Addis Ababa University

Office of Graduate Program

College of Natural and Computational Sciences

Department of Statistics

MULTILEVEL SURVIVAL ANALYSIS OF TIME TO AGE AT FIRST MARRIAGE AMONG WOMEN IN ETHIOPIA

By

Molalign Gualu

Advisor: Dr. Shibru Temesgen

A Thesis submitted to the Office of Graduate Programs of Addis Ababa University in Partial fulfillment of the requirements for the Degree of Master of Science in Statistics (Applied Statistics)

May, 2017
Addis Ababa, Ethiopia

Addis Ababa University

Office of Graduate Program

This is to certify that the thesis prepared by Molalign Gualu, entitled: Multilevel Survival Analysis of time to age at first marriage among women living in Ethiopia and summited in partial fulfillment of the requirements for the Degree of Master of Science in Statistics (Applied Statistics) complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

Examiner__________________________________Signature__________Date _________

Examiner_________________________________Signature___________Date _________

Advisor Dr Shibru Temesgen ________________Signature __________ Date _________

Declaration

I hereby declare that the thesis is my original work, to the best of my knowledge, this thesis has not been presented for degrees in any other University and all sources of materials used for the thesis have been duly acknowledged with proper citation.

Name: Molalign Gualu

Signature: ------------------------

Place of submission: Department of Statistics, College of Natural and computational Sciences, Addis Ababa University

Date of submission: May, 2017

This thesis has been submitted for examination with my approval as a University advisor.

Dr. Shibru Temesgen --------------------------------

Advisor's Name signature

TABLE OF CONTENT

Contents……………….……………………………..……………………………………Page

Abstract

Multilevel Survival Analysis of time to age at first Marriage among Women in Ethiopia.

Molalign Gualu
Addis Ababa University, 2017

A marriage can be viewed as a social union or legally recognized union between a man and a woman in which they are united sexually; cooperate economically, and may have children through birth or adoption. (Ikamari, 2005). Age at marriage is the age at which individuals get married and this varies across communities and individuals. (UNICEF,2005). The main objective of this study is to model time to age at first marriage amongst women in Ethiopia. The data set in this study were obtained from Demography and Health survey conducted in Ethiopia in 2011. Women's work status, religion, place of residence, head education level, women education level, head occupation, access to media and wealth index are variables which were considered as the potential determinant of time to age at first marriage in this study. In this study, we used multilevel survival models to account for the loss of independence that arises from the clustering of women in region of Ethiopia and also we used AIC to compare two different multilevel survival models. Of all 12066 women aged 15-49, 9466(78.45%) were married and the median & mean age at first marriage for women living in Ethiopia were 16 years and 16.2 years respectively, while the minimum and maximum age at first marriage observed were 8 years and 49 years respectively. Based on the result of selected model (Log logistic-Gamma shared frailty model), place of residence of women, religion of women, education level of women, wealth index of household and head education were significant at 5% level of significance. In contrast work status of women, head occupation and access to media were not significant at 5% level of significance. The clustering effect was significant for modeling time-to-age at first marriage dataset and there was heterogeneity among the regions on age at first marriage.

Keywords: Time-to-age at First Marriage, Risk Factors, Heterogeneity, Frailty, Laplace transformations

Acknowledgement

First, and foremost, I would like to thank the almighty God for giving me the opportunity to pursue my graduate study at Department of Statistics, Addis Ababa University.

I owe the deepest gratitude to Dr.Shibru Temesgen, my thesis advisor for his valuable and constructive comments and encouragements throughout my thesis writing. His stimulating support helped me in preparing my thesis within the time-frame. I also thank Central Statistical Agency of Ethiopia for making available the data used in this thesis.

List of Abbreviations (Acronyms)

ACPF	African Child Policy Forum
AIC	Akaike's Information Criterion
CNN	Cable News Network
CSA	Central Statistics Agency
DHS	Demographic and Health Survey
EAs	Enumeration Areas
EDHS	Ethiopian Demographic and Health Survey
IPPF	International Planned Parenthood Federation
IOM	Institute of Medicine
IWHP	International Women's Health Program
MGDs	Millennium Development Goals
NCTPE	National Committee on Harmful Traditional Practice of Ethiopia
NRC	National Research Council
UBOS	Uganda Bureau of Statistics
UN	United Nations
UNFPA	United Nations Population Fund
UNICEF	United Nations Children's Fund

List of Tables

page

List of Figures

CHAPTER ONE

INTRODUCTION

1.1 Background of the study

A marriage can be viewed as a social union or legally recognized union between a man and a woman in which they are united sexually; cooperate economically, and may have children through birth or adoption. Marriage is also an important family institution for the individual and the society at large. For the individual, it is a significant and memorable event in one's life cycle as well as the most important foundation in the family formation process. It is also a rite of passage that marks the beginning of an individual's separation from the parental unit, even if generations continue to be socially and economically interdependent. For the society as a whole, it unites several individuals from different families and represents the creation of a production and consumption unit as well as one for the exchange of goods and services. In addition, marriage marks the beginning to an end of the transition to adulthood as the individual separates from the parental home, even if generations continue to be socially and economically interdependent through the extended family. (Ikamari, 2005)

Marriage is usually formalized by a wedding or celebration ceremony. People marry for various reasons ranging from social, legal, love, emotional, financial, spiritual to religious reasons. There is a general call world-wide to delay marriage and to discourage premarital sex because early marriage, especially among girls, is often associated with adolescent motherhood, school dropouts, maternal morbidity and mortality, and forfeited future life opportunities for the affected individual (Pathfinder International Report, 2006; Green, Makuria and Rubin, 2009).

Age at marriage is the age at which individuals get married and this varies across communities and individuals. The term "age at early marriage" is used to refer both formal marriages and informal unions in which girl lives with a partner as if married before age of 18 (UNICEF, 2005). According to UNFPA (2006), early marriage, also known as Child marriage, is defined as "any marriage carried out below the age of 18 years, before the girl is physically, physiologically, and psychologically ready to shoulder the responsibilities of marriage and childbearing". Age at Child marriage, on the other hand, involves either one or both spouses being children and may take place with or without formal registration, and under civil, religious, or customary laws.

Early marriage has been an area of concern in Ethiopia. According to Alemu (2008), marriage in Ethiopia can be broken down into three types namely: Promissory, Child Marriage and Adolescence. However, what is clear is that the forms of marriage may vary from one community to another and circumstances (Dercon and Hoddinott, 2011). But, the payment of either bride price or receipt of dowry is important in formalizing the process in some cultures (Dercon and Hoddinott, 2011). There is a high prevalence of early marriages in Ethiopia although proportions vary by region (Alemu, 2008). One reason why the high prevalence of early marriages is a source of concern is the negative implications that they could have on health and coupled with the socio-economic consequences for Ethiopia. For example, loss of income (in the absence of alternative sources) can lead to lower investment in education (Beegle et al., 2006; Beegle and Krutikova, 2008) and as such have negative implications on good health. The problem with most of these marriages is that they are imposed on the children, especially the girls (Win, 2009). For example, some are literary abducted by either their suitors or the suitor's family (CNN, 2013). One drawback with most writings on this topic is the focus on the girl child. This would seem to imply that male children enjoy the prospect of early parenthood, which brings with it a lot of additional responsibilities, especially after offspring arrive. So far there is insufficient evidence to prove this. But again, as found by Fafchamps and Quisumbing (2005) male children tend to wait longer as they prefer to accumulate basic assets prior to marriage. Sometimes, when a parent dies the surviving children may become substitutes for the deceased adult's labor (Beegle et al., 2006; Senbet, 2010). But, it could be argued that such consequences can exist even in the absence of orphan-hood. This is plausibly true in communities that have a low life expectancy, where again children are taught household chores at very tender ages. There are household income effects (Cosic and Deb, 2010; Deb and Rosati, 2002; Jensen and Nielsen, 1997) as well as cultural ones, where a child is expected to help out with certain household tasks regardless of whether they are orphans or not.

Marriage is also seen as a way to share risk especially following an income shock (Rosenzweig and Stark, 1989). As argued prior, early marriage could lead to a child dropping out of school even without a parental death occurring in the household. The health and income effects come into play here too. Female orphans are more likely to get married early especially, if originating from poorer households. (Pathfinder International Report, 2006; Green, Makuria and Rubin, 2009).

In this study, we used multilevel survival models to account for the loss of independence that arises from the clustering of women in region of Ethiopia. Multilevel survival models were used to investigate the relationship between different potential covariates and time-to-age at first marriage for clustered survival data with a random right censoring.

1.2 Statement of the Problem

Many scholars recommend the need to conduct in-depth studies on the risk factors of age at marriage among women for both developing and developed countries. Age at early marriage is a health issue as well as a human right violation. A recent review show that girls who marry before the age of 18 were disproportionately affected by complicated pregnancies that may lead to maternal mortality and morbidity: girls aged 10–14 were five times more likely to die in pregnancy or childbirth than women aged 20– 24; girls aged 15–19 were twice as likely to die (UNICEF, 2011). A pregnancy too early in life before a girl's body is not fully mature is a major risk to both mother and baby. Also, they were more likely to experience complications of childbirth including obstetric fistula and hemorrhaging (IWHP, 2009). Mortality rates for babies born to mothers under age 20 were almost 75 percent higher than for children born to older mothers in Ethiopia. Teenage women were also twice as likely as older women to die due to complications during pregnancy and childbirth. Infants born from teenage mothers were more likely to suffer from low birth weight, and were at higher risk of dying in its first year by 60% compared with infants of mothers in their twenties (Nour, 2006). Age at first marriage had health implication for women and their under-five children (Adebowale, 2012).

Many of the studies conducted used logistic regression analysis and Cox proportional hazard models to estimate the effect of covariates on the age at first marriage; which restricts attention to the events that occur within the shortest time observed and the correct inference based on Cox's models needs identically and independently distributed samples respectively. Logistic regression does not account the censoring observations i.e., does not hold for time-to event data; however, survival analysis is more powerful than Logistic framework that takes censoring into considerations. But here we want to use multilevel survival models since these models permit the analysts to account for the loss of independence that arises from the clustering of subjects in higher level units. Similarly, it allows researchers to make valid inferences when examining the effect of both subject characteristics and cluster characteristics on the risk of the occurrence of the outcome. Multilevel survival model used to model survival data when there are repeated measures on a subject, subjects nested within some other hierarchy, or some other reason to have both fixed and random effects (Crowther et al,2014).

The study focuses on the modeling and identifying the impact of demographic and socio economic factors on age at first marriage. The research questions are:-

- What are the determinants or risk factors of age at first marriage amongst women in Ethiopia?
- How to compare various parametric shared frailty model that is used in modeling time to age at first marriage for the dataset?
- Which baseline distributional assumption among the Exponential, Weibull and log-logistic well describe the age at first marriage dataset?
- Which frailty distribution among the Gama, Positive stable and inverse Gaussian distributions that best describe the age at first marriage for the dataset?

1.3 Objective of the Study

1.3.1 General Objective

The general objective of this study is to model time to age at first marriage.

1.3.2 Specific Objective

The specific objectives of this study are to:-

- ➢ Identify the appropriate parametric distribution for baseline hazard of interest that is used in modeling time to age at first marriage for the data set.
- ➢ Identify the appropriate distribution for frailty that is used in modeling time to age at first marriage for the data set.
- ➢ Identify significant factors or covariates that are associated with time-to-age at first marriage for the data set.
- ➢ Compare two different multilevel survival models that are used in modeling time to age at first marriage for the data set.

1.4 Significance of the Study

The results are expected to give some knowledge about:-

- Prolonging time to age at first marriage for women in Ethiopia based on each categorical predictor.
- The estimated variability (unobserved heterogeneity) in the population (women aged 15-49 in Ethiopia) of clusters (regions) based on selected model.
- The key socio-economic and demographic predictors of age at first marriage in Ethiopian women.
- Policy and strategies designation for government and other concerned bodies.

More generally the study provides information on marriage in Ethiopian women by analyzing the impact of different covariates on survival of age at first marriage and the study will also add to the existing literature on the determinant of time to age at first marriage, that is, it provides an input for further study in Ethiopia.

CHAPTER TWO

LITERATURE REVIEW

2.1 An Overview of Marriage

According to Demographic and Health Surveys (DHS), which provide much of the current country-level child marriage data, age at child marriage is most common in the world's poorest countries. The highest rates are in sub-Saharan Africa and South Asia as well as parts of Latin America and the Caribbean (NRC/IOM, 2005). A UNICEF study found that 48 percent of women were married before 15 years and 24 were married before 18 years in South Asia. The prevalence of age at early marriage is 42 percent in Africa (UNICEF, 2005) and more than 60 percent in some parts of East and West Africa (IPPF and UNFPA, 2006). In Latin America and the Caribbean, prevalence is 29 percent, though some individual countries have much higher rates of age at early marriage (UNICEF, 2005). Also age at child marriage is common in the Middle East, where nearly half of girls younger than 18 in Yemen and Palestine are married (IPPF and UNFPA, 2006). In sub-Saharan Africa, for example, 21 of 30 countries have seen an increase in the national age at marriage over the past several decades (Westoff, 2003). However, this increase in the age at marriage is occurring slowly and unevenly within countries. According to (UNICEF's, 2011) figures, 66 percent of Bangladesh girls are married before the age of 18 and approximately a third of women were married by the age of 15 ; although the legal age at first marriage for females in Bangladesh is 18 years. The highest rates of child marriage are found in West Africa, in countries such as Niger, Chad, and Mali. However, in East Africa, the numbers of girls married in countries such as Ethiopia, Zambia, and Tanzania is also substantial. In rural Tanzania, median age at marriage is 18.5. The Demographic and Health Survey (DHS) for 1995 to 2003 shows that in Niger, 47 percent of women aged between 20 and 24 were married before the age of 15, and 87 percent before the age of 18, a total of 53 percent had also had a child before the age of 18.

The 1992, 2000, and 2006/7 Namibia DHS report showed that mean age at marriage was 24 in 1992, 26.2 in 2000 and 28.6 in 2006/7.In Uganda, marriage is almost universal sooner or later, everyone marries, an early age at first marriage is observed for both males and females. According to the 1995, 2000/01 and 2006 Uganda Demographic and Health Surveys, the age at first marriage has been 17.5, 17.8 and 17.8 respectively and coupled with a low contraceptive prevalence rate of 24%, they have led to a high total fertility rate of 6.9. In the effort to increase the age at first marriage, Uganda has tried to intervene by setting the minimum legal age for a woman to get married at 18 years and through emphasis on educating the girl child through a number of educational reforms instituted since 1990 (Ministry of Education and Sport,2003). However, not all girls of school going age are enrolled in schools, there are high girl child drop-out rates and entry into marriages at early ages is still high. According to the 2002 Uganda Population and Housing Census 6,308,849 girls marry below 16 years (UBOS and Macro International Inc, 2007) and this leads to low education attainment among women and unplanned pregnancies and high fertility.

Table1, Legal age at first marriage for females in selected sub-Sahara African countries (ACPF,2012)

No	Country	Legal age at first marriage	No	Country	Legal age at first marriage
1	Angola	18	12	Eritrea	18
2	Benin	18	13	Ethiopia	18
3	Botswana	18	14	Gabon	15
4	Burkina Faso	17	13	Malawi	15
5	Cameron	15	14	Mozambique	18
6	Cape Verde	18	15	Namibia	18
7	Chad	17	16	Rwanda	21
8	Comoros	18	17	Senegal	16
9	Congo(Brazzaville)	18	18	Tanzania	15

10	Democratic republic of Congo	15	19	Uganda	18
11	Djibouti	18	20	Zimbabwe	18

Bayisenge (2010) observed that African women in general marry at a much earlier age than their non-African counterparts, leading to early pregnancies. In average, age at first marriage is relatively high, compared with developed countries and many other developing countries. Ethiopia has one of the highest rates of age at early marriage in Sub-Saharan Africa. A study by the National Committee on Harmful Traditional Practices of Ethiopia (NCTPE) estimated the proportion married before the age of 15 are 57 percent. The same study shows that the practice occurs in its most extreme forms in northern Ethiopia, where girls are married as young as eight or nine years of age. Although age at early marriage is widely practiced in many parts of the country, rates in Amhara and Tigray region are much higher than the national average (82 percent in Amhara, 79 percent in Tigray, 64 percent in Benshangul, 64 percent in Gambella and 46 percent in Afar) (NCTPE, 2003). A recent study conducted in two woreda's of the Amhara region also shows that 14 percent of women were married before age of 10 years, 39 percent before age of 15 years, and 56 percent before age of 18 years (Population Council, 2004). Age before 18 years marriage stands in a direct conflict with the objectives of the Millennium Development Goals (MDGs) (Mathur *et al.*, 2003). It threatens the achievement of MDGs such as eradicating extreme poverty and hunger, achieving universal primarily education, promoting gender equality and empowering women, reducing child mortality, improving maternal health and combating HIV/AIDs, malaria and other diseases (UN, 2007).

