda > 0$, has a density function for t_i which can also take on the following form:

$$f(t_i|\alpha,\lambda) = \alpha t_i^{\alpha-1} \exp - [\lambda t_i]^{\alpha}, \quad 0 < t_i < \infty. \tag{6.4}$$

The survival function is given by $S(t_i|\alpha,\lambda) = \exp(-\exp(\lambda) t_i^{\alpha})$. The likelihood function of the unknown parameters (α,λ) given the data can be written as:

$$L(\alpha,\lambda|\boldsymbol{D}) = \prod_{i=1}^{n} f(t_i|\alpha,\lambda) S(t_i|\alpha,\lambda)^{(1-\delta_i)}, \tag{6.5}$$

$$= \alpha^{\sum \delta_i} \exp\{\lambda \sum_{i=1}^{n} \delta_i + \sum_{i=1}^{n} (\delta_i (\alpha-1) \log(t_i) - \exp(\lambda) t_i^{\alpha})\}, \tag{6.6}$$

where δ_i is the indicator variable taking value 1 if t_i is the failure time and 0 if t_i is right censored.

To incorporate covariates in order to build a regression model, we do it through λ. We therefore write $\lambda = \boldsymbol{X}_i^T\boldsymbol{\beta}$. Where $\boldsymbol{X}_i$ and $\boldsymbol{\beta}$ are $p \times 1$ vectors of covariates and regression coefficients respectively.

Assuming Gamma prior with parameters (α_0, κ_0) for α and normal prior with parameters $\left(\mu_0, \sigma_0^2\right)$ for λ. The joint posterior distribution of (α, λ) is given by:

$$\pi\left(\alpha, \lambda|\boldsymbol{D}\right) \propto L\left(\alpha, \lambda|\boldsymbol{D}\right)\pi\left(\alpha_0, \kappa_0\right)\pi\left(\lambda|\mu_0, \sigma_0\right). \tag{6.7}$$

If we assume a normal prior $N_p\left(\mu_0, \sum_0\right)$ for $\boldsymbol{\beta}$, the joint posterior is given by:

$$\pi\left(\boldsymbol{\beta}, \alpha|\boldsymbol{D}\right) \propto \alpha^{\alpha_0+d+1}\exp\{\sum_{i=1}^{n}\left(\delta_i + \boldsymbol{X}_i^T\boldsymbol{\beta} + \delta_i\left(\alpha-1\right)\log\left(t_i\right)\right) \tag{6.8}$$

$$-t_i^{\alpha}\exp\left(\boldsymbol{X}_i^T\boldsymbol{\beta}\right) - \kappa_0\alpha - \frac{1}{2}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_0\right)\Sigma_0^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_0\right)\}, \tag{6.9}$$

where $\boldsymbol{D} = (n, t, \boldsymbol{X}, \boldsymbol{\delta})$ denotes the observed data for regression model and $\boldsymbol{X}$ is the $n \times p$ matrix of covariates with the i^{th} row as $\boldsymbol{X}_i$ and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)^T$.

6.1.2 The Weibull frailty model

Let t_{ij} be the survival time for the j^{th} individual in the i^{th} cluster , $i = 1, \ldots, n$ and $j = 1, \ldots m_i$. Here the $m_i^{'s}$ represent the number of individuals in the i^{th} cluster. We assumed that these t_{ij} are independently and identically distributed random variables. Weibull distribution such that

$$t_{ij} \sim Weibull\left(\alpha, \eta_{ij}\right), \quad \alpha > 0.$$

For frailty models the conditional hazard function of t_{ij} given the unobserved frailty z_i, a covariate vector $\boldsymbol{X}_{ij}$ and the Weibull parameter α is given by:

$$h\left(t_{ij}|\boldsymbol{X}_{ij}, z_i, \alpha\right) = \alpha t_{ij}^{\alpha-1}\exp\left[\boldsymbol{\eta}_{ij}\right], \tag{6.10}$$

where $\boldsymbol{\eta}_{ij} = \boldsymbol{\beta}_0 + \boldsymbol{X}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_i$, $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients and $\boldsymbol{\beta}_0$ is denoted intercept and $\boldsymbol{X}_{ij}$ is a $p \times 1$ covariate vector. The complete data likelihood is given by:

$$L\left(\boldsymbol{\beta}, \alpha | \boldsymbol{D}\right) = \prod_{i=1}^{n}\prod_{j=1}^{m_i}\left(\alpha t_{ij}^{\alpha-1}\exp\left(\boldsymbol{\eta}_{ij}\right)\right)^{\delta_{ij}}\exp\left(-\exp\left(\boldsymbol{\eta}_{ij}\right)t_{ij}^{\alpha}\right), \tag{6.11}$$

where δ_{ij} is the censoring indicator having a value 1 if the individual in the j^{th} cluster dies and 0 otherwise and $\boldsymbol{D} = (t, \boldsymbol{X}, \boldsymbol{\delta}, \boldsymbol{b})$ denotes the complete data set with $t = (t_{11}, \ldots, t_{nm_n})^T$, $\boldsymbol{X} = (X_{11}, \ldots, X_{nm_n})$, $\boldsymbol{\delta} = (\delta_{11}, \ldots, \delta_{nm_n})^T$ and $\boldsymbol{b} = (b_1, \ldots, b_n)$.

6.1.3 The INLA approach

Survival analysis consists of a great body of work using latent Gaussian models. Akerkar et al. (2010) recommends the work by (Berzuini and Clayton, 1994, Brezger et al., 2003, Fahrmeir et al., 1994). Latent Gaussian models are a subset of the Bayesian additive models with a structured additive predictor in this case denoted by η_i. With these models, the observation variable Y_i is assumed to belong to an exponential family where the mean μ_i is linked to this structured additive predictor η_i through a link function $(g\left(\mu_i\right))$. The additive predictor η_i accounts for effects of various covariates in an additive way:

$$\eta_i = \alpha + \sum_{j=1}^{n_f} f^{(j)}\left(u_{ji}\right) + \sum_{k=1}^{n_{\boldsymbol{\beta}}}\beta_k Z_{ki} + \epsilon_i,$$

where $\{f^j\left(.\right)\}^s$ are unknown functions of the covariates $\boldsymbol{U}$, $\boldsymbol{\beta}_k^s$ the linear effects of covariates $\boldsymbol{Z}$ and $\boldsymbol{\epsilon}_i^s$ are the unstructured terms. We assign a Gaussian prior to $\alpha, f^j\left(.\right), \beta_k$ and ϵ_i. Let us denote $\pi\left(.|.\right)$ as the conditional density of its arguments and let $\boldsymbol{X}$ denote the vector of all the n Gaussian variables $\eta_i, \alpha, f^j\left(.\right)$ and $\boldsymbol{\theta}$ denotes the vector of the hyper parameters which are not necessary Gaussian. The density $\pi\left(\boldsymbol{X}|\theta_1\right)$ is Gaussian with zero mean and precision matrix $Q\left(\theta_1\right)$ with hyper parameters θ_1.

The distribution for the n_d observational variables $Y = \{Y_i : i \in I\}$ is denoted by $\pi\left(\boldsymbol{Y}|\boldsymbol{X}, \theta_2\right)$ and we assume the $\{Y_i : i \in I\}$ are conditionally independent given $\boldsymbol{X}$ and θ_2. For simplicity, denote $\boldsymbol{\theta} = \left(\theta_1^T, \theta_2^T\right)^T$ with $dim\left(\boldsymbol{\theta}\right) = m$. The posterior is therefore written as:

$$\pi\left(\boldsymbol{X},\boldsymbol{\theta}|\boldsymbol{Y}\right) \approx \pi\left(\boldsymbol{\theta}\right)\pi\left(\boldsymbol{X}|\boldsymbol{\theta}\right),$$
$$\approx \pi\left(\boldsymbol{\theta}\right)\left|Q\left(\boldsymbol{\theta}\right)\right|^{\frac{n}{2}}\exp\left(-\frac{1}{2}\boldsymbol{X}^{T}Q\left(\boldsymbol{\theta}\right)\boldsymbol{X}+\sum_{i\in I}\log\pi\left(\boldsymbol{Y}_{i}|\boldsymbol{X}_{i},\boldsymbol{\theta}\right)\right).$$

The main aim is to approximate the posterior marginals of the latent field, $\pi\left(\boldsymbol{X}_{i}|\boldsymbol{Y}\right)$ and the posterior marginals of the hyper parameters, $\pi\left(\boldsymbol{\theta}|\boldsymbol{Y}\right)$ and $\pi\left(\theta_{j}|\boldsymbol{Y}\right)$.