In Ethiopia, although there remain distinctive ethnic differences in access to education, rural – urban migration and marriage practices (age at marriage and the prevalence of polygamy and divorce), entry into marriage is near universal among all groups, with only 1 percent of men and women age 35 and above having never married (CSA, 2001:77). Marriage is of central importance to all aspects of life in Ethiopia; in one way or another, practically all essentials are organized, procured, and guaranteed through the institution of marriage (Weissleder, 1974:72). A strict sexual division of labor that makes the performance of tasks not of one's gender almost taboo provides a compelling pragmatic rationale for entry into marriage. For women, in particular, being single or in a household without a man is associated with marginalized social status, dependence on kin, and greater vulnerability (Pankhurst, 1992).

Among the Amhara, who for centuries have been the most dominant cultural and political group, very early age at marriage is common. According to the 2005 Ethiopia DHS the median age at first marriage for women in the Amhara region ages 20-49 was 14.4 years compared to a median of 17.1 years among women in the Oromiya region (CSA, 2006:83).

But the generally acknowledged minimum age-at-first marriage in Ethiopia is currently 18 years (Erulkar and Muthengi, 2009).

Parents view early marriage strategically because it provides a means to extend the family's social networks, which are a critical source of aid during times of crisis and household need.

Because first marriages generally involve a bond between households, a bride's virginity is not simply a matter of honor; it has an economic value to parents and to the young women themselves (Pankhurst, 1992:122). In societies, such as Ethiopia, where family networks function as mutual support groups, how well a young woman and man marries has long-term consequences for the families involved as well as for the bride and groom.

According to Dagne (1994:36) the competition to find desirable partners for one's own children means that the earlier a marriage is arranged, the less parents have to worry about. At the same time, depletion of family resources associated with war, political turmoil and economic and environmental crisis has made it more difficult for families to secure a suitable husband for their daughters, and for young men to attain the economic independence desirable in a marriage partner. To the extent that marriage is delayed, individual autonomy in partner selection is likely to be greater for both men and women.

Because grooms bring most of the assets into a marriage, their outcome in the marriage market is not as important in determining their future economic well-being as it is for brides (Fafchamps and Quisumbing, 2005a).

Marriages in many parts of Ethiopia can be divided into six types: ceremonial marriage (serg), religious marriage (k'urban), civil marriage (semanya), marriage proceeded by the provision of labor (k'ot'assir), paid labor marriage (gered or demoz), and marriage by abduction (t'ilf). The types of marriages differ in terms of the involvement of parents in the match; the level of formality, ceremony and expense; and expectations of labor exchanges (Pankhurst,1992:106-07). Marriage by abduction and civil marriage are now the standard forms of marriage, although ceremonial marriage which involves considerable expense remains common in urban areas. In rural areas arranged marriages are the norm whereas abduction marriage provides a socially acceptable way to circumvent the parents' or the bride's disapproval of a match (Fafchamps and Quisumbing, 2002).

While there are strong social and economic pressures on girls to comply with their parents' desires, there are also opportunity costs and risks associated with early marriage and the early initiation of sexual intercourse, especially premarital sexual intercourse that does not lead to marriage. Very early age at first marriage and premarital first sex are associated with marital instability and divorce, multiple partners; poverty, and subsequent drift into prostitution or paid domestic work (Duncan *et al*., 1993).

2.2 Early Marriage and Health Consequence

Evidence from South Africa in a study by Yamauchi (2007) indicates that "education reduces the probability of early marriage but increases the probability of contracting HIV". However, it can be argued that the correlation could be influenced by unobservable factors such as the culture and norms of the community under study. The findings may not hold in highly conservative societies, where pre-marital sex is frowned upon (Lindstrom et al., 2009).

In a study on ‘education and health’ Vogl (2012) posits a positive link between parental education and a child’s health. The argument is that educated parents tend to be healthier than the less educated ones and therefore, are more likely to have healthier children. Therefore, one of the benefits of increasing the level of education is that, it assists in delaying a child’s transition into an early marriage arrangement. By implication there is therefore, a benefit in a child avoiding early marriage by extending their stay in school.

Early marriage increases the risk of divorce (Andersson, 1997) according to economic theories as they argue the partners do not spend enough time and energy for finding an optimal spouse and they do not possess the necessary emotional, educational and economic resources required for a marriage (Martin-Bumpass, 1989)

The propensity to marry, the stability and duration of marriage have considerable implications for the organization of family life. The age at first marriage may also influence population growth, labour supply, consumption, wage rates, mortality, migration and to some extent fertility (Mensch, Singh and Casterline, 2005).

Women who marry early will have, on average; a longer period of exposure to the risk of pregnancy, often leading to higher completed fertility. Variation in age of entry into marriage helps explain differences in fertility across populations and also helps explain trends in fertility within individual populations over time (UN,1990 ; Ezeh,A.C.,&Dodoo,F.N.,2001).

Therefore, age at first marriage has a direct bearing on fertility behavior (Davis,K.&Blake,J., 1956;Coale,A.,1971;Bongaarts,J.,1983).

CHAPTER THREE

DATA AND METHODOLOGY

3.1 Data Source

The data set in this study was obtained from Demographic and Health Survey conducted in Ethiopia in 2011, which was the third comprehensive survey conducted as part of the worldwide Demographic and Health Surveys project. The data provide in-depth information on marriage, fertility, family planning, infant, child, adult and maternal mortality, maternal and child health, gender, nutrition, malaria, knowledge of HIV/AIDS and other sexually transmitted diseases.

3.2 Sample Design

The 2011 EDHS sample was selected using two stage cluster design and census enumeration areas (EAs) were the sampling units for the first stage. The sample included 624 EAs, 187 in urban areas and 437 in rural areas. Households comprised the second stage of sampling. A complete listing of households was carried out in each of the 624 selected EAs from September 2010 through January 2011. A representative sample of 17,817 households was selected for the 2011 EDHS. Of these, 16,702 were successfully interviewed. In the interviewed households 17,385 eligible women were identified for individual interview; complete interviews were conducted for 16,515. Women whose current ages are 15-49 years are included in the survey. After a certain rearrangement and reorganization the total number of women with complete information became 12,066.

3.3 Variable in the Study

3.3.1 Dependent Variable

The dependent variable is the time to age at first marriage. It is measured as the length of time from birth until the age at first marriage which is measured in years. On a sample of all Ethiopian women aged 15-49, we retrospectively observe the timing to first marriage since birth. Hence we have to consider two things. First, all cases with no observed events are right censored. Therefore the women who had not yet experienced the event of interest resulting in right censoring of the data. There is no reason for this censoring pattern to be dependent on the survival times and we consider it uninformative. Second, in order to make censoring valid, we have to assume that all women marry before the age of 50.

3.3.2 Independent Variable

The independent variables considered in this study are respondents work status, religion, type of residence, Head education level, women education level, Head occupations, access to media and wealth index. For description and categories of explanatory variable see appendix VIIII.

3.4 Method of Data Analysis

3.4.1 Survival Analysis

Survival analysis consists of studies of the survival time of a subject (usually measured in days, weeks, months, or years), which is the time that elapses between the baseline and the moment an adverse event occurs, or the subject drops out of the trial. Sometimes the survival time is called a lifetime or an event time.

The survival times for subjects who dropped out of the trial (called dropouts or lost to follow-up subjects*)* are right-censored (or, more simply, censored). The survival times of the subjects who remain in the trial until it ends are censored *as* well. This term applies to situations when it is known that the subject survived a certain length of time and was healthy, but the later health condition for this subject is not recorded. Censored survival times represent very important information and should be kept in the database. Retained censored survival times increase the overall survival rate of the subjects-that is, the percentage of people who are alive for a given period of time. For example, if a subject drops out after being in a study for 5 months, the subject is still included in calculation of the survival rate up to 5 months. Naturally, a higher survival rate implies a better treatment efficacy.

In what follows, each uncensored observation is termed "death," regardless of whether a death or a different adverse event has occurred. Denote by T the random variable representing the survival time of a subject. Let f(t), $t \geq 0$, denote the probability density function (pdf) of T , and let $F(t) = P(T \leq t) = \int_0^t f(x)\,dx$, $t \geq 0$, be the cumulative distribution function (cdf) of T. The distribution of T is called the survival time distribution (or the lifetime distribution).

The objective of survival analysis is to estimate and model the following functions:

- The survival function, S(t), defined as the probability that a subject survives up to time *t:*

$$S(t) = P\,(T > t) = \int_t^\infty f(x)\,dx = 1 - F(t), \quad t \geq 0 \quad \ldots\ldots\ldots\ldots\ldots..(3.1)$$

- The hazard function, h(t), defined as the following ratio:

$$h(t) = \frac{f(t)}{s(t)} \quad , \quad t \geq 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.2)$$

It is interpreted as an instantaneous death rate, since the probability that the event occurs within small time interval [t, t+dt), given that the subject survived up to time t, $t \geq 0$, is equal to

$$P\,(T < t + dt\,|T > t) = \frac{P(t < T < t + dt)}{P(T > t)} = \frac{f(t)dt}{s(t)} = h(t)dt \quad \ldots\ldots(3.3)$$

- The cumulative hazard function, H(t), defined by

$$H(t) = \int_0^t h(x)dx \quad , \quad t \geq 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.(3.4)$$

3.4.2 Non-Parametric Survival Analysis

Non-parametric survival analyses are more widely used in situations where there is doubt about the exact form of distribution. In survival analysis, the data are conveniently summarized through estimates of the survival function and hazard function. The estimation of the survival distribution provides estimates of descriptive statistics such as the median survival time. These methods are said to be non-parametric methods since they require no assumptions about the distribution of survival time. The Kaplan-Meier, Nelson-Aalen and Life Tables are the most widely used to estimate the survival and hazard functions (Collet, 1994).

3.4.3 Estimation of Survival function by the Kaplan-Meier Method

A widely used method for estimation of the survival function is the Kaplan Meier method. This method produces the Kaplan-Meir estimator, a nonparametric estimator, which does not assume any known algebraic form of the estimated survival function. The Kaplan-Meier estimator is also referred to as the KM estimator or the product-limit estimator. Suppose k distinct survival times are observed. Arranged in increasing order, they are $t_1 < t_2 < ... < t_k$. At time t_i, there are n_i subjects who are said to be at risk-that is, they survived up to this time (not including it) and were not censored. Denote by d_i the number of subjects who have an event at time t_i. To simplify notation, let $t_o = 0$ and $d_o = 0$. Then the Kaplan-Meier estimator of the survival function S (t) is

$$\hat{S}(\mathrm{t}) = \prod_{i:t_i \leq t} (1 - \frac{d_i}{n_i}) , \mathrm{t} \geq 0 \qquad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots. (3.5)$$

3.4.4 The Kaplan-Meier Survival Curve

The Kaplan-Meier survival curve is the plot of the Kaplan-Meier estimator of the survival function $\hat{S}(\mathrm{t})$ against time t. This curve is a step-function that decreases at the times of events. The censored times are usually marked by a cross (x). If an event and a censoring occur at the same time, a cross for the censored observation is put at the bottom of the step.

3.4.5 Median Survival Times

Use of the Kaplan-Meier estimator is not restricted to estimating survival probabilities for given times t. It may also be used to estimate fractiles such as the median survival time. Consider the p^{th} fractile ξ_p of the cumulative distribution function $F(t) = 1-S(t)$, and assume that F(t) has positive density $f(t)=F'(t)=-S'(t)$ in a neighborhood of ξ_p. Then ξ_p is uniquely determined by the relation $F(\xi_p)= p$, or equivalently, $S(\xi_p)=1-p$. The Kaplan-Meier estimator is a step function and hence does not necessarily attain the value $1-p$. Therefore a similar relation cannot be used to define the estimator $\widehat{\xi p}$ of p^{th} fractile. Rather we define $\widehat{\xi p}$ to be the smallest value of t for which $\hat{S}(t) \leq 1-p$, that is, the time t where $\hat{S}(t)$ jumps from a value greater than $1-p$ to a value less than or equal to $1-p$.

Hence the median survival times ($\xi_{0.5}$) to be the smallest value of t for which $\hat{S}(t) \leq 0.5$, that is, the time t where $\hat{S}(t)$ jumps from a value greater than 0.5 to a value less than or equal to 0.5.

3.5 Multilevel survival analysis

Multilevel survival analysis is the statistical technique that can apply for clustered (grouped) survival times. Researchers often encounter grouped or multilevel data like individuals are nested within families, and families are nested within neighborhoods. In our study also encountered such kinds of data. For instance women aged 15-49 nested with in region. Analyzing such data requires special treatment because most multivariate models assume that observations are independent, and grouped data clearly violate this assumption. Statisticians and biomedical researchers identified adverse consequences of applying the Cox regression to grouped survival times (Andersen & Gill, 1982; Prentice, Williams, & Peterson, 1981). They noted that when the independent assumption of the Cox model is violated, the tests of statistical significance are biased and in ways that cannot be predicted beforehand (Wei, Lin, & Weissfeld, 1989).

Mixed effects cox regression models, mixed effect piecewise exponential survival models and discrete time survival models with mixed effects are the statistical models for multilevel survival analysis. Hence in this study we concentrate on mixed effects cox regression models and discrete time survival models with mixed effects. Thus, one specific objective of this study were to compare a special case of mixed effects cox regression models (parametric shared frailty model) and discrete time survival models with mixed effects.

3.5.1 Mixed Effects Cox regression Models

Mixed effects cox regression models are used to model survival data when there are repeated measures on an individual, individuals nested within some other hierarchy, or some other reason to have both fixed and random effects. Mixed effect model allow the model to have multiple random effects, whereas frailty models allow model with only random intercept (Crowther et al, 2014). That is why they say parametric shared frailty model is a special case of mixed effects cox regression models. Moreover, parametric shared frailty model is a special case of mixed effects cox regression models due to the fact that as it assume a parametric distribution for baseline hazard function and it consider shared frailty as cluster-specific random effects. Suppose individuals are nested in one of G groups or clusters. A mixed effects cox regression model can be formulated as:-

$$h_i(t) = h_0(t) exp^{(x_i\beta+\alpha_j)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.6)$$

Where α_j denotes the random effects associated with the j^{th} cluster. Robe-Hesketh and Skrondal use the term 'shared frailty' to denote the exponential of the random effect: $\exp(\alpha_j)$. The random effect can be thought of as a random intercept that modifies the linear predictor, while the shared frailty term has a multiplicative effect on the baseline hazard function:

$$h_i(t) = h_0(t)\exp(\alpha_j)\, exp^{(x_i\beta+\alpha_j)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.7)$$

For further detail about parametric shared frailty model see on 3.6.

3.5.2 Discrete Time Survival Models with Mixed Effects

Discrete time survival models with mixed effects is a survival model in which survival time is measured in discrete values or when it is continuous the time scale is divided into intervals and the hazard function is not assumed to be constant with in each interval. Additionally, in fitting the discrete time survival model with mixed effects, each subject's duration of exposure (at risk time) during the interval is does not taken into account and simply consider whether the subject experienced an event within the given interval.These models use a discrete version of the hazard function. Binomial regression models, with a logit, probit or complementary log–log link function can be used to model the probability that the event occurred at a specified discrete time point, conditional on the fact that it had not yet occurred (Rabe-Hesketh & Skrondal, 2012b). Moreover, suppose that we have M intervals, defined by M+1 cut point's; τ_0 , τ_1…….,τ_m (where $\tau_0 = 0$ and $\tau_m = \infty$). In interval M, given by [τ_{m-1} , τ_m), the hazard function for a given subject is not assumed to be constant (i,e it allow to vary). In this study we were considered five discrete intervals by dividing age at first marriage into the interval [5-15), [15-25), [25-35), [35-45) & [45-55) to construct discrete time survival model with mixed effects. If each women experience the event of interest (marriage) in one of each mutual exclusive intervals the event indicator would be 1 and otherwise 0. It has a binomial sense, and it is the reason why we were used Complementary log-log binomial linear model with mixed effects among various choice of discrete time survival model with mixed effects. A regression model for binary outcomes can then be used to model the probability of the occurrence of an event within each interval. Possible link functions for the generalised linear model are the logit link function, the probit link function and complementary log–log link function (Rodriguez, 2001; Allison, 1995; Goldstein, 2010a). An advantage to the latter is that the resultant regression coefficients are identical to those of an underlying proportional hazards regression model (Allison, 1995; Rabe-Hesketh & Skrondal, 2012b). Thus, the estimated coefficients can be interpreted as having a relative effect on the hazard of the occurrence of the event. (Austin,P.C, 2017). In this study, we were used the complementary log–log model to model the occurrence of an event during each time interval.