According to Håvard and Nicolas (2009), INLA computes posterior marginals for each component in the model and it is from these that the posterior expectations and standard deviations can be found. The survival models can be expressed as a latent Gaussian model on which the integrated nested Laplace approximations (INLA) can be applied (Akerkar et al., 2010). The posterior marginals are given by:

$$\pi\left(\boldsymbol{X}_{i}|\boldsymbol{Y}\right) = \int \pi\left(\boldsymbol{X}_{i}|\boldsymbol{\theta},\boldsymbol{Y}\right)\pi\left(\boldsymbol{\theta}|\boldsymbol{Y}\right)d\theta, \tag{6.12}$$
$$\pi\left(\theta_{j}|\boldsymbol{Y}\right) = \int \pi\left(\boldsymbol{\theta}|\boldsymbol{Y}\right)d\theta_{-j}, \tag{6.13}$$

where $\pi\left(\boldsymbol{X}_{i}|\boldsymbol{Y}\right), \pi\left(\boldsymbol{\theta}|\boldsymbol{Y}\right)$ and $\pi\left(\theta_{j}|\boldsymbol{Y}\right)$ are the posterior marginals to be approximated by the Latent Gaussian models. We recommend Akerkar et al. (2010) for thorough reading on the Latent Gaussian models.

The forms mentioned in equation (6.12) are used to construct nested approximations and this makes the Laplace approximations very accurate when applied to latent Gaussian models.

$$\tilde{\pi}\left(\boldsymbol{X}_{i}|\boldsymbol{Y}\right) = \int \tilde{\pi}\left(\boldsymbol{X}_{i}|\boldsymbol{\theta},\boldsymbol{Y}\right)\tilde{\pi}\left(\boldsymbol{\theta}|\boldsymbol{Y}\right)d\theta, \tag{6.14}$$

$$\tilde{\pi}\left(\theta_{j}|\boldsymbol{Y}\right) = \int \tilde{\pi}\left(\boldsymbol{\theta}|\boldsymbol{Y}\right)d\theta_{-j}, \tag{6.15}$$

where $\tilde{\pi}\left(.|.\right)$ represents the conditional approximated density of its arguments.

To obtain the approximations to $\pi\left(\boldsymbol{X}_{i}|\boldsymbol{Y}\right)$ we use approximations to $\pi\left(\theta_{j}|\boldsymbol{Y}\right)$ and $\pi\left(\boldsymbol{X}_{i}|\boldsymbol{\theta},\boldsymbol{Y}\right)$ and we use numerical integration to integrate out $\boldsymbol{\theta}$.

The posterior marginal $\pi(\boldsymbol{\theta}|\boldsymbol{Y})$ of the hyper parameters θ is approximated using a Laplace approximation.

$$\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{Y}) \approx \frac{\pi(\boldsymbol{X},\boldsymbol{\theta},\boldsymbol{Y})}{\tilde{\pi_G}(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{Y})}|_{\boldsymbol{X}=\boldsymbol{X}^*(\boldsymbol{\theta})},$$

where $\tilde{\pi_G}(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{Y})$ is the Gaussian approximation to the full conditional of $\boldsymbol{X}$ and $\boldsymbol{X}(\boldsymbol{\theta})$ is the mode of the full conditional for $\boldsymbol{X}$ for a given $\boldsymbol{\theta}$ (Tierney and Kadane, 1986). We use the approximate sign because the normalising constant for $\pi(\boldsymbol{X},\boldsymbol{\theta}|\boldsymbol{Y})$ is unknown. The approximations of the posterior marginals of the latent field are obtained using the finite sum

$$\tilde{\pi}(\boldsymbol{X}_i|\boldsymbol{Y}) = \sum_K \tilde{\pi}(\boldsymbol{X}_i|\theta_k,\boldsymbol{Y})\,\tilde{\pi}(\theta_k|\boldsymbol{Y})\,\Delta_k.$$

The sum is evaluated at support points θ_k using appropriate weights Δ_k.

Rue and Martino (2007) discussed three different approaches with their features to approximate $\tilde{\pi}(\boldsymbol{X}_i|\theta_k,\boldsymbol{Y})$, namely a Gaussian, a full Laplace and a simplified Laplace approximation. INLA has a great improvement in speed compared to the other MCMC and also a higher level of accuracy (Akerkar et al., 2010). An R package called R-INLA works as an interface for INLA and it is used just as the other R functions. The INLA programme and the R package for INLA are freely available. For more information about the R- INLA project we recommend Rue et al. (2012).

6.1.3.1 Data analysis using INLA for Bayesian inference

For the Uganda DHS data 2011, we assume that the time to death (t_i) of children under the age of five follows a Weibull distribution. This is due to the fact that the Weibull distribution is a flexible distribution which can fit any form of life time data. Given that $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_n)'$ is the vector of coefficients of the covariates considered for analysis, β_0 is the intercept and n the number of covariates, we assume that all these coefficients have a normal prior with mean 0 and variance 0.001. We also assume a gamma prior with parameters 1 and 0.001 for the shape parameter α of the Weibull distribution α.

$$t_i \sim Weibull(\alpha, \lambda_i),$$

where i = 1,... 6692.

To draw inference about the posterior distribution, the information on the prior stated above and the likelihood are used and fitted in INLA. The results are given in the table below.

TABLE 6.1: The results for Bayesian inference by assuming that the time to death of children under-five follows a Weibull distribution.

Results from a Weibull model

Factors	Mean	SD	$[95\%CI]$
Intercept	**-3.52**	**0.233**	$[-3.99, -3.08]$
Fixed effects			
Father's education			
Illiterate	**1**	...	
Complete Primary	**-0.08**	**0.09**	$[-0.28, 0.11]$
Secondary and higher	**-0.330**	**0.18**	$[-0.69, 0.00]$
Sex of the child			
Male	**1**	...	
Female	**-0.19**	**0.09**	$[-0.37, -0.01]$
After at first birth			
Less than 20 years	**1**	...	
20-29 years	**-0.14**	**0.18**	$[-0.36, 0.07]$
30-39 years	**0.49**	**0.58**	$[-0.78, 1.52]$
Births in the past one year			
No-births	**1**	...	
One birth	**0.19**	**0.09**	$[0.00, 0.39]$
Two births	**0.98**	**0.45**	$[0.02, 1.79]$
Sex of the household head			
Male	**1**	...	
Female	**0.29**	**0.1**	$[0.09, 0.49]$
Mother's age group			
Below 20 years	**1**	...	
20- 29 years	**-0.37**	**0.19**	$[-0.75, 0.03]$
30-39 years	**-0.27**	**0.20**	$[-0.66, 0.15]$
40 $^+$ years	**-0.06**	**0.24**	$[-0.52, 0.42]$
Source of drinking water			
Piped water	**1**	...	
Borehole	**0.12**	**0.14**	$[-0.15, 0.39]$
Well	**0.06**	**0.16**	$[-0.25, 0.37]$
Surface/Rain/ Pond/Lake/Tank	**0.24**	**0.17**	$[-0.09, 0.58]$
Others	**0.18**	**0.24**	$[-0.31, 0.65]$
Alpha parameter for Weibull	**0.33**	**0.01**	$[0.30, 0.36]$

Results of a Cox-ph model

Factors	Mean	SD	$[95\%CI]$
Intercept	**-5.67**	**0.23**	$[-6.13, -5.23]$
Fixed effects			
Father's education			
Illiterate	**1**	...	
Complete Primary	**-0.09**	**0.09**	$[-0.28, 0.11]$
Secondary and higher	**-0.34**	**0.18**	$[-0.69, 0.01]$
Sex of the child			
Male	**1**	...	
Female	**-0.19**	**0.09**	$[-0.37, -0.01]$
Age at first birth			
Less than 20 years	**1**	...	
20-29 years	**-0.11**	**0.11**	$[-0.33, 0.09]$
30-39 years	**0.59**	**0.58**	$[-0.69, 1.62]$
Births in the past one year			
No-births	**1**	...	
One birth	**0.25**	**0.09**	$[0.05, 0.44]$
Two births	**1.19**	**0.45**	$[0.22, 1.99]$
Sex of the household head			
Male	**1**	...	
Female	**0.29**	**0.10**	$[0.09, 0.49]$
Mother's age group			
Below 20 years	**1**	...	
20- 29 years	**-0.61**	**0.19**	$[-0.98, -0.19]$
30-39 years	**-0.53**	**0.21**	$[-0.92, -0.11]$
40 $^+$ years	**-0.37**	**0.24**	$[-0.84, 0.10]$
Source of drinking water			
Piped water	**1**	...	
Borehole	**0.12**	**0.14**	$[-0.15, 0.39]$
Well	**0.06**	**0.16**	$[-0.25, 0.37]$
Surface/Rain/ Pond/Lake/Tank	**0.24**	**0.17**	$[-0.09, 0.58]$
Others	**0.17**	**0.24**	$[-0.33, 0.63]$
Precision for baseline Hazard	**18629.97**	**18409.77**	$[1271.41, 67216.46]$
Marginal Likelihood		**-3312.13**	

Results

The mean values presented in Table (6.1), are the means of the estimated coefficients $\hat{\beta}$ of the covariates included in the model. The exponent of these values gives us the hazard ratios, since the exponent of zero is one. Effects whose confidence intervals include zero indicates that they are not significant because they contain the hazard ratio of 1. The null hypothesis is $\hat{\beta} = 0(HR = 1)$ and the alternative hypothesis is that $\hat{\beta} \neq 0(HR \neq 1)$.

From the results presented in Table (6.1), the fixed effects that have confidence intervals which include zero imply that the factors are not significant. Such factors include; father's education, age of the mother at first birth, mother's age group and source of drinking water. These results are consistent with the results got from the models

presented in chapter two and three. The factors that are strongly associated to under-five child survival according to the results presented in the Table (6.1) are; sex of the child, number of births in the past one year and the sex of the household head.

6.1.4 Bayesian frailty modelling

In this section,

the model of interest is the Weibull frailty model with additive frailties.

Let time to death of children under the age of five in Uganda represented by a sample of 6692 children presented in the UDHS 2011 data set follow a Weibull distribution, t_{ij} is the survival time of the j^{th} child in the i^{th} cluster. In our data set there are 404 clusters at the community level and at household level there are 4285 clusters. This implies that $i = 1, 2, \ldots, 404$ at community level and $i = 1, 2, \ldots, 4285$ at household level.

$$t_{ij} \sim weibull(\alpha, \eta_{ij}).$$

The conditional hazard function given the unobserved frailty $\boldsymbol{z}_i$, a covariate vector $\boldsymbol{X}_{ij}$ and the Weibull parameter α is given by:

$$h\left(t_{ij}|\boldsymbol{X}_{ij}, \boldsymbol{z}_i, \alpha\right) = \alpha t_{ij}^{\alpha-1} \exp\left[\boldsymbol{\eta}_{ij}\right], \tag{6.16}$$

where $\boldsymbol{\eta}_{ij} = \boldsymbol{\beta}_0 + \boldsymbol{X}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_i$.

The likelihood function can then be obtained as described in chapter three. We further assume that the coefficients of the fixed effects have a normal prior with parameters 0 and 0.001, the frailty term $z_i \sim N\left(0, \tau\right)$ and a Gaussian prior for τ and α with parameters 1 and 0.001. The assumptions used on the prior parameters and the frailty term are the same assumptions used by (Akerkar et al., 2010, Rue et al., 2012).

6.1.4.1 Data analysis for the Weibull frailty model

TABLE 6.2: The results for Bayesian inference, a model with frailty at community and household level.

Results for frailty at community level				**Results for frailty at household level**			
Factors	**Mean**	**SD**	[95%*CI*]	**Factors**	**Mean**	**SD**	[95%*CI*]
Intercept	**-3.49**	**0.24**	$[-3.97, -3.05]$	**Intercept**	**-3.89**	**0.26**	$[-4.42, -3.39]$
Fixed effects				**Fixed effects**			
Father's education				**Father's education**			
Illiterate	**1**	...		**Illiterate**	**1**	...	
Complete Primary	**-0.08**	**0.09**	$[-0.28, 0.11]$	**Complete Primary**	**-0.08**	**0.14**	$[-0.28, 0.12]$
Secondary and higher	**-0.33**	**0.18**	$[-0.69, 0.00]$	**Secondary and higher**	**-0.34**	**0.18**	$[-0.70, 0.01]$
Sex of the child				**Sex of the child**			
Male	**1**	...		**Male**	**Ref**	...	
Female	**-0.19**	**0.09**	$[-0.37, -0.01]$	**Female**	**-0.19**	**0.09**	$[-0.38, -0.01]$
Age at first birth				**Age at first birth**			
Less than 20 years	**1**	...		**Less than 20 years**	**1**	...	
20-29 years	**-0.14**	**0.11**	$[-0.36, 0.06]$	**20-29 years**	**-0.16**	**0.11**	$[-0.38, 0.06]$
30-39 years	**0.49**	**0.59**	$[-0.78, 1.52]$	**30-39 years**	**0.56**	**0.62**	$[-0.78, 1.67]$
Births in the past one year				**Births in the past one year**			
No-births	**1**	...		**No-births**	**1**	...	
One birth	**0.19**	**0.09**	$[0.00, 0.39]$	**One birth**	**0.19**	**0.10**	$[-0.01, 0.39]$
Two births	**0.99**	**0.45**	$[0.12, 1.79]$	**Two births**	**1.189**	**0.50**	$[0.12, 2.10]$
Sex of the household head				**Sex of the household head**			
Male	**1**	...		**Male**	**1**	...	
Female	**0.29**	**0.10**	$[0.09, 0.49]$	**Female**	**0.30**	**0.10**	$[0.09, 0.51]$
Mother's age group				**Mother's age group**			
Below 20 years	**1**	...		**Below 20 years**	**1**	...	
20- 29 years	**-0.37**	**0.19**	$[-0.75, 0.04]$	**20- 29 years**	**-0.35**	**0.21**	$[-0.75, 0.08]$
30-39 years	**-0.26**	**0.21**	$[-0.65, 0.15]$	**30-39 years**	**-0.24**	**0.21**	$[-0.65, 0.19]$
40 $^+$ years	**-0.05**	**0.24**	$[-0.52, 0.43]$	**40 $^+$ years**	**-0.02**	**0.25**	$[-0.51, -0.49]$
Source of drinking water				**Source of drinking water**			
Piped water	**Ref**	...		**Piped water**	**1**	...	
Borehole	**0.12**	**0.14**	$[-0.15, 0.39]$	**Borehole**	**0.12**	**0.14**	$[-0.15, 0.41]$
Well	**0.06**	**0.16**	$[-0.25, 0.37]$	**Well**	**0.06**	**0.17**	$[-0.26, 0.39]$
Surface/Rain/ Pond/Lake/Tank	**0.24**	**0.17**	$[-0.10, 0.58]$	**Surface/Rain/ Pond/Lake/Tank**	**0.24**	**0.18**	$[-0.11, 0.59]$
Others	**0.18**	**0.24**	$[-0.31, 0.65]$	**Others**	**0.20**	**0.25**	$[-0.31, 0.69]$
Random effects				**Random effects**			
Precision for frailty term	**61.33**	**64.88**	$[11.66, 63.47]$	**Precision for frailty term**	**1.42**	**0.35**	$[0.91, 2.28]$
Alpha parameter for Weibull	**1.18**	**0.19**	$[1.07, 1.29]$	**Alpha parameter for Weibull**	**0.33**	**0.01**	$[0.30, 0.36]$
Marginal likelihood		**-2951.26**		**Marginal Likelihood**		**-2945.52**	

6.1.4.2 Results

From Table (6.2), the precision of the frailty term is 61.3. Precision is defined as the reciprocal of the variance. The variance of the frailty term is 0.016, which is very close to zero and therefore confirms the results that the community frailty is not significant. This implies that all the differences among the mortality rates of the children under the age of five are explained by the observed fixed covariates stated in the model.

The precision for the household frailty term is 1.42 with a tight confidence interval presented in the table indicates that the frailty at household level is significant. Children under the age of five in different households were exposed to different risks of death whereby some of the children were more exposed to the risk than others.

The factors that were strongly associated to a high under-five child mortality were sex of the child , number of births in the past one year and sex of the child just as concluded earlier on.

Generally the confidence intervals for all the fixed effects considered in the household effects model appear wider than in the community effects model.