3.6 Modeling Frailty and Shared Frailty Model

Expanding proportional hazards model to include a random effect, called a frailty, allows for modeling association between individual failure times within a group. A frailty acts multiplicatively on the hazard function and the model that incorporates this random effect into the hazard function is called the frailty model. There are two different, but related, connotations of frailty. First, frailty is the missing covariates that are not known to us and consequently they are unobservable. More specifically, let *Z* denote the covariate vector that is known to us and w denote the covariate vector that is unknown.

The hazard function for a given individual is

$$h(t \mid Z) = h_0(t) \exp(\beta Z + \psi w), \ldots \ldots \ldots . \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots . (3.8)$$

where ψ is the regression coefficient of unknown covariates.

To simplify Eq. (3.8), let u = exp(ψw), then the hazard function has the form

$$h(t \mid Z) = h_0(t) u \exp(\beta Z), \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (3.9)$$

where u is a random variable assumed to have a one-dimensional distribution q.

In Eq. (3.9), the frailty u represents the total effect on failure of the covariates not measured when collecting information on individuals. Eq. (3.9) is known as the frailty model.

Clayton (1978) suggested the other connotation of the frailty when individuals in a study are divided into distinct groups. Here, the frailty denotes unobservable common covariates shared by members in a group, and the frailty model handles the dependence generated by those common covariates. For example, for a study including husband and wife, each couple shares common environmental factors; for menozygotic twins study, twins share a common genotype as well as common environmental factors. Specifically, suppose there are G groups with n_i individuals in the i th group; Z_{ij} is the observable covariate vector for the j^{th} individual in the i^{th} group. Let w_i be the unobservable covariates for the i^{th} group and ψ be its regression coefficient.

The hazard function of the j^{th} individual in the i^{th} group is

$$h_{ij}(t/Z_{ij}) = h_0(t)\exp\left(\beta Z_{ij} \ + \psi W_i\right), i = 1, \ldots, G, j = 1, \ldots, n_i \ldots \quad \ldots\ldots\ldots (3.10)$$

Replacing $\exp(\psi w_i)$ by u_i, which is the frailty of the i^{th} group, the hazard function incorporating frailty reduces to

$$h_{ij}(t/Z_{ij}) = h_0(t)u_i\exp\left(\beta Z_{ij}\right), i = 1, \ldots, G, j = 1, \ldots, n_i \ . \ . \quad \ldots\ldots\ldots\ldots (3.11)$$

Here it is assumed that $u_1, \ldots, u_G$ are random variables with the common probability density function q. The model (3.11) can be considered as a random effects model with two sources of variation. There is a group variation, described by the random variable u with the probability density function q. Secondly, there is the individual variation described by the hazard function $h_0(t) \exp(\beta Z_{ij})$.

In (3.11), members in a group share the same frailty, so the frailty model under this circumstance is known as the shared frailty model. Also, in this model, groups with a large value of the frailty will experience the failure at earlier times than groups with small values of the frailty. Hougaard (2000)

3.7 Inference for the Shared Frailty Model

As in most contexts, provided that one trusts the model (3.10), maximum likelihood method is the method of choice. To derive the general form of the likelihood function, it is assumed that the common factor causes dependence between individuals in a given group, and conditional on that, all individuals within the group are independent. Thus, for one group of *n* individuals, the conditional joint survival distribution of failure times $T_1, T_2, \ldots, T_n$ is given by

$$P(T_1 > t_1, \ldots, T_n > t_n|u) = P(T_1 > t_1|u)P(T_2 > t_2|u) \ldots . P(T_n > t_n|u)$$

$$= exp\left\{-u\sum_{j=1}^{n} H_0(t_j)\exp(\beta Z_j)\right\} \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots . \ldots \ldots \ldots \ldots \ldots \quad (3.12)$$

Note that we omit the group index *i* in Eq. (3.12). The above joint conditional survival distribution holds for any group. Integrating the frailty out, we get the joint survival function for this group as

$$S(t_1, \ldots, t_n) = P(T_1 > t_1, \ldots, T_n > t_n) = \int_0^\infty P(T_1 > t_1, \ldots, T_n > t_n|u)q(u)\,\mathrm{d}u$$

$$= \int_0^\infty \exp\left\{-u\sum_{j=1}^{n} H_0(t_j)\exp(\beta Z_j)\right\} q(u)\,\mathrm{d}u$$

$$= \mathrm{LP}\left[\sum_{j=1}^{n} H_0(t_j)\exp(\beta Z_j)\right] \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots . (3.13)$$

where *LP* is the Laplace transform of the density function *q* and $H_0(t) = \int_0^t h0(u)\,\mathrm{d}u$. From Eq. (3.13) it is clear that the joint survival function for one group is the Laplace transform of the frailty density function *q* with parameter $\sum_{j=1}^{n} H_0(t_j)\exp(\beta Z_j)$. In principle, any distribution on the positive numbers can be applied as a frailty distribution. In this paper, we concentrate on the gamma, the inverse Gaussian and the positive stable distributions.

From Eq. (3.13) one can derive the likelihood function for one group as follows: If the failure time is observed for the j th individual at time t_j, its probability is given by

$$P\left(T_j = t_j, T_1 > t_1, T_2 > t_2, \ldots.\right) = -\frac{\partial S(t_1, \ldots., t_n)}{\partial t_j}$$

$$= -h_0\left(t_j\right)\exp(\beta Z_j) LP^{(1)}\left(\sum_{j=1}^{n} H_0(t_j)\exp(\beta Z_j)\right) \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (3.14)$$

where $LP^{(1)}(s)$ denotes the first derivative of $LP(s)$ with respect to s. Let $D. = \sum \delta j$, the total number of failures in the group, and θ be the parameter of the frailty distribution. Then, using Eq. (3.14), the likelihood for one group is given by

$$= (-1)^{D.}\left\{\prod_{j=1}^{n} h_0(t_j)^{\delta_j} \exp(\delta_j \beta Z_j)\right\} LP^{(D.)}\left(\sum_{j=1}^{n} H_0(t_j)\exp(\beta Z_j)\right) \ldots \ldots \ldots \ldots \ldots \ldots .\ldots .. (3.15)$$

The likelihood function for all individuals is constructed by multiplying the group likelihoods together. Specifically, if D_i denotes the number of failures in the i th group, and $D = \sum_{i=1}^{G} Di$, then the likelihood function is given by

$$= (-1)^{D} \prod_{i=1}^{G}\left\{\prod_{j=1}^{n_i} h_0(t_{ij})^{\delta_{ij}} \exp(\delta_{ij} \beta Z_{ij})\right\} LP^{(D_i)}\left(\sum_{j=1}^{n} H_0(t_{ij})\exp(\beta Z_{ij})\right) \ldots \ldots .. (3.16)$$

If we assume a parametric form for h_0, we can handle the estimation in the usual way by differentiating the log likelihood function. If a parametric form is not assumed for h_0, there are several estimation methods like full conditional approach and EM algorithm available to handle this semi-parametric model.

3.7.1 The Gamma Frailty Model

Clayton (1978) proposed the gamma distribution for the frailty. Since then, the gamma frailty model has been used extensively because the derivatives of its Laplace transformation are quite simple. From a computational and analytical point of view, it fits very well to failure data. It is widely used due to mathematical tractability (Wienke, 2011).

The density function of the frailty is

$$q(u) \; = \frac{u^{1/\theta-1}\exp(-u/\theta)}{\Gamma\,(1/\theta)\theta^{1/\theta}} \; \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots..(3.17)$$

Where θ>0 and u>0 indicates that individuals in group are frail, whereas u<0 indicates that individuals are strong and have lower risk. The corresponding Laplace transform is given by; *LP(s)* $=(1+\theta s)^{-1/\theta}$. Usually, we use the one-parameter gamma distribution denoted by Gamma*(θ)*. Thus the mean of the frailty is 1, which is the desired property of the frailty distribution; the variance is θ, which reflects the degree of dependence in the data. Large θ indicates strong dependence. The conditional survival function of the gamma frailty distribution is given by: (Gutierrez, 2002).

$$S_\theta(\mathrm{t}) = [(1-\theta\ln\{\mathrm{S(t)}\})]^{-1/\theta}\,, \theta > 0 \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(3.18)$$

The conditional hazard function of the gamma frailty distribution is given by: (Gutierrez, 2002)

$$h_\theta(\mathrm{t}) = \mathrm{h(t)}[1-\theta\ln\{\mathrm{S(t)}\}]^{-1} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots.\dots\dots\dots\dots..(3.19)$$

where S (t) and h(t) are the survival and the hazard functions of the baseline distributions.

The gamma model has predictive hazard ratios which are time invariant (Fine et al., 2003)

For the Gamma distribution, the Kendall's Tau (Hougaard, 2000), measures the association between any two event times from the same cluster in the multivariate case and given by:-

$$\tau = \frac{\theta}{(\theta+2\;)}\,,\text{where } \tau\epsilon\ (0,1) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \; (3.20)$$

One can derive the likelihood function as follows. The p^{th} derivative of the Laplace transform is

$$LP^{(p)}(s) = (-1)^p\theta^p(1+\theta s)^{-\frac{1}{\theta}-p}\Gamma\frac{\frac{1}{\theta}+p}{\Gamma}\left(\frac{1}{\theta}\right)\dots\dots\dots\dots\dots\dots\dots\dots\dots..\dots..(3.21)$$

Following Eq. (3.16), the likelihood for all individuals is given by

$$= (-1)^D \prod_{i=1}^{G} \frac{\theta^{D_i}\Gamma\,(\frac{1}{\theta} + D_i)}{\Gamma\,(\frac{1}{\theta})} \left\{ \prod_{j=1}^{n_i} h_0(t_{ij})^{\delta_{ij}} \exp(\delta_{ij}\beta Z_{ij}) \right\}$$

$$\times \left\{ 1 + \theta \sum_{j=1}^{n_i} H_0(t_{ij}) \exp(\beta Z_{ij}) \right\}^{-\frac{1}{\theta}-D_i} \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots .\,(3.22)$$

For gamma frailty model the marginal hazards are not proportional over time.

3.7.2 The Positive Stable Frailty Model

The positive stable frailty model assumes the positive stable distribution for the frailty, see Hougaard (2000). For most frailty models, the marginal hazard functions are not proportional. The positive stable frailty model is an exception. This is an advantage of the positive stable frailty model. Suppose the frailty has the positive stable distribution with parameter θ. We restrict $0 < \theta \leq 1$ to get a distribution with positive numbers. Here a small value of θ indicates a strong dependence in the data. If θ is close to 1, individuals in the study are independent of each other, and the frailty model is not necessary.

Note that the positive stable (PS) model has the attractive feature that predictive hazard ratio decrease to 1 over time (Oakes, 1989).

The density is

$$q(u) \;=\; -\frac{1}{\pi u} \sum_{l=1}^{\infty} \frac{\Gamma\,(l\theta \;+\; 1)}{l!} \left(-u^{-\theta}\right)^{l} \sin\,(\theta l \pi), \dots \dots \dots \dots \dots \dots \dots \dots \dots .\,(3.23)$$

with Laplace transform $L(s) = \exp(-s^{\theta})$. The marginal distribution of T_j is given by

$$P\left(T_j > t\right) = exp\left\{-H_0\left(t_j\right)^{\theta} \exp(\theta\beta Z)\right\} \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots .\,.\, \dots \dots \dots \dots \dots \;(3.24)$$

Thus, the integrated hazard and the hazard function are $H_0\left(t_j\right)^{\theta} \exp(\theta\beta Z)$ and $\theta h_0(t_j) H_0\left(t_j\right)^{\theta-1} \exp(\theta\beta Z)$ respectively.

The p^{th} derivative of Laplace transform is

$$L^{(P)}(s) = (-1)^{\theta} \exp\left(-s^{\theta}\right) \sum_{m=1}^{p} C_{p,m} \theta^m s^{m\theta - p} \ldots\ldots\ldots\ldots\ldots\ldots.\ldots\ldots\ldots\ldots\ldots. (3.25)$$

where $C_{p,m}$ is a polynomial in θ of degree m and is defined recursively by

$$C_{p,p} = 1, C_{p,1} = \Gamma\,(p - \theta)/\Gamma\,(1 - \theta),\ \&\ C_{p,m} = C_{p-1,m-1} + C_{p-1,m}\{(p - 1) - m\theta\}$$

Following Eq. (3.16), the likelihood function for all individuals is given by

$$= \prod_{i=1}^{G} \left\{ \prod_{j=1}^{n_i} h_0(t_{ij})^{\delta_{ij}} \exp\left(\delta_{ij}\beta Z_{ij}\right) \right\} exp \left\{ -\left[\sum_{j=1}^{n_i} H_0\left(t_{ij}\right) \exp\left(\beta Z_{ij}\right) \right]^{\theta} \right\}$$

$$\times \sum_{m=1}^{D_i} C_{D_i,m} \theta^m \left[\sum_{j=1}^{n_i} H_0\left(t_{ij}\right) \exp\left(\beta Z_{ij}\right) \right]^{\theta m - D_i} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. (3.26)$$

3.7.3 Inverse Gaussian Frailty Model

Similar to the gamma frailty model, simple closed-form expressions exist for the unconditional survival and hazard functions, this makes the model attractive. The probability density function of an inverse Gaussian shared distributed random variable with parameter θ > 0 is given by

$$f_u(u) = \left(\frac{1}{2\pi\theta}\right)^{1/2} u^{-3/2} exp\left(\frac{-(u-1)^2}{2\theta u}\right), \theta > 0, u > 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (3.27)$$

For identifiability, we assume u has expected value equal to one and variance θ.

The Laplace transformation of the inverse Gaussian distribution is:-

$$LP(s) = exp\left[\frac{1-(1+2\theta s)^{1/2}}{\theta}\right], \theta > 0, s > 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (3.28)$$

For the inverse Gaussian frailty distribution the conditional survival function is given by: (Gutierrez, 2002).

$$S_\theta(t) = exp\left\{\frac{1}{\theta}\left(1-[1-2\theta ln\{S(t)\}]^{1/2}\right)\right\}, \theta > 0 \; \ldots\ldots\ldots\ldots\ldots.\ldots\ldots\ldots\ldots \; (3.29)$$

For the inverse Gaussian frailty distribution the conditional hazard function is given by: (Gutierrez, 2002).

$$h_0(t) = h(t)[1-2\theta ln\{S(t)\}]^{-1/2}, \theta > 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..\ldots\ldots\ldots\ldots\ldots\ldots\ldots. (3.30)$$

where S (t) and h(t) are the survival and the hazard functions of the baseline distributions.

With multivariate data, an Inverse Gaussian distributed frailty yields a Kendall's Tau given by:-

$$\tau = \frac{1}{2} - \frac{1}{\theta} + 2\frac{exp\left(\frac{2}{\theta}\right)}{\theta^2}\int_{\frac{2}{\theta}}^{\infty}\frac{exp(-u)}{u}du\;, where\; \tau\epsilon\left(\frac{0,1}{2}\right)$$

3.8 Baseline Hazard Distribution for Parametric Frailty Models

(i). Baseline Exponential Distribution

The exponential distribution, with only one unknown parameter and it is the simplest of all life distribution models. In the exponential model, the conditional probability is constant over time. In other words, the main feature of exponential distribution is that the instantaneous hazard does not vary over time. Modeling the dependency of the hazard rate on covariates entails constructing a model that ensures a non-negative hazard rate (or non-negative expected duration time). The exponential PH model is a special case of the Weibull model when $\gamma = 1$. The hazard function under this model is to assume that it is constant over time.

Table 2: Baseline Exponential distribution for survival and hazard functions

f(t)	S(t)	h(t)	H(t)	Parameter space
$\lambda exp(-\lambda t)$	$exp(-\lambda t)$	Λ	λt	$\lambda > 0$

(ii). Baseline Weibull Distribution

Weibull distribution is one of the parametric distributions which are used for the analysis of life time data and mostly used in literature for modeling life time data (Ibrahim et al., 2001 and Yu, 2006). The Weibull distribution is more general and flexible than the exponential distribution and allows for hazard rates that are non-constant but monotonic. It is a two parameter model (λ *and* γ), where λ is the scale parameter and γ is the shape parameter because it determines whether the hazard is increasing, decreasing, or constant over time i.e., the hazard rate increases when, $\gamma > 1$ and decreases when $\gamma < 1$ as time goes on. When $\gamma = 1$, the hazard rate remains constant, which is the special case of exponential.