Chapter 7

Discussion and conclusions

Two broad methods were used for inference in this thesis namely the Frequentist and Bayesian approaches. Each of these approaches has its advantages and disadvantages. To use the Bayesian approach, one has to accept that parameters of interest may be treated as random variables with the so called prior. Advantages of the Frequentist inference over the Bayesian inference are listed by Hall (2012) and these include; Frequentists models are easier to implement because there is no requirement to set any prior distributions. All aspects of parameter estimation are data driven unlike the Bayesian models where prior distributions and initial values for approximations have to be specified.

In addition to the above, the run time for the Frequentists models are shorter than for those of Bayesian models via the MCMC. The problem becomes even more compounded when one is dealing with complex models with larger sample sizes taking weeks in Bayesian inference via MCMC. Note: However the recent Laplace approximations when used in Bayesian estimation give run-times that are as fast as the Frequentists Maximum likelihood estimation even in case of a large data set. This is one of the facts on which the INLA package in R was built. Further more, the Frequentist models are able to include large data sample sizes well as Bayesian models via MCMC have been restricted to small sample sizes.

Note: This ristriction however has been overcome by running the MCMC Algorithms in Laplace Approximation functions. These do not loop through records and are vectorised allowing for handling large data sets.

Hall (2012) also presents the advantages of Bayesian inference over the Frequentist inference and these include; (a) Bayesian models allow for informative prior so that the prior knowledge can be used to inform the current models. (b) Bayesian inference assumes the

data to be fixed and the unknown parameters to be random which is true and this is the opposite with Frequentists inference. The Frequentists estimation is therefore not based on the data at hand but data at hand plus hypothetical repeated sampling in future with similar data. (c) There is no Frequentists probability distribution associated with the unknown parameters or hypotheses. Bayesian inference therefore estimates a full probability model. (d) Bayesian inference estimates the probability of the hypothesis given the data where as the Frequentists estimate the probability of the data given the hypothesis. Hypothesis testing itself suggests that one should test for the hypothesis given the data. (e) Bayesian inference uses observed data only while Frequentists uses observed and future data that is unobserved and hypothetical.

As per what type of inference gives the most precise results remains as a choice and preference of the researcher. For this thesis the conclusions got by using Bayesian inference and Frequentists inference were not in anyway different.

In this thesis one has to note that some of the variables that were excluded from the analysis was due to the selection of variables and model selection criteria used but this does not mean that these variables have no effect on the survival time of the children under the age of five years in Uganda. Given the fact that the models used in this section have got defined assumptions, there was not enough evidence to include some of these covariates in the final analysis.

There was no evidence to prove that the frailty effect was significant at the community level but there was enough evidence from the data to show the significance of the frailty effect at the household level. This implies that survival times of children under the age of five in a given community in a country are different from the survival time of the children under the age of five in any other community in the country whereas children under the age of five who live in the same household have correlated survival times. This indicates that more efficient interventions may be those that target individual households rather than communities. This may be expensive but may realistically help reduce on infant and child mortality and hence see the way to achieving the millennium development goal.

The results in this thesis do not deviate much from the findings of other sub-Saharan African countries such as Zimbabwe, Kenya, Tanzania among others (Hobcraft et al., 1984). The results indicate that sex of the child, sex of the household head and number of births in the last one year are strongly associated to the survival of children under the age of five. The results also indicated that children born in broken families were exposed to a higher risk of death than those in families that had both the parents and the father as the family head. A female child was at a lower risk of death than a male child and this is attributed to the value most of the tribes in Uganda attach to the girl

child. The female child is seen as a source of wealth to the family, thus the female child is preferentially given more care or attention.

Mothers who had more than one child born in the last one year exposed their children to a higher risk of death before reaching the age of five. This covariate also shows that short birth intervals are associated with a high child mortality rate. The short birth intervals portrayed by the results can also be as a result of the low use of modern family planning methods. The data shows that only 25.3% of the women were using modern family planning methods, 3.4% were using traditional methods, 47.5% were non-users and intend to use later and 23.7% of the women did not intend to use any family planning method in the near future at the time of the survey. The data also shows that the health facilities where these women go for their maternal health failed to inform women on the several methods available for family planning because 50% of the women confessed that the health facilities did not inform them about family planning and only 27.0% of these women claimed to have been told about family planning and 22.7% of the women didnot answer the question and therefore had missing information.

Covariates such as previous and preceding birth interval which could have given us good information on the birth intervals of the women included in this sample were not included in the analysis. This is as a result of the high level of missing information with these covariates. Most mothers did not know their preceding and succeeding birth intervals.

Other factors like fathers education, source of drinking water, mother's age at first birth among others, have an effect on the survival of children under the age of five but the effect seemed insignificant based on the results. The results however give us some clues on how some of these factors are associated to under-five children mortality. For example, the results show that Families with fathers who had attained secondary and higher education, had their children under the age of five exposed to a low risk of death than those in households where the father had not even attained the basic primary education.

Recommendations from this study are that, more efficient interventions may be those that target individual households rather than communities. Mothers should be made more aware about modern family planning methods to enable them plan for their families well and to also avoid short preceding and succeeding birth intervals. Parents should also be made aware that all children require equal treatment regardless of their sex. This will reduce on the death of children due to negligence by parents basing on the sex of the child. Issues on who to take care of the child after the marriage is broken should be handled carefully by the state to stop making children under the age of five from such families vulnerable or exposed to high risk of death.

The DHS survey is associated with the following critics and they include; the DHS data collection involves collection of data from women age between the age $15 - 49$ who are alive in a given household but for mothers who have died, no information is collected, this creates a bias in the results. More to that, the DHS survey uses a retrospective technique. This is associated to many problems and challenges. The major problem with it is that the quality of the results depends on the completeness and correctness of the birth and death history. In most cases incomplete and inaccurate information is provided which affects the final conclusions and recommendations. A follow up study may be a good alternative to the DHS surveys, in this case a cohort of women is followed up for five years, recording the death and birth histories of their children.

Bibliography

O. Aalen, O. Borgan and H. Gjessing. *Survival and event history analysis: a process point of view*. Springer, 2008a.

O. Aalen, O. Borgan and H. K. Gjessing. Nonparametric analysis of survival and event history data. *Survival and Event History Analysis*, pages 69–130, 2008b.

W. W. M. Abeysekera and M. R. Sooriyarachchi. Use of schoenfeld's global test to test the proportional hazards assumption in the cox proportional hazards model: an application to a clinical study. *Journal of the National Science Foundation of Sri Lanka*, 37(1):41–51, 2009.

H. Akaike. Information theory and an extension of the maximum likelihood principle. In *International Symposium on Information Theory, 2 nd, Tsahkadsor, Armenian SSR*, pages 267–281, 1973.

H. Akaike. Factor analysis and aic. *Psychometrika*, 52:317–332, 1987. ISSN 0033-3123. doi: 10.1007/BF02294359. URL http://dx.doi.org/10.1007/BF02294359.

R. Akerkar, S. Martino, and H. Rue. Implementing approximate bayesian inference for survival analysis using integrated nested laplace approximations. Technical report, Technical report 1, Department of mathematical sciences, Norwegian University of Science and Technology, 2010.

D. Arthur, L. Nan, and R. Donald. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, pages 1–38, 1977.

R. Ayiko, D. Antai, A. Kulane, et al. Trends and determinants of under-five mortality in uganda. *East African journal of public health*, 6(2):136, 2009.

M. Bailey. Factors affecting infant and child mortality in rural sierra leone. *Journal of tropical pediatrics*, 34(4):165–168, 1988.

C. Berzuini and D. Clayton. Bayesian analysis of survival on multiple time scales. *Statistics in medicine*, 13(8):823–838, 1994.

W. M. Bolstad and S. O. Manda. Investigating child mortality in malawi using family and community random effects. *Journal of the American Statistical Association*, 96 (453):12–19, 2001.

A. Brezger, T. Kneib, and S. Lang. Bayesx: Analysing bayesian structured additive regression models. Technical report, Discussion paper//Sonderforschungsbereich 386 der Ludwig-Maximilians-Universität München, 2003.

J. Bryce, N. Terreri, C. G. Victora, E. Mason, B. Daelmans, Z. A. Bhutta, F. Bustreo, F. Songane, P. Salama, and T. Wardlaw. Countdown to 2015: tracking intervention coverage for child survival. *The Lancet*, 368(9541):1067–1076, 2006.

M.H. Chen, Q.M. Shao, and G. Joseph. Ibrahim (2000) monte carlo methods in bayesian computation springer series in statistics, 2000.

D. G. Clayton. A monte carlo method for bayesian inference in frailty models. *Biometrics*, pages 467–485, 1991.

D.G. Clayton. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1):141–151, 1978.

M. A. Cleves, W. W. Gould, and R. G. Gutierrez. *An introduction to survival analysis using Stata*. Stata Corp, 2008.

D. Collett. Modelling survival data in medical research. *Texts in statistical science*, 1994.

D. Collett. *Modelling survival data in medical research*, volume 57. Chapman & Hall/CRC, 2003.

D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 187–220, 1972.

K. Croke. The political economy of child mortality decline in tanzania and uganda, 1995–2007. *Studies in Comparative International Development (SCID)*, pages 1–23, 2012.

S. L. Curtis and J. W. McDONALD. Birth spacing and infant mortality in brazil. *Journal of biosocial science*, 23(3):343–352, 1991.

M. DHS. Demographic and health survey. 2012.

A. J. Dobson. *An introduction to generalized linear models*. Chapman & Hall, 2002.

L. Duchateau and P. Janssen. the frailty model (statistics for biology and health), 2007.

L. Duchateau and P. Janssen. *The frailty model*. Springer Verlag, 2008.

L. Fahrmeir, G. Tutz, and W. Hennevogl. *Multivariate statistical modelling based on generalized linear models*, volume 2. Springer New York, 1994.

T. R. Fleming and D. P. Harrington. *Counting processes and survival analysis*, volume 169. Wiley, 2011.

A. E. Gelfand and A. F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409, 1990.

S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):721–741, 1984.

W. R. Gilks, S. Richardson, D. J. Spiegelhalter, et al. Markov chain monte carlo in practice. 1996. *New York: Chapman Hall/CRC*, 486, 1996.

P. M. Grambsch and T. M. Therneau. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81(3):515–526, 1994.

B. Hall. Bayesian inference. *Statisticat, LLC*, 2012.

D. D. Hanagal. *Modeling Survival Data Using Frailty Models*. Chapman & Hall/CRC, 2011.

M. Håvard, R. Sara and C. Nicolas. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2):319–392, 2009.

J. N. Hobcraft, J. W. McDonald, and S. O. Rutstein. Socio-economic factors in infant and child mortality: A cross-national comparison. *Population studies*, 38(2):193–223, 1984.

D. W. Hosmer Jr, S. Lemeshow, and S. May. *Applied survival analysis: regression modeling of time to event data*, volume 618. Wiley-Interscience, 2011.

R. I. Jennrich and S. M. Robinson. A newton-raphson algorithm for maximum likelihood factor analysis. *Psychometrika*, 34:111–123, 1969. ISSN 0033-3123. doi: 10.1007/BF02290176. URL http://dx.doi.org/10.1007/BF02290176.

G. Jones. Markov chain monte carlo methods for inference in frailty models with doubly-censored data. *Journal of data science*, 2:33–47, 2004.

E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

L. Kazembe, A. Clarke, and N. B. Kandala. Childhood mortality in sub-saharan africa: cross-sectional insight into small-scale geographical inequalities from census data. *BMJ open*, 2(5), 2012.

J. P. Klein and P. K. Goel. *Survival analysis: state of the art*, volume 211. Springer, 1992.

D. G. Kleinbaum, M. Klein, and E. R. Pryor. *Logistic regression: a self-learning text*. Springer Verlag, 2002.

B. Kyaddondo. Uganda population stabilisation report. *Population Trends and Policy Options in Selected Developing Countries*, page 202, 2012.

C. T. Le and Le. *Applied survival analysis*. Wiley New York, 1997.

W. Limin. Determinants of child mortality in ldcs: Empirical findings from demographic and health surveys. *Health Policy*, 65(3):277 – 299, 2003. ISSN 0168-8510. doi: 10.1016/S0168-8510(03)00039-3. URL http://www.sciencedirect.com/science/article/pii/S0168851003000393.

ORC. Macro. Demographic and health survey. *ORC Macro. http://www. measuredhs. com*, 2004.

A. D. Martin and K. M. Quinn. Dynamic ideal point estimation via markov chain monte carlo for the us supreme court, 1953–1999. *Political Analysis*, 10(2):134–153, 2002.

A. K. Mbonye, M. Sentongo, G. K. Mukasa, R. Byaruhanga, O. Sentumbwe-Mugisa, P. Waiswa, H. N. Sengendo, P. Aliganyira, M. Nakakeeto, J. E. Lawn, et al. Newborn survival in uganda: a decade of change and future implications. *Health Policy and Planning*, 27(suppl 3):iii104–iii117, 2012.

J. S. Mehta and R. Srinivasan. Estimation of the mean by shrinkage to a point. *Journal of the American Statistical Association*, 66(333):86–90, 1971.

R. G. Miller Jr. *Survival analysis*. Wiley-Interscience, 2011.

N. I. Mondai, A. R. Khan, and J. Chakma. Factors affecting child mortality and schooling of rural chalona in cht, bangladesh. *Soc. Sci*, 3(2):189–192, 2008.

W. H. Mosley, L. C. Chen, et al. Child survival: strategies for research continued. *Population and development review*, 10(Supplement), 1984.

F. Nuwaha, J. Babirye, and N. Ayiga. Why the increase in under five mortality in uganda from 1995 to 2000? a retrospective analysis. *BMC public health*, 11(1):725, 2011a.

F. Nuwaha, J. Babirye, O. Okui, and N. Ayiga. Understanding socio-economic determinants of childhood mortality: a retrospective analysis in uganda. *BMC research notes*, 4(1):484, 2011b.

G. W. Oehlert. A note on the delta method. *The American Statistician*, 46(1):27–29, 1992.

G.W. Oehlert. Estimating the mean of a positive random variable. *unpublished Ph. D. dissertation, Yale University, Dept. of Statistics*, 1981.

Uganda Bureau of Statistics Kampala. *Uganda: Demographic and Health Survey 2006*. Macro International, 2007.

P. Okwero, A. Tandon, S. Sparkes, J. McLaughlin, and J. G Hoogeveen. *Fiscal space for health in Uganda*, volume 186. World Bank Publications, 2010.

D. W. Omariba, R. Beaujot, and F. Rajulton. Determinants of infant and child mortality in kenya: an analysis controlling for frailty effects. *Population Research and Policy Review*, 26:299–321, 2007. ISSN 0167-5923. doi: 10.1007/s11113-007-9031-z. URL http://dx.doi.org/10.1007/s11113-007-9031-z.

B. Phipson and H. Mwambi. Incorporating frailty effects in the cox proportional hazards model using two independent methods in independent data sets. *South African Statistical Journal*, 44(1):61–72, 2010.

U. Protectorate. Uganda census 1959: African population. *Statistics Branch. Ministry of Economic Affairs*, 1961.

R. Reinikka and P. Collier. *Uganda's recovery: the role of farms, firms, and government*. World Bank Publications, 2001.

G. O. Roberts and N. G. Polson. On the geometric convergence of the gibbs sampler. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 377–384, 1994.

H. Rue and S. Martino. Approximate bayesian inference for hierarchical gaussian markov random field models. *Journal of statistical planning and inference*, 137(10):3177–3192, 2007.

H. Rue, S. Martino, and F. Lindgren. The r-inla project, 2012.

S. O. Rutstein et al. Factors associated with trends in infant and child mortality in developing countries during the 1990s. *BULLETIN-WORLD HEALTH ORGANIZATION*, 78(10):1256–1270, 2000.

M. J. Schervish and B. P. Carlin. On the convergence of successive substitution sampling. *Journal of Computational and Graphical statistics*, 1(2):111–127, 1992.

D. Schoenfeld. Partial residuals for the proportional hazards regression model. *Biometrika*, 69(1):239–241, 1982.

J. H. Shih and T. A. Louis. Assessing gamma frailty models for clustered failure time data. *Lifetime Data Analysis*, 1(2):205–220, 1995.

C. G. Small and J. J. Wang. *Numerical methods for nonlinear estimating equations*, volume 29. Oxford University Press, USA, 2003.

G. F. Ssengonzi, R. De Jong and S. C. Shannon. The effect of female migration on infant and child survival in uganda. *Population Research and Policy Review*, 21(5):403–431, 2002.

S. Ssewanyana and S. D. Younger. Infant mortality in uganda: Determinants, trends and the millennium development goals. *Journal of African Economies*, 17(1):34–61, 2008.

J. R. Thompson. Some shrinkage techniques for estimating the mean. *Journal of the American Statistical Association*, pages 113–122, 1968.