Table 3 . Baseline Weibull distribution for Survival and Hazard functions.

f(t)	S(t)	h(t)	H(t)	Parameter space
$\gamma\lambda t^{\gamma-1} exp(-\lambda t^{\gamma})$	$exp(-\lambda t^{\gamma})$	$\gamma\lambda t^{\gamma-1}$	λt^{γ}	$\lambda, \gamma > 0$

(iii). Baseline Log-Logistic Distribution

The cumulative distribution function can be written in closed form is particularly useful for analysis of survival data with censoring (Bennett, 1983). The log-logistic distribution is very similar in shape to the log-normal distribution, but is more suitable for use in the analysis of survival data. The log-logistic model has two parameters λ is the scale parameter and γ is the shape parameter which is denoted by log L (γ , λ). The distribution imposes the following functional forms on the density, survival, hazard and cumulative hazard function:

Table 4. Baseline Log-logistic distribution for Survival and Hazard functions.

f(t)	S(t)	h(t)	H(t)	Parameter space
$\dfrac{\lambda\gamma t^{\gamma-1}}{(1+\lambda t^{\gamma})^2}$	$\dfrac{1}{1+\lambda t^{\gamma}}$	$\dfrac{\lambda\gamma t^{\gamma-1}}{1+\lambda t^{\gamma}}$	$ln\left[1+\left(\dfrac{t}{\lambda}\right)^{\gamma}\right]$	$\lambda \in$ R, $\gamma > 0$

By specifying one of the four functions f (t), S(t), h(t) or H(t) specifies the other three functions of the above baselines. The parameter λ is reparameterized in terms of predictor variables and the regression parameters. Typically for parametric models, the shape parameter γ is held fixed.

3.9 Comparisons of Models

Model comparison and selection are among the most common problems of statistical practice, with numerous procedures for choosing among a set of models (Kadane and Lazar, 2001) and (Rao and Wu, 2001). There are several methods of model selection. The most commonly used methods include information criteria. One of the most commonly used model selection criteria is Akaike Information Criterion (AIC). A data-driven model selection method such as an adapted version of Akaike's information criterion AIC (Akaike, 1974) is used to find the truncation point of the series. In some circumstances, it might be useful to easily obtain AIC value for a series of candidate models (Munda et al., 2012). In this study, we used the AIC criteria to compare two different multilevel survival models. The model with the smallest AIC value is considered a better fit.

3.10 Model Diagnostics

3.10.1 Evaluation of the Baseline Parameters

The graphical methods can be used to check if a parametric distribution fits the observed data or not. The model with the Weibull baseline has a property that the log (-log ($\hat{s}(t)$)) is linear with the log of time, where $\hat{s}(t) = exp(-\lambda t^{\gamma})$. Hence, log (-log ($\hat{s}(t)$))=log($\lambda$)+ γlog(t)).

The intercept and slope of the line will be rough estimate of log λ and γ respectively. This property allows a graphical evaluation of the appropriateness of a Weibull model by plotting log (–log ($\hat{s}(t)$))) versus log (t) where $\hat{s}(t)$ is Kaplan-Meier survival estimate(Datwyler and Timon Stucki, 2011).

The appropriateness of the model with the exponential baseline can graphically be evaluated by plotting –log ($\hat{s}(t)$) versus t where $\hat{s}(t)$ is Kaplan-Meier survival estimate. This plot should be linear and goes through the origin (Klein, 1992). Because for exponential distribution, $\hat{S}$(t) = $exp\ (-\lambda t)$, and hence, -log ($\hat{s}(t)$) = λt is linear with time.

The appropriateness of the model with the log logistic baseline can graphically be evaluated by plotting log (1 − $\hat{s}(t)$/ $\hat{s}(t)$) versus log (t), where $\hat{s}(t)$ is Kaplan-Meier survival estimate. The log-failure odd versus log time of the log-logistic model is linear with slope γ then the survival time follows a log-logistic distribution.

Where the failure odds of log-logistic survival model can be computed as:

$$(1-\mathrm{S(t)})/\mathrm{S(t)} = \frac{\frac{\lambda t^{\gamma}}{1+\lambda t^{\gamma}}}{\frac{1}{1+\lambda t^{\gamma}}}$$

Therefore, the log-failure odds can be written as log((1-S(t))/S(t))=log(λt^{γ}) = log(λ) + γlog(t) which is the liner function of log(t). (Datwyler and Timon Stucki, 2011)

3.10.2 The Cox-Snell Residuals

The Cox-Snell residuals method can be applied to any parametric model and the residual plots can be used to check the goodness of fit of the model. For the parametric regression problem, analogs of the semi-parametric residual plots can be made with a redefinition of the various residuals to incorporate the parametric form of the baseline hazard rates (Klein and Moeschberger, 2003). The Cox-Snell residual for the j^{th} individual with observed survival time t_j is given by r_j = $\hat{H}$ (T_j /X_j)=−log$\hat{S}(T_j/X_j)$, where $\hat{H}$ and $\hat{s}$ are the estimated values of the cumulative hazard and survivor function of the j^{th} subject at time t_j respectively. If the model fits the data, then the r_j's should have a standard (λ =1) exponential distribution, so that a hazard plot of r_j versus the Nelson–Aalen estimator of the cumulative hazard of the r_j's should be a straight line with slope unity and zero intercept. If yes, the fitted model is adequate. In general, Cox-Snell residual that provides a check of the overall fits of the model (Cox and Snell, 1968).

The three baseline hazard functions of Cox–Snell residuals that are considered in this study are given below:

Table 5: The three baseline hazard functions of Cox-Snell residuals

Model	r_j
Exponential	$\hat{\lambda}\, t_j exp\left(\widehat{\beta'}\, Xj\right)$
Weibull	$\hat{\lambda}\, t_j \hat{\gamma}\, exp\left(\widehat{\beta'}\, Xj\right)$
Log-logistic	$\left[\frac{1}{1+\hat{\lambda}\, tj\hat{\gamma}\, exp\left(\widehat{\beta'}\, Xj\right)}\right]$

CHAPTER FOUR

RESULTS AND DISCUSSIONS

4.1 Descriptive Statistics

Of all 12066 women aged 15-49, 9466(78.45%) were married and the median & mean age at first marriage for women living in Ethiopia were 16 years and 16.2 years respectively, while the minimum and maximum age at first marriage observed were 8 years and 49 years respectively. The Percentage of married women aged 15-49 was highest for those living in Amhara region (11.14%), followed by those living in Oromiya region (9.45%), while the lowest for those living in Somali region(4.81%) when compared to those living in other regions. The mean age at first marriage is highest in Addis Ababa (19years) and lowest in Amhara (14years). Similarly, the median age at first marriage is highest in Addis Ababa (18years) and lowest in Amhara (13years). The percentage of women aged 15-49 who were ever married was higher for those residing in the rural area(57.97%) than those residing in urban area(20.48%). The median age and mean age at first marriage among women aged 15-49 is 17 years and 18 years respectively for those living in urban area, which is higher than those living in rural area (median=15 years, mean=16 years). Concerning to educational level of women, percentage of women aged 15-49 who were married was highest for uneducated women (50.39%), while the lowest for those women having higher education (2.54%) relative to wome having other level of education. However, median age and mean age at first marriage is similar as well as highest (21 years) for those women achieving higher education, while similar as well as lowest for those with no education women (15 years). The result revealed that the percentage of women aged 15-49 who were married was highest for those women with Muslim religion (31.87%), while the lowest for Catholic religion followers (0.84%) when compared to other religions in Ethiopia. The mean age at first marriage is similar as well as highest for those women except orthodox and Muslim followers (17years), whereas the lowest as well similar for those women with orthodox and Muslim religion (16years). The median age at first marriage is similar as well as highest for those women except orthodox followers, while the lowest for those women with orthodox religion. Regarding to wealth index, the percentage of ever married women aged 15-49 were highest for richest women (21.95%), followed by poorest women (20.45%), while the lowest for those having middle income (11.78%). The percentage of women aged 15-49 who were married were higher for those who have not access to media(48.22%) than those having access to media(30.23%). Both median age and mean age at first marriage for women age 15-49 is higher among those having better access to media and wealthier. From the result we can observe that, relatively the percentage of women aged 15-49 who were married were highest for those women having uneducated head (39.87%) and agriculturalist head(50.91%), while lowest for those having head who achieved higher education (4.91%) and professional head(4.46%). The median age and mean age at first marriage among women aged 15-49 is18 years and 19 years respectively for those women having head who achieved higher education, which is highest, while lowest for those having no educated head relative to those women having other head education level. The median age at first marriage is similar as well as highest for those women having professional and laborers head (17years), while lowest for those having agriculturalist head (15years). The mean age at first marriage is highest for those women having professional head (18years), whereas lowest for those having agriculturalist head (16years). Finally, with regard to work status of women, the percentage of women aged 15-49 who were first married were higher for those who have not work than those having work. The median age and mean age at first marriage is similar, which is 16 years whether the women are working or not.

Table 6:Descriptive summary for women age at first marriage by categories of covariates.

Covariates	Category	No.(Percentage) of Women	No.(Percentage) of Women ever married	Mean (Year)	Median (year)
Region	Tigray	1281(10.62)	959(7.95)	16	15
	Afar	1084(8.98)	885(7.33)	16	16
	Amhara	1630(13.51)	1344(11.14)	14	13
	Oromiya	1576(13.06)	1140(9.45)	16	16
	Somali	741(6.14)	580(4.81)	17	16
	Benshangul-Gumeze	1025(8.49)	867(7.19)	15	15
	SNNP	1436(11.9)	1042(8.64)	17	17
	Gambela	927(7.68)	742(6.15)	16	15
	Harari	768(6.36)	626(5.19)	17	16
	Addis Ababa	846(7.01)	654(5.42)	19	18
	Dire Dewa	752(6.23)	627(5.20)	18	17
Residence	Urban	3145(26.06)	2471(20.48)	18	17
	Rural	8921(73.94)	6995(57.97)	16	15
Education Level	No Education	7647(63.38)	6080(50.39)	15	15
	Primary	3331(27.61)	2580(21.38)	16	16
	Secondary	657(5.45)	499(4.14)	19	19
	Higher	431(3.57)	307(2.54)	21	21
Religion	Orthodox	4765(39.49)	3714(30.78)	16	15
	Catholic	132(1.09)	101(0.84)	17	16
	protestant	2129(17.64)	1642(13.61)	17	16
	Muslim	4838(40.10)	3846(31.87)	16	16
	Others	202(1.67)	163(1.35)	17	16
Wealth Index	Poorest	3116(25.82)	2468(20.45)	16	15
	Poorer	1935(16.04)	1491(12.36)	16	15
	Middle	1798(14.90)	1421(11.78)	16	15
	Richer	1848(15.32)	1438(11.92)	16	15
	Richest	3369(27.92)	2648(21.95)	17	17
Head Education	No Education	6040(50.06)	4811(39.87)	16	15
	Primary	4080(33.81)	3172(26.29)	16	16
	Secondary	1152(9.55)	891(7.38)	17	17
	Higher	794(6.58)	592(4.91)	19	18
Work Status	No	7734(64.10)	6097(50.53)	16	16
	Yes	4332(35.90)	3369(27.92)	16	16

Access to media	No	7300(60.50)	5818(48.22)	16	15
	Yes	4766(39.50)	3648(30.23)	17	16
Head Occupation	Agriculturalist	7822(64.83)	6143(50.91)	16	15
	Professional	706(5.85)	538(4.46)	18	17
	Laborers	1193(9.89)	931(7.72)	17	17
	Business	1384(11.47)	1099(9.11)	17	16
	Others	961(7.96)	755(6.26)	17	16

4.2 The Kaplan-Meier (KM) Survival Curve for Different Groups

The Kaplan-Meier estimate of the survival curve is the best description of times to event of a group of subjects using all the data currently available. The resulting KM survival curve based on EDHS 2011 dataset is shown in the following figure. Note that in this plot survival time is being measured in years.

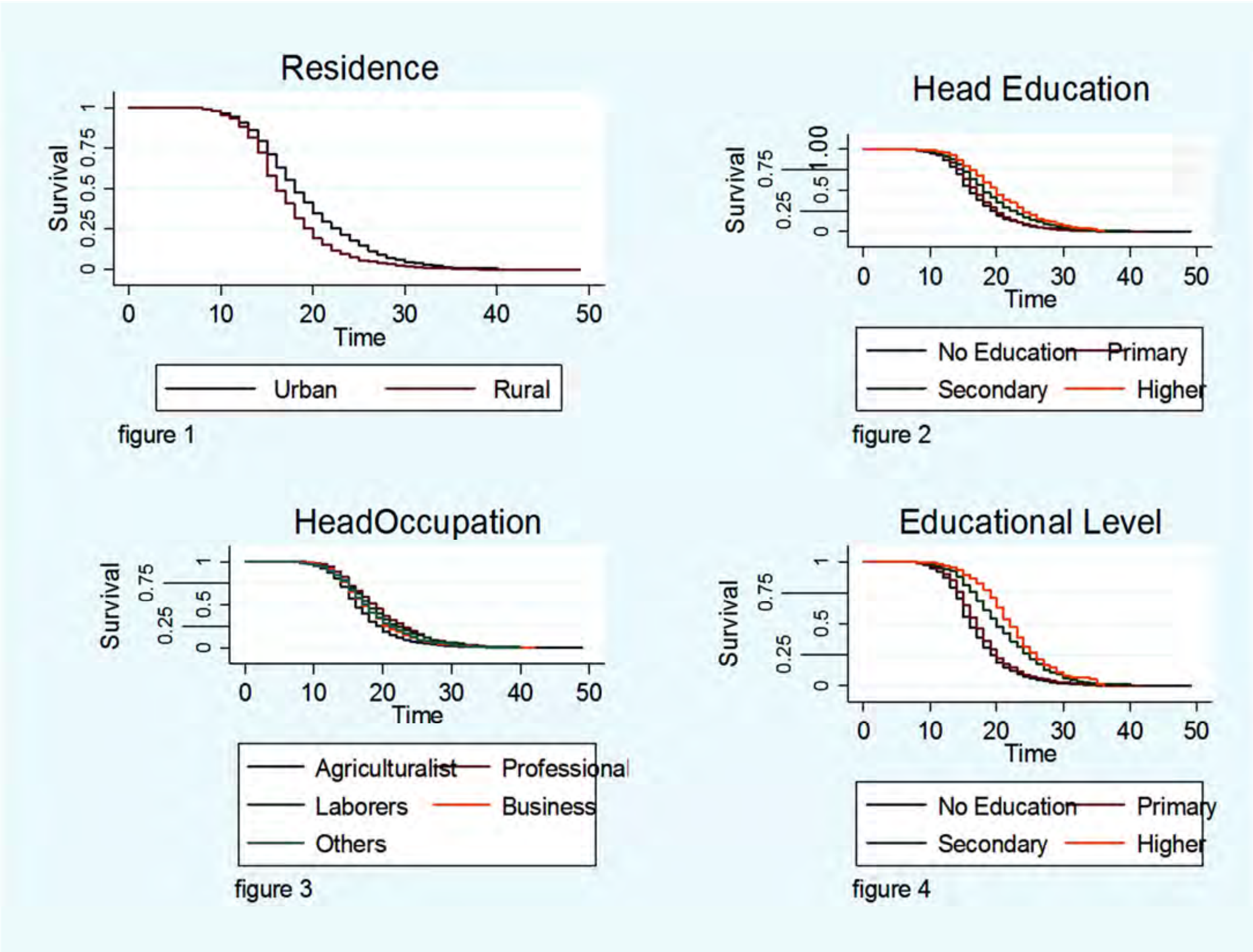

Figure 1 indicates that survival probability by women's residence VS time (in years). This curves starts at one and continues horizontally until age at first marriage happened at 8 years; at this time, it then drops down for both women residing in urban and rural with slightly higher age at first marriage for those living in urban relative to those living rural. The longest survival time (age at first marriage) is for women who in fact married at 49 years.

Figure 2 indicates that survival probability by Head Education of women VS time (in years). In the plot we can see that survival time to marry of women who have no educated head is worst among women having head who achieved higher education.