L. Tierney and J. B. Kadane. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–86, 1986.

Z. Uddin, J. Hossain, M. O. Ullah, et al. Child mortality in a developing country: A statistical analysis. *Journal of Applied Quantitative Methods*, 3:270–283, 2009.

Uganda Ministry of Planning and Economic Development Statistics Division. *Report on the 1969 Population Census: The analytical report*, volume 4. 1976.

W. K. G Vaupel, J. Manton and E. Stallard. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3):439–454, 1979.

A. Wienke. *Correlated Frailty Models in Survival Analysis*, volume 37. Chapman & Hall/CRC, 2009.

P. Wienke, A. Lichtenstein and A. I. Yashin. A bivariate frailty model with a cure fraction for modeling familial correlations in diseases. *Biometrics*, 59(4):1178–1183, 2003.

World Bank Development Data Group. *World Development Indicators 2012*. World Bank-free PDF, 2012.

World Bank Group. *Migration and Remittances Factbook 2011*. World Bank Publications, 2010.

Delta method

According to Oehlert (1992), the delta method was introduced by Thompson (1968) and modified by different authors including Mehta and Srinivasan (1971) and Oehlert (1981). To find the estimate of the variance of the Kaplan-Meier estimate, we need to introduce the concept of the delta method. This method uses the first order Taylor series expansion of a function $h(X)$ around the mean $E(X) = \mu$ of a random variable X.

$$\begin{aligned} h(X) \approx & h(\mu) + h^{'}(\mu)(X-\mu), \\ Var(h(X)) \approx & Var\left(h(\mu) + h^{'}(\mu)(X-\mu)\right), \\ = & h^{'2}(\mu) Var(X-\mu) \\ = & h^{'2}(\mu) Var(X). \end{aligned}$$

provided the function $h(X)$ is first-order differentiable.

In order to get the variance of the estimate to the survival function $\widehat{S(t)}$, we apply the delta method on the estimate $\widehat{S}(t)$ of the survival function.

$$\begin{aligned} \widehat{S}(t) = & \prod_{i=1}^{r}\left(\frac{n_i - d_i}{n_i}\right), \\ \log \widehat{S}(t) = & \log \prod_{i=1}^{r}\left(\frac{n_i - d_i}{n_i}\right) = \sum_{i=1}^{r} \log\left(\frac{n_i - d_i}{n_i}\right), \\ Var\left\{\log \widehat{S}(t)\right\} = & \sum_{i=1}^{r} Var\left\{\log\left(\frac{n_i - d_i}{n_i}\right)\right\}. \end{aligned}$$

by applying the delta method on the right hand side of the equation,

$$Var\left\{\log \widehat{S}(t)\right\} = \sum_{i=1}^{r} \frac{d_i}{n_i (n_i - d_i)}$$

by applying the delta method on the left hand side we get the estimate for variance and then derive the estimate for the standard error as shown below:

$$Var\left\{\log \widehat{S}(t)\right\} = \frac{1}{\widehat{S}(t)^2} Var\left\{\hat{S(t)}\right\}$$

$$Var\left\{\widehat{S}(t)\right\} = \left[\widehat{S}(t)\right]^2 \sum_{i=1}^{r} \frac{d_i}{n_i (n_i - d_i)}.$$

Schoenfeld Residuals

For the proportionality hazard assumption to be true, the schoenfeld residuals centred at zero should be independent of time, Schoenfeld (1982).

The approach considers one covariate at a time i.e. we get one set of residuals and one p-value per covariate in the model. Suppose a failure occurs and that the model we have is correct, the failure would happen to an individual k with probability:

$$P\left(kfails\right) = \frac{e^{\beta x_k}}{\sum_{j \in R(t)} e^{\beta x_j}}.$$

The individual who fails has at time t_i has a covariate x_i, the expected value of this is:

$$\begin{aligned} E\left(X_i\right) &= \sum_{k \in R(t_i)} x_k P\left(kfails\right), \\ &= \sum_{k \in R(t_i)} \frac{x_k e^{\beta x_k}}{\sum_{j \in R(t)} e^{\beta x_j}}. \end{aligned}$$

The difference between the observed and the expected covariates of the person who fails at time t_i is given by:

$$\hat{r}_i = x_i - E\left(X_i\right). \tag{1}$$

If the proportional hazard assumption is true, (1) should be independent of time, Schoenfeld (1982). Alternatively Grambsch and Therneau (1994) proposed the scaled schoenfeld residuals. He scaled the residuals by an estimate of their variance. The scaled schoenfeld residuals are given by:

$$\tilde{r} = m\hat{V}\left(\hat{\beta}\right)\hat{r}_i.$$

Where m is the number of censored observations and this is what the *R* package and other softwares give. The null hypothesis is that the scale residuals are independent of failure times and the test statistic is based on a Chi-square test derived in Grambsch and Therneau (1994) and R automatically calculates it using the Cox.phz command.

The Gibbs Sampler (Alternating conditional sampling)

.1 The Gibbs Sampler (Alternating conditional sampling)

According to Chen et al. (2000), the Gibbs Sampler is one of the best known MCMC sampling algorithm in Bayesian computational literature. They add that the formal term was introduced by Geman and Geman (1984). According to Duchateau and Janssen (2008), Gibbs sampling is based on the posterior density of each parameter conditional on all other parameters.

With Gibbs sampling, the high dimensional problem is reduced to many unidimensional or lower dimensional problems. They continue to say that, the parameter vector is divided into a number of subvectors and in each iteration, sampling cycles through the different subvectors and draws a random sample for the subvector using the posterior density of the subvectors conditional to the current values of the other subvectors. We will design the subvector such that it consists of only one parameter. The advantage of Gibbs sampling is that the posterior densities are easy to obtain and also easy to draw samples from them.

Let $\eta = (\eta_1, \eta_2, \ldots, \eta_p)'$ denote a p-dimensional vector of parameters and let $f(\eta|X)$ be the posterior distribution given the data. The basic Gibbs sampler is given as follows:

Step 0:

- Choose an arbitrary starting point:

$$\eta_0 = (\eta_{1,0}, \eta_{2,0}, \ldots, \eta_{,o})',$$

- Set $i = 0$.

Step 1:

- Generate

$$\eta_{i+1} \sim (\eta_{1,i+1}, \eta_{2,i+1}, \ldots, \eta_{p,i+1})'$$

as follows;

- Generate

$$\eta_{1,i+1} \sim f(\eta_1 | \eta_{2,i+1}, \ldots, \eta_{p,i}, X);$$

- Generate

$$\eta_{2,i+1} \sim f(\eta_2 | \eta_{1,i+1}, \eta_{3,i+1}, \ldots, \eta_{p,i}, X);$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots;$$

- Generate

$$\eta_{p,i+1} \sim f(\eta_p | \eta_{1,i+1}, \eta_{2,i+1}, \ldots, \eta_{p-1,i+1}, X).$$

Step 2: Set $i = i + 1$ and go to step 1.

Therefore each component of η is visited in the natural order and a cycle in this scheme requires generation of p random variables.

Gelfand and Smith (1990) show that under certain regularity conditions, the vector sequence $\{\eta_i, i = 1, 2, \ldots\}$ has a stationary distribution $f(\eta|X)$.

Schervish and Carlin (1992) provide a sufficient condition that guarantees geometric convergence. Other properties regarding geometric convergence are discussed in Roberts and Polson (1994).

R Codes

```
############### The univariate Cox-ph model, testing
for the schoenfield residuals and  Exploratory
 data analysis codes ########################

#### Changing categorical variables into factors####
Data$V149 = as.factor(Data$V149)
Data$V729 = as.factor(Data$V729)
Data$B4 = as.factor(Data$B4)
Data$V201= as.factor(Data$V201)
Data$V025= as.factor(Data$V025)
Data$V190= as.factor(Data$V190)
Data$M4= as.factor(Data$M4)
Data$M18= as.factor(Data$M18)
Data$V130<-as.factor(Data$V130)
#Data$B11= as.factor(Data$B11)
Data$BORD<-as.factor(BORD)
Data$V212<-as.factor(V212)
#Data$previousbirthinterval2<-as.
factor(previousbirthinterval2)
Data$V209<-as.factor(Data$V209)
Data$V208<-as.factor(Data$V208)
Data$V137<-as.factor(Data$V137)
Data$V013<-as.factor(Data$V013)
Data$V151<-as.factor(Data$V151)
Data$V714<-as.factor(Data$V714)
Data$V113<-as.factor(Data$V113)
Data$V717<-as.factor(Data$V717)
Data$B0<-as.factor(Data$B0)
Data$V130<-as.factor(Data$V130)
Data$V116<-as.factor(Data$V116)
################################

#####mother's educ V149#############
Data$Censorstatus
```

```
Data$V149 = as.factor(Data$V149)
#summary(data)
coxfit=coxph(Surv(Data$Survivaltime,Data$Censorstatus)~
Data$V149, data=Data)
coxfit
#plotHR(coxfit)
summary(coxfit)
###Testing for proportionality hazard assumption######
cox.zph(coxfit,transform=rank)
plot(cox.zph(coxfit,transform=rank))
####end mother's education#################

######father's educ V729#######################
Data$V729 = as.factor(Data$V729)
#summary(data)
coxfit1=coxph(Surv(Data$Survivaltime,Data$Censorstatus)~
Data$V729, data=Data)
coxfit1
summary(coxfit1)####satisfies coxph
###Testing for proportionality hazard assumption######
cox.zph(coxfit1,transform=rank)
plot(cox.zph(coxfit1,transform=rank))
######################end#####################

#################Child sex B4 ######################
Data$B4 = as.factor(Data$B4)
coxfit2=coxph(Surv(Data$Survivaltime,Data$Censorstatus)~
Data$B4+frailty(V001) ,data=Data)
coxfit2
summary(coxfit2)
###Testing for proportionality hazard assumption######
cox.zph(coxfit2,transform=rank)
plot(cox.zph(coxfit2,transform=rank))
#################End child sex ############################

############################V201 Children ever born#######
Data$V201= as.factor(Data$V201)
Data$V201
class(Data$V201)
Data$V201
coxfit3=coxph(Surv(Data$Survivaltime,Data$Censorstatus)~
Data$V201,data=Data)
coxfit3
summary(coxfit3)
###Testing for proportionality hazard assumption######
cox.zph(coxfit3,transform=rank)
plot(cox.zph(coxfit3,transform="km", global=TRUE))
########################End Partner's educ#############
```

```
############################Type of place of residence V102##
Data$V025= as.factor(Data$V025)
Data$V025 ###Not significant
coxfit4=coxph(Surv(Data$Survivaltime,Data$Censorstatus)~
Data$V025,data=Data)
coxfit4
summary(coxfit4)
###Testing for proportionality hazard assumption######
cox.zph(coxfit4,transform=rank)
plot(cox.zph(coxfit4,transform="km", global=TRUE))
#################eEnd type of place of residence########

############################V190 Wealth index###############
Data$V190= as.factor(Data$V190)###satisfies coxph
Data$V190
coxfit5=coxph(Surv(Data$Survivaltime,Data$Censorstatus)~
Data$V190,data=Data)
coxfit5
summary(coxfit5)
###Testing for proportionality hazard assumption######
cox.zph(coxfit5,transform=rank) ##doesnot satisfy cox ph
plot(cox.zph(coxfit5,transform="km", global=TRUE))
##############################end V190 Wealth index############

########################M4 Duration of breast feeding############
Data$M4= as.factor(Data$M4)
Data$M4
coxfit6=coxph(Surv(Data$Survivaltime,Data$Censorstatus)~
Data$M4,data=Data)
coxfit6
###Testing for proportionality hazard assumption######
cox.zph(coxfit6,transform=rank)
plot(cox.zph(coxfit6,transform="km", global=TRUE))
#################end of M4 Duration of breast feeding###

######################M18 size of child at birth########
Data$M18= as.factor(Data$M18) #satisfies  cox ph
Data$M18
coxfit7=coxph(Surv(Data$Survivaltime,Data$Censorstatus)~
Data$M18,data=Data)
coxfit7
summary(coxfit7)
###Testing for proportionality hazard assumption######
cox.zph(coxfit7,transform=rank)
plot(cox.zph(coxfit7,transform="km", global=TRUE))
############################end M18 size of child at birth#########
############################birth intervals B11 nad B12 ############
```

```
Data$B11= as.factor(Data$B11)
Data$B11
coxfit8=coxph(Surv(Data$Survivaltime,Data$Censorstatus)~
Data$B11,data=Data)
coxfit8
###Testing for proportionality hazard assumption######
cox.zph(coxfit8,transform=rank)
plot(cox.zph(coxfit8,transform="km", global=TRUE))
##################end birth intervals B11 and B12#####
#######Birth order BORD##############################
Data$BORD<-as.factor(BORD)
Data$BORD= as.factor(Data$BORD)
Data$BORD ####n
coxfit9=coxph(Surv(Data$Survivaltime,Data$Censorstatus)~
Data$BORD,data=Data)
coxfit9
summary(coxfit9)
###Testing for proportionality hazard assumption######
cox.zph(coxfit9,transform=rank)
plot(cox.zph(coxfit9,transform="km", global=TRUE))
############end birth order############################
########Age at first birth V212#################################
Data$V212<-as.factor(V212)
Data$V212= as.factor(Data$V212)#satisfies cox ph
Data$V212 ####n
coxfit10=coxph(Surv(Data$Survivaltime,Data$Censorstatus)~
Data$V212,data=Data)
coxfit10
summary(coxfit10)
###Testing for proportionality hazard assumption######
cox.zph(coxfit10,transform=rank)
plot(cox.zph(coxfit10,transform="km", global=TRUE))
############end birth order############################
#####Previous birth interval B11################################
Data$previousbirthinterval2<-as.factor(previousbirthinterval2)
Data$previousbirthinterval2= as.factor(Data$previousbirthinterval2)
Data$previousbirthinterval2####n
###Testing for proportionality hazard assumption######
coxfit11=coxph(Surv(Data$Survivaltime,Data$Censorstatus)~
Data$previousbirthinterval2,data=Data)
coxfit11
summary(coxfit11)
cox.zph(coxfit11,transform=rank)
plot(cox.zph(coxfit11,transform="km", global=TRUE))
############end previous birth interval#################
#####Births in the past one yr#############################
Data$V209<-as.factor(Data$V209)
Data$previousbirthinterval2= as.factor(Data$V209)
```

```
Data$V209####n
coxfit12=coxph(Surv(Data$Survivaltime,Data$Censorstatus)~
Data$V209,data=Data)
coxfit12
summary(coxfit12)
###Testing for proportionality hazard assumption######
cox.zph(coxfit12,transform=rank)
plot(cox.zph(coxfit12,transform="km", global=TRUE))
############end births in the past one yr#############

#####Births in the past one yr###################
Data$V209<-as.factor(Data$V209)
Data$V209= as.factor(Data$V209)
Data$V209####n
coxfit13=coxph(Surv(Data$Survivaltime,Data$Censorstatus)~
Data$V209,data=Data)
coxfit13
summary(coxfit13)
###Testing for proportionality hazard assumption######
cox.zph(coxfit13,transform=rank)
plot(cox.zph(coxfit13,transform="km", global=TRUE))
############end births in the past one yr#############

#####Births in the past five yrs###############
Data$V208<-as.factor(Data$V208)
Data$V208= as.factor(Data$V208)
Data$V208####n
coxfit13=coxph(Surv(Data$Survivaltime,Data$Censorstatus)~
Data$V208,data=Data)
coxfit13
summary(coxfit13)
###Testing for proportionality hazard assumption######
cox.zph(coxfit13,transform=rank)
plot(cox.zph(coxfit13,transform="km", global=TRUE))
############end births in the past one yr###############

#####Number of childeren underfive in a household V137#######
Data$V137<-as.factor(Data$V137)
Data$V137= as.factor(Data$V137)
Data$V137####n
coxfit14=coxph(Surv(Data$Survivaltime,Data$Censorstatus)~
Data$V137,data=Data)
coxfit14
summary(coxfit14)