Figure 3 indicates that survival probability by Head Occupation VS time (in years). This shows that women whose head occupation is professional are less likely to marry early among those women having head occupation except professional. In contrast, those women having Head that is Agriculturalist are more likely to marry early.

Figure 4 indicates that survival probability by Educational level of women VS time (in years). The differences that are displayed in survival curve shows that; women who attend at least primary school have higher age at first marriage when compared to those women who have no education.

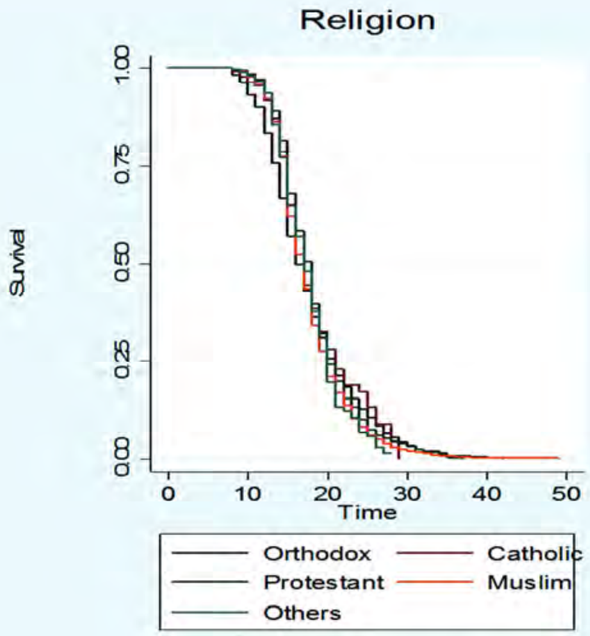

figure 5

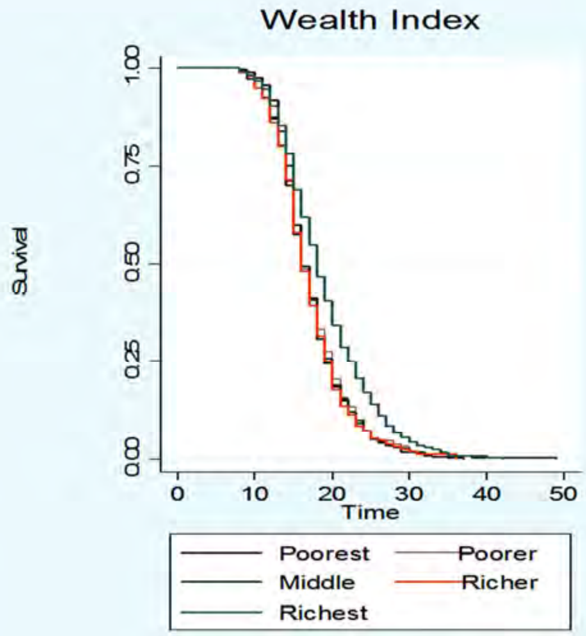

figure 6

Figure 5 indicates that survival probability by religion of women VS time (in years). The figure show that, those women with Catholic religion seems to have highest age at first marriage followed orthodox religion followers and the lowest for those with other religion (religion except orthodox, catholic, protestant & Muslim).

Figure 6 indicates that survival probability by wealth index VS time (in years). From the plot we see that, wealthier women have lowest age at first marriage than others. Here the curve for poorer, middle, richer and richest seems to be overlapped. But it seems as there is least gap among curves when we made a delicate look at.

4.3 Statistical Model for Multilevel Survival Analysis

The two different statistical models that we consider in this study were especial case of mixed effect cox regression model (Parametric shared frailty model) and discrete time survival model with mixed effects (Complementary log-log binomial linear model with mixed effects). First let us try to see multivariable survival analysis using parametric shared frailty model with model comparison among various parametric shared frailty models and then finally we intend to compare between one best selected model among various parametric shared frailty models and discrete time survival models with mixed effects (Complementary log-log binomial linear model with mixed effects) for EDHS 2011 dataset. As a result the elaboration or justification of the study for the latter case should be based on the final selected model.

4.3.1 Multivariable Analysis and Model Comparisons for Parametric shared frailty model

The multivariable survival analysis in this part of the study was done by assuming the exponential, weibull and log-logistic distributions for the baseline hazard function; and the gamma, inverse Gaussian and Positive stable frailty distributions. It was performed using the covariates; residence, education level of women, religion, work status of women, access to media of women, wealth index of household, parent education and parent occupation. In this study, we used the AIC criteria to compare various candidates of parametric shared frailty models. The model with the smallest AIC value is considered a better fit.

Table 7:AIC value and test of unobserved heterogeneity for multivariable parametric shared frailty models, EDHS 2011.

Shared frailty Model	AIC	θ	τ	P-value
Exponential-Gamma	24443.57	0.0073	0.0036	0.000
Exponential-Inverse-Gaussian	24491.36	0.0062	0.0031	0.003
Exponential-Positive Stable	24562.66	0.0059	0.0029	0.003
Weibull-Gamma	6242.53	0.0535	0.0261	0.000
Weibull-Inverse-Gaussian	6638.82	0.0485	0.0237	0.000
Weibull-Positive Stable	7023.42	0.0293	0.0144	0.000
Log logistic-Gamma	4276.55	0.0877	0.0420	0.000
Log logistic-Inverse-Gaussian	4727.21	0.0754	0.0363	0.003
Log logistic-Positive Stable	5273.15	0.0516	0.0252	0.003

P-value= P-value for Likelihood-ratio test of theta=0, θ =theta (variance of random terms), τ=Kendaell's tau

Using all the multivariable parametric shared frailty models, the covariate educational level of women was significant, indicating that it was the most important determinant factor for the time to age at first marriage based on EDHS 2011 dataset. Place of residence of women was also a significant factor in all models except Exeponential -Gamma shared frailty model. Some categories of religion, wealth index and head education were significant in the four models namely, Weibull-Gamma, Log logistic-Gamma, Weibull-Positive Stable and Log logistic –Inverse Gaussian shared frailty models. Work Status of women, access to media of women and head occupation were not a significant factor for time to age at first marriage using all models based on EDHS 2011 dataset. The variance of the random effect (*θ)* was significant for all multivariable parametric shared frailty models. It was highest when we assume the Gamma frailty distribution (θ =0.0877) followed by the Inverse-Gaussian frailty distribution (θ=0.0754) with the Log logistic baseline hazard function. Similarly it was highest when we assume the Gamma frailty distribution (θ=0.0535) relative to other frailty distribution for the weibull baseline hazard function. This term was again highest when we assume a Gamma frailty distribution (θ=0.0073) relative to other frailty distribution for the Exponential baseline hazard function. The Kendall's tau (τ) was highest for the highest θ values. Accordingly the dependency within the clusters for the Log logistic-Gamma shared frailty model (τ=0.0877) was the maximum followed by Log logistic-Inverse-Gaussian shared frailty model (τ=0.0754). The AIC values of the Log logistic-Gamma shared frailty model (4276.55), was the minimum among all the other AIC values of the parametric shared frailty models indicating that it was relatively the most efficient model to describe dataset (Table 7).

Table 8 Multivariable analysis using the Log logistic-Gamma shared frailty model, EDHS 2011

Covariate	Coeff	St.err	P-value	ϕ	[95% CI for ϕ]	
Constant	2.757132	0.014421	0.000	15.75	15.32	16.21
Residence						
Urban(ref)						
Rural	-0.0331585	0.0097971	0.001	0.9673852	0.9489868	0.9861403
Education Level						
No Education(ref)						
Primary	0.0330889	0.005878	0.000	1.0336424	1.0218025	1.0456196
Secondary	0.1684082	0.0131558	0.000	1.1834196	1.1532952	1.2143309
Higher	0.281212	0.0163527	0.000	1.3247344	1.2829488	1.3678810
Religion						
Orthodox(ref)						
Catholic	0.046315	0.0227856	0.042	1.0474043	1.0016574	1.0952405
Protestant	0.0524916	0.0081184	0.000	1.0538937	1.0372571	1.0707973
Muslim	0.0524736	0.0064776	0.000	1.0538747	1.0405795	1.0673400
Others	0.0647476	0.0185256	0.000	1.0668897	1.0288463	1.1063398

Wealth Index						
Poorest(ref)						
Poorer	-0.0099705	0.0072415	0.169	0.9900790	0.9761260	1.0042314
Middle	-0.014243	0.0074423	0.056	0.9858580	0.9715818	1.0003438
Richer	-0.0299688	0.0075612	0.000	0.9704758	0.9561996	0.9849651
Richest	-0.0528073	0.0107207	0.000	0.9485628	0.9288392	0.9687050
Head Education						
No Education(ref)						
Primary	0.0154898	0.0055625	0.005	1.0156104	1.0045979	1.0267436
Secondary	0.02867	0.010067	0.004	1.0290849	1.0089790	1.0495915
Higher	0.04283	0.0140958	0.002	1.0437604	1.0153188	1.0729988
Work Status						
No(ref)						
Yes	-0.0046146	0.0049996	0.356	0.9953960	0.9856898	1.0051979
Access to media						
No(ref)						
Yes	-0.0003079	0.0057177	0.957	0.9996921	0.9885516	1.0109582
Head Occupation						
Agriculturalist(ref)						
Professional	-0.0078202	0.0126729	0.537	0.9922103	0.9678687	1.0171638
Laborers	0.0195623	0.0099795	0.050	1.0197549	1.0000028	1.0398970
Business	0.0117877	0.0085707	0.169	1.0118574	0.9950021	1.0289984
Others	-0.0065789	0.009861	0.505	0.9934427	0.9744262	1.0128299
θ =0.0876576, γ =0.1305663, τ = 0.0420 Likelihood-ratio test of theta=0: chibar2(01)=452.66 prob>=chibar2(01)=0.000						

Analysis based on Log logistic-Gamma shared frailty model showed that place of residence of women, religion of women, education level of women, some categories at wealth index of household and head education were significant at 5% level of significance. In contrast work status of women, head occupation and access to media were not significant at 5% level of significance.

Regarding to education level of women, women who attend primary school, secondary school and higher education had significantly different age at first marriage than the reference category (No Education) since their confidence interval of acceleration factor do not include 1 at 5% level of significance (Table 8). An acceleration factor of greater than 1 indicates prolonging the time to age at first marriage. Therefore women who attend primary school (ϕ=1.034), secondary school (ϕ=1.183) and higher education (ϕ=1.325) had prolonged age at first marriage by a factor of 1.034, 1.183 & 1.325 respectively than the reference category (No Education). The confidence interval of the acceleration factor for catholic, protestant, Muslim and other (category) religion were (1.0017, 1.0952), (1.0373, 1.0708), (1.0406, 1.0673) and (1.0288, 1.1063) respectively, did not include 1, indicating that those categories are also significant determinant factor for time to age at first marriage. Thus, women who are follower of catholic; protestant, Muslim and other religion (religion except orthodox, catholic, protestant & Muslim) had prolonged age at first marriage by a factor of 1.047, 1.05389, 1.05387 and 1.067 respectively when compared to the reference category (orthodox). Head education is also another significant covariate with acceleration factors greater than 1 for all categories. As a result those women having head that attend primary school, secondary school and higher education had prolonged age at first marriage by a factor of 1.016, 1.029 and 1.044 respectively when compared with the reference category (No Education). Categories of significant covariates having acceleration factor less than 1 imply that women characterized by those categories of the same covariate marry early relative to the reference category of the same covariates. For instance, Women residing in rural area of Ethiopia have marry early than those residing in urban area (ref) of Ethiopia (ϕ=0.967). Similarly, richer (ϕ=0.970) and richest women (ϕ=0.949) marry early than reference category (poorest women).

The value of the shape parameter in the Log logistic –Gamma shared frailty model was (γ =0.131). This indicates that non monotonic hazard rates, specifically initially increasing and then decreasing rates. The variability (heterogeneity) in the population of clusters (Region) estimated by this Log logistic –Gamma shared frailty model was θ =0.088, and the dependence within region was about τ = 4.2%.

4.3.2 Multivariable Survival Analysis for Discrete time survival model with mixed effects

In this part of the study we were done multivariable survival analysis using discrete time survival model with mixed effects. This model have the approach of divides follow up time into a finite set of mutually exclusive intervals and fits a survival model that assumes that the hazard function is not constant within each interval. This model also does not take into account the duration of exposure time or at risk within each interval and simply consider whether the subject experienced an event within the given interval. As a result a Complementary log-log generalized linear model with mixed effects can be fit.

This model was fitted by dividing the age at first marriage into five discrete intervals such as [5-15), [15-25), [25-35), [35-45) & [45-55). The aim of discretizing the intervals into mutually exclusive intervals were restructured the dataset that can able to compatible with statistical software and to do discrete time survival model with mixed effects. If each women experience the event of interest (marriage) in one of each mutually exclusive intervals the event indicator would be 1 and 0 otherwise. It has a binomial sense, and it is the reason why we were used Complementary log-log binomial linear model with mixed effects among various choice of discrete time survival model with mixed effects.

Table 9: Multivariable analysis using Complementary log-log model binomial linear model with mixed effects

Covariate	Coeff	St.err	P-valu	ϕ	[95% CI for Coeff]	
Constant	-8.869443	0.269541	0.000	1.406e-4	-9.397734	-8.341152
Residence						
Urban(ref)						
Rural	0.089287	0.233109	0.702	1.093395	-0.367598	0.546173
Education Level						
No Education(ref)						
Primary	-0.137331	0.138179	0.320	0.871681	-0.408156	0.133494
Secondary	-0.592554	0.337402	0.079	0.552914	-1.253849	0.068742
Higher	-0.895417	0.475878	0.060	0.408437	-1.82812	0.037286
Religion						
Orthodox(ref)						
Catholic	-0.196591	0.517820	0.704	0.821526	-1.2115	0.818318
Protestant	-0.798461	0.178497	0.000	0.450021	-1.148308	-0.448614
Muslim	-0.721670	0.120249	0.000	0.485926	-0.957383	-0.486017
Others	-0.631812	0.490733	0.198	0.531628	-1.593631	0.330008
Wealth Index						
Poorest(ref)						
Poorer	0.333864	0.170745	0.051	1.396354	-0.000789	0.668517
Middle	0.347257	0.17378	0.046	1.415181	0.006655	0.687860
Richer	0.377930	0.175840	0.032	1.459261	0.03329	0.722570
Richest	0.503477	0.256316	0.049	1.654464	0.001106	1.005848

Head Education						
No Education(ref)						
Primary	-0.338696	0.129901	0.009	0.712699	-0.593297	-0.084095
Secondary	-0.280998	0.236530	0.235	0.755030	-0.744588	0.182592
Higher	-0.588098	0.363146	0.105	0.555382	-1.299852	0.123655
Work Status						
No(ref)						
Yes	0.066312	0.111853	0.553	1.068559	-0.152916	0.285539
Access to media						
No(ref)						
Yes	-0.022746	0.125891	0.857	0.977511	-0.269487	0.223995
Head Occupation						
Agriculturalist(ref)						
Professional	-0.222438	0.343699	0.518	0.800565	-0.896076	0.451200
Laborers	-0.013009	0.221896	0.953	0.987075	-0.447916	0.421899
Business	-0.208772	0.218009	0.338	0.811580	-0.636061	0.218517
Others	0.263491	0.218027	0.227	1.301466	-0.163833	0.690815
lnsig2u	-17.6391	596.8205			-1187.386	1152.108
Sigma_u	0.0001478	0.044109			1.5e-258	1.5e+250
rho	1.33e-08	7.93e-06			0	
Likelihood-ratio test of rho=0: chibar2(01) = 0.00 prob>=chibar2(01)= 1.000						

Based on multivariable analysis using Complementary log-log binomial linear model with mixed effects, some categories of religion, wealth index and one category of head education level were significant factor in determining age at first marriage at 5% level of significance, whereas residence, education level of women, work status, access to media and head occupation were not significance. Similarly cluster (region) level variance was not significance. This test of significance were also confirmed by likelihood-ratio test of rho=0, where rho is the proportion of total variance contributed by the cluster (region) level variance component.

When rho is zero, the cluster (region) level variance components is not important, and the conditional model did not differ from the marginal model. Thus, the estimated variability among cluster (region) that is estimated by complementary log-log binomial linear model with mixed effects was the square of sigma_u which is approximately zero. The test also indicates that as there is no variation regards to age at first marriage among region of Ethiopia.

4.4 Multilevel Survival Model Comparison

In this part, we were consider comparisons based on significance of covariate and AIC value even if our concentration is mainly on AIC value as we were described in part of methodology for two different multilevel survival model.

Table 10:Comparison based on significance of covariate at 5% level of significance.

Log logistic-gamma shared frailty model	Complementary log-log binomial linear model with mixed effects
Significant covariates	**Significant covariates**
Residence, education level, religion, wealth index & head education level	Religion, wealth index & head education level
Not significant covariates	**Not significant covariates**
Work status, access to media & head occupation	Residence, education level, work status, access to media & head occupation

The covariate religion, wealth index and head education level were the most significance factor in determining age at first marriage in both models at 5% level of significance, while access to media, work status & head occupation were not significant at 5% level of significance in both models. The estimated variability was significant in log logistic-gamma shared frailty model, while it is not in Complementary log-log binomial linear model with mixed effects.

After we had Complementary log-log binomial linear model with mixed effects and one best selected model among various parametric shared frailty model (log logistic-gamma shared frailty model), we can proceed to compare those models based on AIC value. Accordingly, a model having small AIC value taken as a better fit to the dataset.

Table 11: Comparison based on AIC value.

model	Log log-gamma shared frailty model	Complementary log-log binomial linear model with mixed effects
AIC value	4276.55	6002.35

According to table 11, the AIC value for log logistic-gamma shared frailty model was lower. This implies that the dataset were best fitted by log logistic-gamma shared frailty model than Complementary log-log binomial linear model with mixed effects. Therefore, later on elaboration or justification of the study should base on this model.

One important reason why a large value of AIC value for Complementary log-log binomial linear model with mixed effects may be due to the fact that restructuring the dataset and dividing the follow up time (age at first marriage) into discrete intervals. Such discretization process can result in a loss of information and proper unable of explained for the hazard of age at first marriage by the effect of covariates.

4.5 Model Diagnostics

4.5.1 Evaluation of the Baseline Parameters

To check the adequacy of our baseline hazard, the exponential has been plotted against the cumulative hazard function with time-to- age at first marriage (in years). Similarly, the weibull has been plotted against the logarithm cumulative hazard function with the logarithm of time-to-age at first marriage (in years) and log-logistic has been plotted against the logarithm of the failure odds with the logarithm of time-to-age at first marriage (Figure 7). The plot of log-logistic was most linear than the other plots. The patterns suggested that the log-logistic hazard function was appropriate in the model.

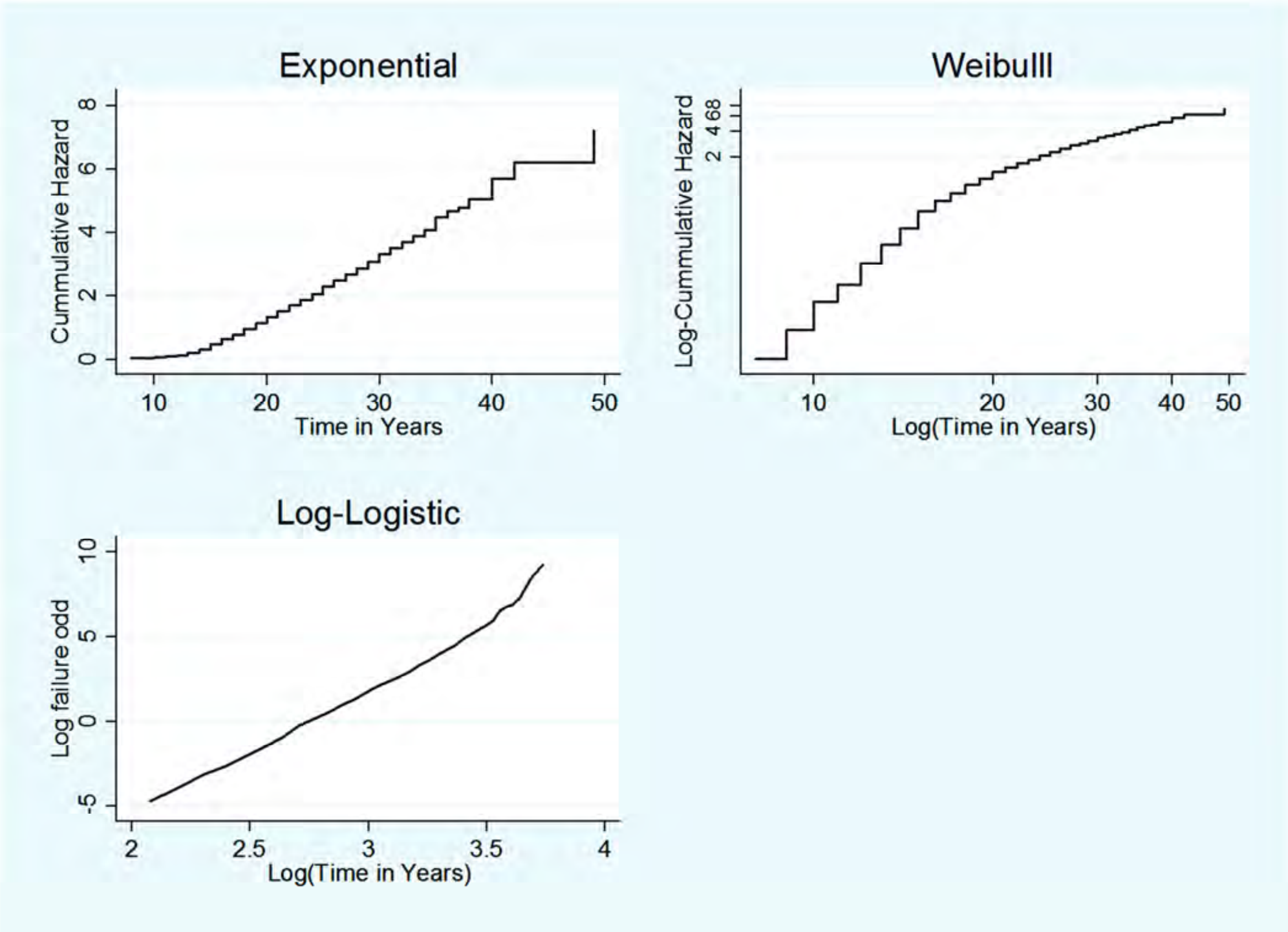

Figure 7 Graphical evaluations of the exponential, weibull and log-logistic assumptions.

4.5.2 The Cox-Snell Residuals

The Cox-Snell residuals together with their cumulative hazard function were obtained by fitting the exponential, weibull and log-logistic models to our dataset, via maximum likelihood estimation (Figure 8). The plots indicate that the Weibull and Log-logistic models seems to fit the data well even if it is better fitted by Log-logistic relative to others. But the exponential model fits poorly relative to others. These results are consistent with our previous results (in table 7) based on Akaike's information criterion. The plots of Cox-Snell residuals VS estimated cumulative hazard function were nearest to the line through the origin for all models.

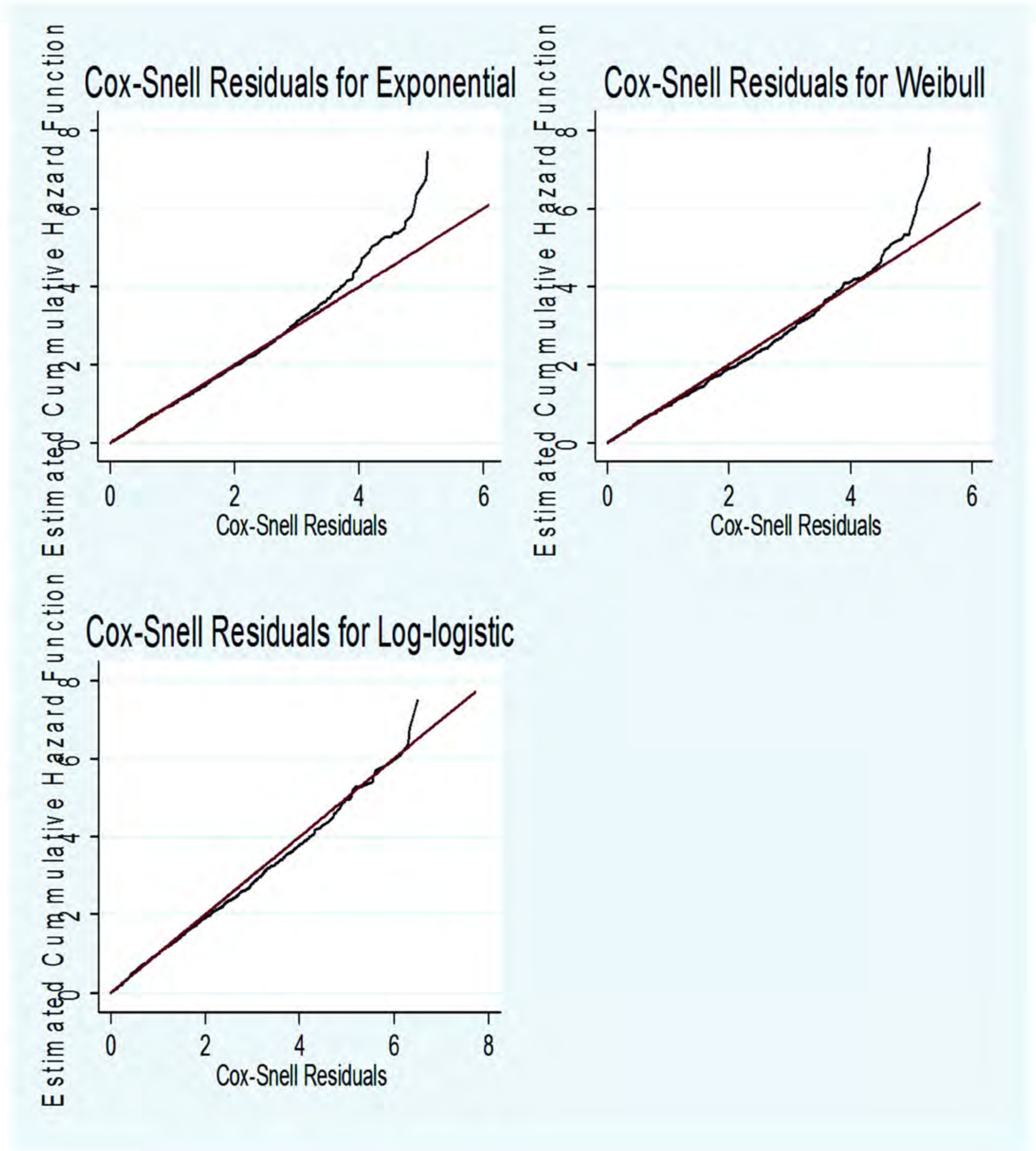

Figure 8 Cox–Snell residuals to evaluate model fit of exponential, weibull and Log-logistic model

4.6 Discussion

The results of this study suggested that place of residences was significant predictive factor for age at first marriage in Ethiopian women. This shows that women who lived in urban areas are more survived on age at first marriage than women who lived in rural areas. This might be due to the fact that rural areas tend to have institutional and normative structures such as the kinship and extended family that promote early marriage and childbearing, but women in urban areas need to develop skills, gain resources, and achieve maturity to manage an independent household and thus they have to delay marriage. A study in Nigeria by Thomas (2010-2011) and Adebowale (2012) found that women who are living in rural area had a higher risk of first marriage than urban area. The results of this study suggest that work status of the women had a significant effect on age at first marriage and age at early marriage was higher for women who had no work status than women having work. This is consistent with the finding of Shapiro (1996), Zahangir *et al.* (2008) and Kamal (2011), they revealed that work status of women have significant effect on age at first marriage and pre-marital work status of women significantly delayed the timing of marriage. But this is not consistent with our study with regards to work status of women. In our study work status of women was not a significant factor in determining time to age at first marriage among women living in Ethiopia. A study conducted in Ethiopian regions by Erulkar (2013) investigated the factors associated with marriage and the result suggested that educational attainments of women had significant effect on marriage and women who were not educated were married earlier than educated. This study is consistent with our finding since our study revealed that educational level of women had a significant effect on time to age at first marriage at 5% level of significance and it prolonged age at first marriage by the factor of ϕ= 1.034, 1.183 and 1.325 for primary, secondary and higher education respectively when illiterate women was used as the reference group.

With regards to wealth index of household, compared to poorest women wealthier had early age at first marriage. These findings seem to suggest that in Ethiopia, richest women may attract male marriage partners. Studies elsewhere suggest the opposite, with poor women having a relatively higher risk of first marriage (Haloi and Limbu, 2013; Sivaram, Richard, and Rao, 1995; Hoq, 2013). Head occupation was not a significant determining factor for time to age at first marriage based on our study. But a study by Kamal (2011) revealed that heads/parents occupations are the important factor for age at first marriage of Ethiopian women. A similar study in western Uganda by Peninah et al. (2011) and Zahangir et al. (2008) in rural Bangladesh found that the occupation of the parents were strong socio-economic determinants of age at first marriage. Also, another study in Bangladesh by Mosammat et al. (2013) showed that the occupations of the parents were important factors for determining age at first marriage.

Religion of women was found to have a significant effect on age at first marriage in our study. The result showed that women who follow religion except orthodox had prolonged time to age at first marriage than those who follow orthodox religion. This finding is consistent with (Okeibunor, 1999; Hoq, 2013). Access to mass media was found to have a significant effect on age at first marriage by [Tezera (2013), Zahangir and Kamal (2011)]. But in our study it was not. In our study Head Education was also found to be the significant factor in determining time to age at first marriage. This finding was consistent with results from (Agaba et al, 2011;Kamchulesi et al, 2011).

The final model of our study also includes some of insignificant covariates such as work status of women, access to media of women and head occupation because these covariates have been found to be most significantly affect the time to age at first marriage among women based on most literature [Tezera (2013), Zahangir and Kamal (2011), Kamal (2011), Peninah et al. (2011) and Zahangir et al. (2008), Mosammat et al. (2013), Thomas (2010-2011) and Adebowale (2012), Shapiro (1996), Zahangir *et al.* (2008) and Kamal (2011)].

CHAPTER FIVE

CONCLUSION AND RECOMMENDATION

5.1 Conclusion

The main objective of this study was to model time-to- age at first marriage among women living in Ethiopia using appropriate multilevel survival model by comparing two different multilevel survival models. The comparison of two different multilevel survival models for the dataset was performed using the AIC criteria, where a model with minimum AIC is accepted to be the best. Accordingly, the log-logistic-Gamma shared frailty model which has smallest AIC value was the most appropriate model to describe the dataset. This study also showed that there was a clustering (frailty) effect on modeling time-to- age at first marriage among women living in Ethiopia due to the fact that heterogeneity in Region from which the women live in, assuming women living in the same Region share similar risk factors related to marriage. Therefore, it was important considering the clustering effect in modeling the hazard function. The dataset was also best described by the log-logistic baseline as compared to the exponential and weibull hazard functions. According to the diagnostic plots the log failure odds of log-logistic baseline with log time was more linear as compared to the plots of exponential (cumulative hazard versus time) and weibull (log cumulative hazard versus log time), showing the dataset was best described by the log-logistic baseline. This result was also confirmed by the cumulative hazard plots for the Cox-Snell residuals of the exponential, weibull and the log-logistic models. The plot was more approached to the line in case of the log-logistic model, indicating that the log-logistic was relatively best.

The determinant factors considered were residence of women, educational level of women, religion of women, work status of women, access to mass media, wealth index of household, head educational level and head occupation. Analysis using the best model, Log logistic-Gamma shared frailty model showed that residence of women, educational level of women, religion of women, wealth index and head educational level were the most significance factors for the time-to-age at first marriage. Women residing in urban part of Ethiopia had prolonged age at first marriage as compared to those residing in rural part of Ethiopia. Concerning educational level of women, women having better education had prolonged age at first marriage than illiterate women. Similarly, women having Head who achieved higher education had prolonged age at first marriage compared to women having illiterate head.

With regards to religion of women, those women who follow religion except orthodox had prolonged age at first marriage compared to those who follow orthodox religion. In case of wealth index of household, richer and richest women marry early than reference category (poorest women). This study also revealed that, of all 12066 women age 15-49, 9466(78.45%) were married and the median & mean age at first marriage for women living in Ethiopia were 16 years and 16.2 years respectively, while the minimum and maximum age at first marriage observed were 8 years and 49 years respectively. The mean ages at first marriage among women in Ethiopia were not equal in most regions of Ethiopia. It is lowest for Amhara and Benshanguel-Gumeze regions, while highest for Addis Ababa administration city when compared to other regions in Ethiopia.

5.2 Recommendation

Awareness has to be given for the society on age at the marriage. The education sector can play an effective role in this regard and the awareness need to follow the ordinance of the legal age of marriage because it is the most determinants of health for women and child borne. Moreover, it is advisable to target young women, particularly those with no or little education including primary school girls, with information on reproductive health and to provide them to avoid ultimately early age marriage.