###Testing for proportionality hazard assumption######
cox.zph(coxfit14,transform=rank)
plot(cox.zph(coxfit14,transform="km", global=TRUE))
####end Number of childeren underfive in a household######
```

```
#####Mother's age group V013######################
Data$V013<-as.factor(Data$V013)
Data$V013= as.factor(Data$V013)
Data$V013####n
coxfit15=coxph(Surv(Data$Survivaltime,Data$Censorstatus)~
Data$V013,data=Data)
coxfit15
summary(coxfit15)
###Testing for proportionality hazard assumption######
cox.zph(coxfit15,transform=rank)
plot(cox.zph(coxfit15,transform="km", global=TRUE))
############end Mother's age group V013#################

#####Sex of household head V151#########################
Data$V151<-as.factor(Data$V151)
Data$V151= as.factor(Data$V151)
Data$V151####n
coxfit16=coxph(Surv(Data$Survivaltime,Data$Censorstatus)~
Data$V151,data=Data)
coxfit16
summary(coxfit16)
###Testing for proportionality hazard assumption######
cox.zph(coxfit16,transform=rank)
plot(cox.zph(coxfit16,transform="km", global=TRUE))
############end Number of childeren underfive in a househo####

#####Current working status of respondent V714############
Data$V714<-as.factor(Data$V714)
Data$V714= as.factor(Data$V714)
Data$V714####n
coxfit17=coxph(Surv(Data$Survivaltime,Data$Censorstatus)~
Data$V714,data=Data)
coxfit17
summary(coxfit17)
###Testing for proportionality hazard assumption######
cox.zph(coxfit17,transform=rank)
plot(cox.zph(coxfit17,transform="km", global=TRUE))
############end Current working status of
respondent V714###
####Source of drinking water V113##########
Data$V113<-as.factor(Data$V113)
Data$V113= as.factor(Data$V113)
Data$V113####n
coxfit18=coxph(Surv(Data$Survivaltime,Data$Censorstatus)~
Data$V113,data=Data)
coxfit18
summary(coxfit18)
```

```
###Testing for proportionality hazard assumption######
cox.zph(coxfit18,transform=rank)
plot(cox.zph(coxfit18,transform="km", global=TRUE))
#######end Source of drinking water V113###############


####Mother's occupation V717##########
Data$V717<-as.factor(Data$V717)
Data$V717= as.factor(Data$V717)
Data$V717####n
coxfit19=coxph(Surv(Data$Survivaltime,
Data$Censorstatus)~
Data$V717,data=Data)
coxfit19
summary(coxfit19)
###Testing for proportionality hazard assumption######
cox.zph(coxfit19,transform=rank)
plot(cox.zph(coxfit19,transform="km", global=TRUE))
############end Mother's occupation V717#################

####Type of birth B0#############################
Data$B0<-as.factor(Data$B0)
Data$B0= as.factor(Data$B0)
Data$B0####n
coxfit20=coxph(Surv(Data$Survivaltime,Data$Censorstatus)~
Data$B0,data=Data)
coxfit20
summary(coxfit20)
###Testing for proportionality hazard assumption######
cox.zph(coxfit20,transform=rank)
plot(cox.zph(coxfit20,transform="km",
global=TRUE))
############end Type of birth B0###################

####Religion V130################################
Data$V130<-as.factor(Data$V130)
Data$V130= as.factor(Data$V130)
Data$V130####n
coxfit21=coxph(Surv(Data$Survivaltime,
Data$Censorstatus)~
Data$V130,data=Data)
coxfit21
summary(coxfit21)
###Testing for proportionality hazard assumption######
cox.zph(coxfit21,transform=rank)
plot(cox.zph(coxfit21,transform="km",
global=TRUE))
############end Religion V130####################
```

```
####Type of Toilet Facility V116############
Data$V116<-as.factor(Data$V116)
Data$V116= as.factor(Data$V116)
Data$V116####n
coxfit22=coxph(Surv(Data$Survivaltime,
Data$Censorstatus)~
Data$V116,data=Data)
coxfit22
summary(coxfit22)
###Testing for proportionality hazard assumption######
cox.zph(coxfit22,transform=rank)
plot(cox.zph(coxfit22,transform="km", global=TRUE))
############end Toilet Facility V160######################

#############Fitting the overall Cox-PH model after data
 cleaning and  testing for the PH assumption
 with the selected variables#####################

###variables that satisfy the propazard assumption#####
coxfitO1=coxph(Surv(Data$Survivaltime,
Data$Censorstatus)~
Data$V729 +Data$B4+Data$V212
+Data$V209 +Data$V151 +Data$V013 +
Data$V113 +Data$V130 ,data=Data)
coxfitO1
summary(coxfitO1)

##model selection #################
f1=step(coxfitO1,.~Surv(Data$Survivaltime,
Data$Censorstatus)~Data$V729
+Data$B4 +Data$V212
+Data$V209 +Data$V151 +Data$V013
+Data$V113 +Data$V130
,direction="both",data=Data)
################################################

#################final cox model################
m=coxph(Surv(Data$Survivaltime, Data$Censorstatus) ~
Data$V729 + Data$B4 +
  Data$V209 + Data$V151+ Data$V013, data=Data)
summary(m)
########################################
############# Frailty Modelling ########
###############Frailty###################
####comunity frailty#########
pen=frailtyPenal(Surv(Data$Survivaltime,
```

```
Data$Censorstatus)
~Data$V729 +
Data$B4
+Data$V212 +Data$V209
+Data$V151 +Data$V013
+Data$V113+Data$V130 +cluster(Data$V001),
data=Data, Frailty=TRUE,n.knots=6,
kappa1=5000,recurrentAG=FALSE,
cross.validation=TRUE
,hazard="Splines")
pen

summary(pen)
f2=step(pen,.~Surv(Data$Survivaltime,
Data$Censorstatus)~
Data$V729 +Data$B4
+Data$V212 +Data$V209 +Data$V151
+Data$V013 +Data$V113 +Data$V130
+cluster(Data$V001),Frailty=TRUE,
n.knots=6,kappa1=5000, recurrentAG=FALSE,
cross.validation=TRUE,hazard="Splines"
,direction="both",data=Data)
######## Household frailty ###############
pen=frailtyPenal(Surv(Data$Survivaltime,
Data$Censorstatus)~
Data$V729 +Data$B4
+Data$V212 +Data$V209 +Data$V151
+Data$V013 +Data$V113 +Data$V130
 +cluster(Data$HHI),data=Data, Frailty=TRUE,
 n.knots=6,kappa1=5000,recurrentAG=FALSE
 ,cross.validation=TRUE,hazard="Splines")
pen
summary(pen)
########################################

#############Model selection#########

f1=step(coxfitO1f,.~Surv(Data$Survivaltime,
Data$Censorstatus)~
Data$V729 +Data$B4
+Data$V212 +Data$V209 +Data$V151 +Data$V013
 +Data$V113 +Data$V130
+frailty(V001),direction="both",data=Data)
###################### End Frailty ###############

#########Bayesian  modelling and inference ########
############INLA CODE#########
###INLA CODE################
```

```
library(INLA)

########Weibull Model##########
inla.surv(Data$Survivaltime, Data$Censorstatus)
formula = inla.surv(Data$Survivaltime,
Data$Censorstatus)~Data$V729+
Data$B4+Data$V212
+Data$V209+Data$V151+Data$V013+Data$V113
model=inla(formula, family="weibull", data= Data,
verbose=TRUE,keep=TRUE )
summary(model)
#################End Weibull################

########semiparametric models############
formula = inla.surv(Data$Survivaltime,
Data$Censorstatus)~Data$V729+Data$B4+
Data$V212+Data$V209+
Data$V151+Data$V013+Data$V113+Data$V130
model1 = inla(formula,family="coxph",
control.hazard=list(model="rw1", cutpoints =
c(0,900,1800)),data=Data,keep=T)
summary(model1)
############end semiparametric models############

############ Community  frailty term################
formula = inla.surv(Data$Survivaltime,
Data$Censorstatus)~ Data$V729+Data$B4+
Data$V212+Data$V209+Data$V151+Data$V013+
Data$V113+Data$V130+f(Data$V001, model="iid",
 param =c(1, 0.001), initial=0.01)
model = inla(formula, family="weibull",
data =Data,verbose=T,control.family
=list(param=c(1,0.001),fixed=F,initial = 0.3),
 control.inla = list(int.strategy="grid",
 diff.logdens=20, dz=0.5), keep=T )
summary(model)

###############Household Frailty###############
formula = inla.surv(Data$Survivaltime,
 Data$Censorstatus)~ Data$V729+Data$B4+
 Data$V212+Data$V209+Data$V151+Data$V013+
Data$V113+Data$V130+f(Data$HHI, model="iid", param
=c(1, 0.001), initial=0.01)
model2 = inla(formula,
family="weibull",data =Data,verbose=T,
control.family=list(param=c(1,0.001),
fixed=F,initial = 0.3),
 control.inla = list(int.strategy="grid",
```

```
 diff.logdens=20, dz=0.5), keep=T )
summary(model2)
###################################
```