Religious leaders can also play an important role to delay age at first marriage among women in Ethiopia. Especially it is advisable to orthodox religion leaders to delay early marriage of women for their followers by giving basic information regards to marriage and by developing the perception of women.

Another recommendation that emerge from the study is that as it is crucial to improving women's age at first marriage in regions of Ethiopia especially at Amhara and Benshanguel-Gumeze regions.

Further studies should be conducted in each region of Ethiopia and identify other factors that are not identified in this study. Based on that study regional government should takes an action on age at first marriage.

Reference

1) ACPF (2012). Minimum Age of Marriage: The International/Regional Framework. http:// www.africanchildforum.org

2) Adebowel A., Fagbamigbe A., Okareh O. and Lawal O. (2012). Survival Analysis of Timing of First Marriage among Women of Reproductive age in Nigeria: *African Journal of Reproductive Health.*

3) Agaba, P., Atuhaire, L.K., Rutaremwa, G. (2011). Determinants of age at first marriage among women in Western Uganda, Proceedings of the 6th African Population conference, Ouagadougou, Burkina Faso, 5-9 December 2011.

4) Akaike H. (1974). A new look at the statistical model identification. IEEE Trans Automatic Control.

5) Alemu, B. (2008). Early Marriage in Ethiopia: Causes and Health Consequences. *Exchange on HIV/AIDS, Sexuality and Gender*, (1):4-6.

6) Allison, P. D. (1995). Survival analysis using the SAS system. Cary, NC: SAS Institute.

7) Andersen, P. K., & Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Annals of Statistics*, 10, 1100–1120.

8) Andersson, Gunnar [1997]: The Impact of Children on Divorce Risks of Swedish Women. *European Journal of Population* 13: 109-145.

9) Austin,P.C (2017). A Tutorial on Multilevel Survival Analysis: Methods, Models and Applications. *International Statistical Review*. Canada

10) Bayisenge J. (2010). Early Marriage as a Barrier to Girl's Education Rwanda: *National University of Rwanda.*

11) Beegle, K., De Weerdt, J., and Dercon, S. (2006). Orphanhood and the longrun impact on children. *American Journal of Agricultural Economics*, 88(5):1266-1272.

12) Beegle, K. and Krutikova, S. (2008). Adult mortality and children's transition into marriage. *Demographic Research*, 19(42):1551-1574.

13) Bennett S. (1983). Analysis of survival data by the proportional odds model. *Statistics in Medicine*, 2, 273.

14) Bongaarts, J. (1983). *A framework for analyzing the proximate determinants of fertility Population and Development Review*, 4, (1): 105-132.

15) CNN (2013). Girl Rising - Stories of girls fighting to go to school. *Film* aired on (6.17.2013).

16) CSA [Centeral Statisical Authority, Ethiopia and ORC Macro]. Ethiopia Demographic and Health Survey 2000, Addis Ababa, Ethiopia and Calverton. Central Statistical Authority and ORC Macro; Maryland: 2001.

17) CSA [Centeral Statistical Authority, Ethiopia and ORC Macro]. Ethiopia Demographic and Health Survey 2005, Addis Ababa, Ethiopia and Calverton. Central Statistical Authority and ORC Macro; Maryland: 2006.

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

19) Coale, A. (1971). Age patterns of marriage. *Journal of Population Studies*, 25, (2): 193-214.

20) Collett D., Modelling Survival Data in Medical Research(1994), Chapman & Hall/CRC, London, UK, 2nd edition.

21) Cosic, D. and Deb, P. (2010). The Effects of Mortality Shocks on Household Spending on Education, Health and Nutrition. In Fuentes-Nieva, R. and Seck, P., editors, *Risk, Shocks and Human Development On The Brink*, 207-220. Palgrave Macmillan.

22) Cox D. R. and Snell E. J. (1968). A general definition of residuals with discussion. *Journal of the Royal Statistical Society. Series B 30 (1968), 248-275.*

23) Crowther, M.J., Look, M.P. & Riley, R.D. (2014). Multilevel mixed effects parametric survival models using adaptive Gauss-Hermite quadrature with application to recurrent events and individual participant data meta-analysis. Stat.Med., 33, 3844–3858.

24) Dagne HG. Early Marriage in Northern Ethiopia. Reproductive Health Matters. 1994; 4:35–38.

25) Davis, K, & Blake, J. (1956). Social structure and fertility: an analytic framework. *Economic Development and Cultural Change*, 4, (3): 211-235.

26) Deb, P. and Rosati, F. (2002). Determinants of Child Labor and School Attendance: The Role of Household Unobservables. *Technical Report*, Hunter College: Department of Economics.

27) Dercon, S. and Hoddinott, J. (2011). The Ethiopian Rural Household Survey 1989-2009: Introduction. *mimeo*, International Food Policy Research Institute (IFPRI).

28) Duncan, ME.; Mehari, L.; Tibaux, G.; Pelzer, A. Social Aspects of Obstetrics and Gynecology. In: Kloos, H.; Zein, ZA., editors. The Ecology of Health and Disease in Ethiopia. Westview Press; Boulder: 1993. p. 307-318.

29) Erulkar, A. S., & Muthengi, E. (2009). Evaluation of Berhane Hewan: a program to delay child marriage in rural Ethiopia. International Perspectives on Sexual and Reproductive Health, 6-14.

30) Erulkar A. (2013). Early Marriage, Marital Relations and Intimate partner violence in Ethiopia.

31) Ezeh, A.C., & Dodoo, F.N. (2001). Institutional change and African fertility transition

32) Fafchamps M, Quisumbing A. Control and Ownership of Assets within Rural Ethiopian Households. *The Journal of Development Studies.* 2002; 38(6):47–82.

33) Fafchamps M, Quisumbing A. Assets at marriage in rural Ethiopia. *Journal of Development Economics.* 2005a; 77:1–25.

34) Fafchamps, M. and Quisumbing, A. R. (2005). Marriage, bequest, and assortative matching in rural Ethiopia. *Economic Development and Cultural Change*, 53(2):347{380.

35) Fine, J.P., Glidden, D.V., Lee, K.E. (2003) A simple estimator for a shared frailty regression model. *Journal of the Royal Statistical Society* (B) 65, 317 – 329

36) Goldstein, H. (2010a) A general model for the analysis of multilevel discrete time survival data. Submitted for publication.

37) Green, C. A., Mukuria, A. and Rubin D. (2009). Addressing early marriage in Uganda, Washington D.C. Futures Group, Health Policy Initiative, Task Order 1.

38) Gutierrez, R. G., S. L. Carter, and D. M. Drukker. 2002. sg160: On boundary-value likelihood ratio tests. Stata Technical Bulletin 60: 15–18. Reprinted in Stata Technical Bulletin Reprints, vol. 10, pp. 269–273. College Station, TX: Stata Press.

39) Haloi, A. and Limbu, D.K. (2013). Socio-economic factors influencing age at first marriage of Muslim women of a remote population from North East India, *Antrocom Online Journal of Anthropology* 9 (1) 75-79.

40) Hoq, M. (2013).Regional differentials in age at first marriage among women in Bangladesh, Asian Journal of Applied Science and Engineering 2 (2)76-83.

41) Hougaard P. (2000). *Analysis of Multivariate Survival Data, Springer-Verlag,* Newyork.

42) Ibrahim J.G., Chen M. and Sinha D. (2001). Bayesian survival analysis. Springer Verlag, New York.

43) Ikamari,L.D.(2005). The effect of education on the time of marriage in Kenya. Demographic Research 12(1):1-28.doi;10.4054/DemRes.2005.12.1.

44) IPPF and UNFPA. (2006). Ending Child Marriage: A Guide for Global Policy Action, *International Planned Parenthood Federation (IPPF)*, London

45) IWHP (2009). Early Marriage: The society of obstetricians and Gynecologists of Canada.

46) Jensen, P. and Nielsen, H. S. (1997). Child Labour or School Attendance? Evidence from Zambia. *Journal of Population Economics*, 10(4):407-424.

47) Kadane J. and Lazar N. (2001). Methods and Criteria for Model Selection. *Technical Report*, 759, Carnegie Mellon University.

48) Kamal S.M.Mostafa (2011). Socio-Economic Determinants of Age at First Marriage of the Ethnic Tribal Women in Bangladesh, Asian Population studies.

49) Kamchulesi, G., Palamuleni, M. and Kalule-Sabiti, I. (2011). Factors affecting age at first marriage in Malawi, Proceedings of the 6th African Population conference, Ouagadougou, Burkina Faso, 5-9 December 2011.

50) Klein J. (1992). Survival analysis: techniques for censored and truncated data. *Medical College of Wisconsin.*

51) Klein P. and Moeschberger L. (2003). Survival Analysis Techniques for Censored and Truncated Data – Google Books.htm

52) Lindstrom, D.P., Kiros, G.E., and Hogan, D.P. (2009). Transition into first intercourse, marriage, and child-bearing among Ethiopian women. *Genus*, 65(3).

53) Martin, Teresa Castro - Larry L. Bumpass [1989]: Recent Trends and Differentials in Marital Disruption. Demography, 26. 1: 37-51.

54) Mathur, S.,Greene, M.,&Malhotra, A (2003). Too Young to Wed: The lives, rights and health of young married girls. Washington, DC: International center for Research on Women.

55) Mensch, B. S., Singh, S. and Casterline, J.B. (2005). Trends in the timing of first marriage among men and women in the developing world, Population Council, Policy Research Division, Working Paper 202.

56) Ministry of Education and Sports (2003). The proposed strategic plans for higher education 2003-2015.

57) Mosammat Zamilun Nahar, Mohammad Salim Zahangir and S.M. Shafiqul Islam (2013). Age at first marriage and its relation to fertility in Bangladesh, *Chinese Journal of Population Resources and Environment, 11:3, 227-235,DOI:10.1080/10042857.2013.835539*

58) Munda M., Rotolo F. and Legrand C. (2012). Parametric Frailty Models in R. *Journal of American Statistical Association: 55:1-21.*

59) Namibia Demographic and Health Survey Report 1992

60) Namibia Demographic and Health Survey Report 2006/7

61) Namibia Demographic and Health Survey Report, 2000

62) National Laws and Policies on minimum ages in Namibia (1993).

63) NCTPE (2003). Old Beyond Imagings. Ethiopian Harmful Traditional Practices. Addis Ababa, Ethiopia, United printers.

64) Niger Demographic and Health Survey Report 1995 to 2003

65) Nour N. M. (2006). Health Consequences of Child Marriage in Africa. *Emergence Infection Disease* 12(11):1644-9.

66) NRC/IOM (2005). Growing up Global; The changing Transitions to adulthood in Developing Countries. Panel on Transitions to Adulthood in Developing Countries, Cynthia B. Lloyd Washington, DC: National Academies Press.

67) Oakes, D. (1989) Bivariate survival models induced by frailties. *Journal of the American Statistical Association* 84, 487 – 493

68) Okeibunor, J.C. (1999). Religious differentials in trends in age at first marriage in Enugu State, Nigeria, *Journal of Humanities*, 10, 209-226.

69) Pankhurst, H. Gender, Development and Identity: An Ethiopian Study. Zed Books; London:1992.

70) Pathfinder International Report (2006). *Report on the causes and consequences of early marriage in Amhara region,* Addis Ababa, Ethiopia.

71) Peninah A., Leonard K. Atuhaire. and Gideon Rutaremwa (2011). Determinants of age at first marriage among women in western Uganda. Presented in European Population Conference 2010, Vienna 1-4 September 2010

72) Population Council (2004). Supporting Married Girls: Calling Attention to a Neglected Group.

73) Prentice, R. L., Williams, B. J., & Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, 68, 373–379.

74) Rabe-Hesketh, S. & Skrondal, A. (2012b). Multilevel and Longitudinal Modeling Using Stata, Volume 2: Categorical Responses, Counts, and Survival, 3rd ed. Vol. 2: Stata Press.

75) Rao C. and Wu Y. (2001). On Model Selection (with discussion). In P. Lahiri (Ed.), Model Selection, Volume 38 of IMS Lecture Notes - Monograph Series. Institute of Mathematical Statistics.

76) Rodriguez, G. and Goldman, N. (2001) Improved estimation procedures for multilevel models with binary response: a case study. Journal of the Royal Statistical Society, Series A, 164, 339–356.

77) Rosenzweig, M. R. and Stark, O. (1989). Consumption Smoothing, Migration, and Marriage: Evidence from Rural India. *Journal of Political Economy*, 97(4):905-926.

78) Senbet, D. (2010). Determinants of child labor versus schooling in rural Ethiopia. *European Journal of Social Sciences*, 17(3):414-425.

79) Shapiro, D. (1996). Fertility Decline in Kinshasa, Population Studies Vol50, No.1, pp 89- 103.

80) Sivaram, M., Richard, J., and Rao, P.S. (1995). Early marriage among rural and urban females of South India, *Journal of Biosocial Science*, 27 (3) 325-331.

81) Tezera A. (2013). Determinants of Early Marriage among Women in Ethiopia.

82) Thomas E. (2010-2011). Multilevel Survival Analysis of the Determinants of Age at First marriage Among Women living in Nigeria. Thesis for Master of Statistical Data Analysis, Gent University.

83) Uganda Bureau of Statistics (UBOS) and Macro International Inc. (2007): Uganda Demographic and Health Survey 2006. Calverton, Maryland, USA: UBOS and Macro International Inc

84) Uganda Demographic and Health Survey Report 1995 to 2006.

85) United Nations. (1990). Patterns of First Marriage: Timing and Prevalence. New York: United Nations, Department of International Economic and Social Affairs.

86) UN. (2007). The Millennium Development Goals, Report 2007, United Nations, New York. Retrieved 12, August, 2013 http://www.un.org/mi1lenniumgoals/pdf/mdg2007.pdf

87) UNFPA. (2006). In ending child marriage, a guide for global policy action International Planned Parenthood Federation and the Forum on Marriage and the Rights of Women and Girls. U.K.

88) UNICEF. (2005). Early Marriage: A Harmful Traditional Practice: A Statistical Exploration. UNICEF: New York, NY.

89) UNICEF. (2011). Child Protection from Violence, Exploitation and Abuse.

90) Vogl, T. S. (2012). Education and Health in Developing Economies. *Technical Report*, Princeton University.

91) Weissleder W. Amhara Marriage: The Stability of Divorce. The Canadian Review of Sociology and Anthropology. 1974; 11(1):67–85.

92) Wei, L. J., Lin, D. Y., & Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association*, 84(408), 1065–1073.

93) Westoff C. F. (2003). Trends in Marriage and Early Childbearing in Developing Countries. *DHS Comparative Reports No.5*. Macro International Inc.: Calverton, Maryland.

94) Wienke, A. (2011). Frailty Models in Survival Analysis. Taylor & Francis Group.

95) Win, H. I. (2009). Mothers of Ethiopia Part II: Escaping Early Marriage. http://www.hu_ngtonpost.com/hanna-ingber-win/mothers-of-ethiopia-partb 301245.html. Accessed: 7.31.2013.

96) Yamauchi, F. (2007). Marriage, schooling, and excess mortality in prime-age adults: Evidence from South Africa. *Technical Report*, International Food Policy Research Institute (IFPRI).

97) Yu B. (2006). Estimation of shared gamma frailty models by a modified EM algorithm. *Computational Statistics and Data Analysis 50, 463-474.*

98) Zahangir M. S., M. A. Karim, M. R. Zaman, M. I. Hussain and M. S. Hossain, 2008. Determinants of age at first marriage of rural women in Bangladesh: *A cohort analysis Trends Applied Science. Res. 4(3):335-343, Academic journal*, Department of Statistics.

99) Zahangir M.S. and Kamal M.M. (2011**).** Several Attributes Linked with Child Marriage of Females' in Bangladesh, *International Journal of Statistics and Systems, volume 6.*

Appendixes

Appendix I: Multivariable analysis using the Exponential-Gamma shared frailty model

Covariate	Coeff	St.err	P-value	ϕ	[95% CI for ϕ]	
Constant	-3.031104	0.0596437	0.000	0.0482623	0.0429378	0.0542471
Residence						
Urban(ref)						
Rural	0.030879	0.0452287	0.495	1.031361	0.9438695	1.126962
Education Level						
No Education(ref)						
Primary	-0.0423994	0.0270127	0.117	0.9584869	0.9090607	1.0106
Secondary	-0.2030898	0.0563942	0.000	0.8162049	0.7307964	0.9115952
Higher	-0.3363576	0.0731408	0.000	0.7143676	0.6189623	0.8244785
Religion						
Orthodox(ref)						
Catholic	-0.0013278	0.1045671	0.990	0.9986731	0.8136088	1.225832
Protestant	0.002006	0.038528	0.958	1.002008	0.9291291	1.080603
Muslim	-0.0341618	0.0301313	0.257	0.9664151	0.9109949	1.025207
Others	0.0146763	0.0841229	0.862	1.014785	0.8605345	1.196684
Wealth Index						
Poorest(ref)						
Poorer	-0.0047556	0.0338891	0.888	0.9952557	0.931297	1.063607
Middle	0.0243746	0.0348199	0.484	1.024674	0.9570774	1.097045
Richer	0.0341618	0.0353392	0.334	1.034715	0.9654722	1.108923
Richest	0.0829609	0.049664	0.095	1.086499	0.9857242	1.197577
Head Education						
No Education(ref)						
Primary	-0.0141782	0.0255151	0.578	0.9859218	0.9378298	1.03648
Secondary	-0.0425375	0.044501	0.339	0.9583545	0.878308	1.045696
Higher	-0.0719367	0.0625857	0.250	0.9305898	0.8231619	1.052038

Work Status						
No(ref)						
Yes	-0.0009102	0.0226749	0.968	0.9990902	0.9556608	1.044493
Access to media						
No(ref)						
Yes	-0.0241269	0.0261615	0.356	0.9761619	0.9273702	1.027521
Head Occupation						
Agriculturalist(ref)						
Professional	0.0162436	0.0577841	0.779	1.016376	0.9075459	1.138257
Laborers	-0.0245564	0.044406	0.580	0.9757427	0.8944105	1.064471
Business	-0.0067989	0.0389451	0.861	0.9932242	0.9202316	1.072006
Others	0.0162699	0.0445211	0.715	1.016403	0.9314713	1.109079
θ =0.0073, τ = 0.0036 Likelihood-ratio test of theta=0: chibar2(01)=49.80 prob>=chibar2(01)=0.000						

Appendix II: Multivariable analysis using the Weibull-Gamma shared frailty model

Covariate	Coeff	St.err	P-value	ϕ	[95% CI for ϕ]	
Constant	-12.48746	0.1259836	0.000	0.0000038	0.000003	0.000005
Residence						
Urban(ref)						
Rural	0.1775751	0.0469282	0.000	1.194318	1.089368	1.309378
Education Level						
No Education(ref)						
Primary	-0.0542326	0.0271327	0.046	0.9472118	0.8981559	0.9989471
Secondary	-0.641945	0.056457	0.000	0.5262678	0.4711406	0.5878454
Higher	-0.9080798	0.0740494	0.000	0.4032979	0.3488149	0.4662909
Religion						
Orthodox(ref)						
Catholic	-0.0213532	0.1054976	0.840	0.9788731	0.796025	1.203722
Protestant	-0.0390061	0.0395694	0.324	0.9617449	0.8899761	1.039301
Muslim	-0.0640181	0.0324551	0.049	0.937988	0.8801799	0.9995928
Others	-0.0560462	0.084963	0.509	0.9454954	0.8004584	1.116812
Wealth Index						
Poorest(ref)						
Poorer	0.0096157	0.0341398	0.778	1.009662	0.9443134	1.079533
Middle	-0.0272211	0.0351624	0.439	0.9731461	0.9083386	1.042577
Richer	0.0920278	0.0357719	0.010	1.096395	1.022158	1.176024
Richest	0.1026037	0.0514976	0.046	1.108052	1.001672	1.225731
Head Education						
No Education(ref)						
Primary	0.0322923	0.0256369	0.208	1.032819	0.9822051	1.086042
Secondary	-0.0691344	0.0446491	0.122	0.9332012	0.8550076	1.018546
Higher	-0.1615589	0.0628472	0.010	0.8508165	0.752212	0.9623465

Work Status						
No(ref)						
Yes	0.0124179	0.0228768	0.587	1.012495	0.9681002	1.058926
Access to media						
No(ref)						
Yes	0.0003885	0.0258996	0.988	1.000389	0.9508739	1.052482
Head Occupation						
Agriculturalist(ref)						
Professional	0.0902687	0.0590315	0.126	1.094468	0.9748897	1.228714
Laborers	-0.088559	0.0453177	0.051	0.9152491	0.8374615	1.000262
Business	-0.0604247	0.0395577	0.127	0.9413647	0.8711367	1.017254
Others	0.0541288	0.0448364	0.227	1.055621	0.9668142	1.152584

θ =0.0535, γ =4.228144, τ = 0.0261

Likelihood-ratio test of theta=0: chibar2(01)=398.29 prob>=chibar2(01)=0.000

Appendix III: Multivariable analysis using the Exponential-Positive Stable shared frailty model

Covariate	Coeff	St.err	P-value	ϕ	[95% CI for ϕ]	
Constant	-3.373	0.285	0.000	0.034280	0.020	0.600
Residence						
Urban(ref)						
Rural	-0.244	0.093	0.009	0.783	0.652	0.941
Education Level						
No Education(ref)						
Primary	-0.160	0.110	0.143	0.852	0.687	1.056
Secondary	-0.663	0.232	0.004	0.515	0.327	0.812
Higher	-0.742	0.254	0.003	0.476	0.290	0.783
Religion						
Orthodox(ref)						
Catholic	-0.264	0.464	0.570	0.768	0.309	1.908
Protestant	0.057	0.158	0.717	1.059	0.777	1.442
Muslim	0.076	0.123	0.535	1.079	0.848	1.373
Others	-0.156	0.593	0.793	0.856	0.268	2.734
Wealth Index						
Poorest(ref)						
Poorer	0.009	0.125	0.943	1.009	0.790	1.289
Middle	0.095	0.127	0.453	1.100	0.858	1.411
Richer	-0.117	0.136	0.388	0.890	0.681	1.161
Richest	0.202	0.184	0.271	1.224	0.854	1.755
Head Education						
No Education(ref)						
Primary	-0.003	0.099	0.978	0.997	0.822	1.211
Secondary	-0.030	0.179	0.866	0.970	0.683	1.378
Higher	0.001	0.188	0.996	1.001	0.692	1.448

Work Status						
No(ref)						
Yes	-0.066	0.084	0.433	0.936	0.793	1.104
Access to media						
No(ref)						
Yes	-0.060	0.098	0.542	0.942	0.777	1.141
Head Occupation						
Agriculturalist(ref)						
Professional	0.231	0.195	0.237	1.260	0.859	1.846
Laborers	0.024	0.183	0.893	1.024	0.717	1.466
Business	0.066	0.144	0.647	1.068	0.805	1.417
Others	0.177	0.163	0.278	1.194	0.867	1.645
θ =0.0059, τ = 0.0029						
Likelihood-ratio test of theta=0: prob>=chibar2(01)=0.003						

Appendix IV: Multivariable analysis using the Weibull-Positive Stable shared frailty model

Covariate	Coeff	St.err	P-value	ϕ	[95% CI for ϕ]	
Constant	-14.409	0.527	0.000	5.53e-07	1.97e-07	1.55e-06
Residence						
Urban(ref)						
Rural	0.650	0.234	0.005	1.916	1.211	3.032
Education Level						
No Education(ref)						
Primary	-0.165	0.111	0.139	0.848	0.682	1.055
Secondary	-1.258	0.240	0.001	0.284	0.178	0.455
Higher	-1.106	0.316	0.001	0.331	0.178	0.615
Religion						
Orthodox(ref)						
Catholic	-0.424	0.500	0.396	0.654	0.246	1.743
Protestant	0.579	0.228	0.011	1.784	1.142	2.788
Muslim	-0.034	0.148	0.818	0.967	0.723	1.292
Others	0.563	0.628	0.370	1.756	0.513	6.011
Wealth Index						
Poorest(ref)						
Poorer	-0.034	0.130	0.794	0.967	0.749	1.247
Middle	0.062	0.133	0.644	1.064	0.819	1.381
Richer	-0.274	0.141	0.052	0.940	0.576	1.003
Richest	0.525	0.201	0.009	1.690	1.140	2.504
Head Education						
No Education (ref)						
Primary	0.376	0.170	0.027	1.456	1.045	2.032
Secondary	-0.030	0.188	0.872	0.970	0.672	1.401
Higher	-0.039	0.189	0.837	0.962	0.680	1.492

Work Status						
No(ref)						
Yes	-0.158	0.088	0.073	0.854	0.718	1.015
Access to media						
No(ref)						
Yes	0.046	0.096	0.632	1.047	0.868	1.263
Head Occupation						
Agriculturalist(ref)						
Professional	0.420	0.223	0.06	1.522	0.983	2.359
Laborers	0.007	0.201	0.973	1.007	0.680	1.492
Business	0.135	0.154	0.382	1.145	0.846	1.546
Others	0.220	0.184	0.232	1.246	0.869	1.786
θ =0.0293, γ =4.500, τ = 0.0144						
Likelihood-ratio test of theta=0: prob>=chibar2(01)=0.000						

Appendix V: Multivariable analysis using the Log logistic-Positive Stable shared frailty model

Covariate	Coeff	St.err	P-value	ϕ	[95% CI for ϕ]	
Constant	2.914	0.069	0.000	18.432	16.111	21.086
Residence						
Urban(ref)						
Rural	0.968	0.291	0.001	2.633	1.490	4.656
Education Level						
No Education(ref)						
Primary	-0.234	0.121	0.053	0.791	0.624	1.003
Secondary	-1.367	0.257	0.001	0.255	0.154	0.422
Higher	-1.220	0.349	0.001	0.295	0.149	0.585
Religion						
Orthodox(ref)						
Catholic	-0.519	0.616	0.40	0.595	0.178	1.991
Protestant	0.387	0.231	0.094	1.473	0.936	2.314
Muslim	0.150	0.146	0.304	1.162	0.872	1.548
Others	0.390	0.626	0.534	1.477	0.433	5.035
Wealth Index						
Poorest(ref)						
Poorer	-0.039	0.140	0.779	0.962	0.730	1.265
Middle	0.191	0.146	0.191	1.210	0.909	1.612
Richer	-0.202	0.155	0.194	0.817	0.603	1.108
Richest	0.231	0.237	0.331	1.260	0.791	2.003
Head Education						
No Education(ref)						
Primary	0.047	0.112	0.674	1.048	0.842	1.305
Secondary	0.184	0.193	0.341	1.202	0.823	1.754
Higher	-0.090	0.208	0.666	0.914	0.608	1.374

Work Status						
No(ref)						
Yes	-0.164	0.121	0.174	0.849	0.670	1.075
Access to media						
No(ref)						
Yes	-0.011	0.106	0.916	0.989	0.803	1.217
Head Occupation						
Agriculturalist(ref)						
Professional	0.420	0.223	0.060	1.522	0.983	2.359
Laborers	0.259	0.206	0.209	1.296	0.865	1.941
Business	0.214	0.166	0.197	1.239	0.895	1.717
Others	0.220	0.184	0.232	1.246	0.869	1.786
θ =0.0516, γ =6.207, τ = 0.0252 Likelihood-ratio test of theta=0: prob>=chibar2(01)=0.003						

Appendix VI: Multivariable analysis using the Exponential-Inverse-Gaussian shared frailty model

Covariate	Coeff	St.err	P-value	ϕ	[95% CI for ϕ]	
Constant	-3.018282	0.075822	0.000	0.048885	0.0421343	0.0567175
Residence						
Urban(ref)						
Rural	0.505298	0.346959	0.016	1.65748	1.099682	2.498215
Education Level						
No Education(ref)						
Primary	-0.046384	0.035815	0.195	0.954676	0.8899588	1.0240990
Secondary	-0.206743	0.076796	0.007	0.813229	0.6995898	0.9453263
Higher	-0.313103	0.098642	0.002	0.731175	0.6026381	0.8871268
Religion						
Orthodox(ref)						
Catholic	-0.000193	0.153703	0.999	0.999807	0.7397479	1.351290
Protestant	-0.023830	0.051498	0.644	0.976452	0.882704	1.080155
Muslim	-0.035922	0.038937	0.356	0.964716	0.8938321	1.041220
Others	0.006252	0.119793	0.958	1.006272	0.7956959	1.272575
Wealth Index						
Poorest(ref)						
Poorer	0.002255	0.043812	0.959	1.002258	0.9197856	1.092125
Middle	0.034211	0.044976	0.447	1.034803	0.9474884	1.130165
Richer	0.000197	0.046367	0.997	1.000197	0.9133084	1.095351
Richest	0.104721	0.065707	0.111	1.110400	0.9762238	1.263018
Head Education						
No Education(ref)						
Primary	-0.010136	0.033382	0.761	0.989915	0.92722	1.056849
Secondary	-0.052083	0.060229	0.387	0.949250	0.8435552	1.068188
Higher	-0.089061	0.084319	0.291	0.914790	0.775441	1.079180

Work Status						
No(ref)						
Yes	-0.020801	0.030061	0.489	0.979414	0.9233749	1.038854
Access to media						
No(ref)						
Yes	-0.035061	0.034200	0.305	0.965547	0.902946	1.032487
Head Occupation						
Agriculturalist(ref)						
Professional	0.035354	0.075583	0.640	1.03599	0.8933414	1.201408
Laborers	0.011596	0.059432	0.845	1.011663	0.9004242	1.136645
Business	0.013997	0.051760	0.787	1.014095	0.9162645	1.122371
Others	0.044146	0.060906	0.469	1.045134	0.9275325	1.177647
θ =0.0062, τ = 0.0031						
Likelihood-ratio test of theta=0: chibar2(01) =26.63 prob>=chibar2(01)=0.003						

Appendix VII: Multivariable analysis using the Weibull-Inverse-Gaussian shared frailty model

Covariate	Coeff	St.err	P-value	ϕ	[95% CI for ϕ]	
Constant	-13.93654	0.468038	0.000	8.86e-07	3.54e-07	2.22e-06
Residence						
Urban(ref)						
Rural	0.558741	0.227003	0.014	1.748471	1.120554	2.728249
Education Level						
No Education(ref)						
Primary	-0.1275015	0.112221	0.256	0.880292	0.706487	1.096855
Secondary	-1.2706800	0.233997	0.000	0.280641	0.177408	0.443945
Higher	-1.123771	0.323940	0.001	0.325052	0.172272	0.613326
Religion						
Orthodox(ref)						
Catholic	-0.477261	0.501841	0.342	0.620481	0.2320386	1.659190
Protestant	0.155690	0.574596	0.320	1.168464	0.4456917	3.063349
Muslim	-0.012138	0.139950	0.931	0.987936	0.7509369	1.299732
Others	0.343990	0.496803	0.489	1.410564	0.5327373	3.734844
Wealth Index						
Poorest(ref)						
Poorer	-0.01617	0.133220	0.903	0.983960	0.7578450	1.277540
Middle	0.184408	0.136564	0.177	1.202506	0.9201195	1.571558
Richer	-0.171629	0.143796	0.233	0.842292	0.635423	1.116509
Richest	0.342976	0.197388	0.082	1.409134	0.9570526	2.074765
Head Education						
No Education(ref)						
Primary	0.0476394	0.103832	0.646	1.048792	0.8556728	1.285498
Secondary	0.042415	0.182545	0.816	1.043327	0.7295204	1.492119
Higher	-0.015821	0.190918	0.934	0.9843036	0.6770478	1.430997

Work Status						
No(ref)						
Yes	-0.171188	0.149108	0.333	0.842663	0.5957107	1.1919900
Access to media						
No(ref)						
Yes	0.0853311	0.097672	0.382	1.089078	0.8993326	1.3188560
Head Occupation						
Agriculturalist(ref)						
Professional	0.3570767	0.197426	0.071	1.429145	0.9705713	2.104386
Laborers	0.028705	0.198687	0.885	1.029121	0.6971782	1.51911
Business	0.078951	0.152717	0.605	1.082151	0.8022235	1.459757
Others	0.224765	0.220372	0.202	1.252028	0.8867319	1.767812
θ =0.0485, γ =4.540, τ = 0.0237						
Likelihood-ratio test of theta=0: chibar2(01) = 74.24 prob>=chibar2(01)=0.000						

Appendix VIII: Multivariable analysis using the Log logistic-Inverse-Gaussian shared frailty model

Covariate	Coeff	St.err	P-value	ϕ	[95% CI for ϕ]	
Constant	2.856438	0.0590654	0.000	17.399439	15.4973958	19.5349272
Residence						
Urban(ref)						
Rural	-0.088903	0.0385016	0.021	0.914934	0.8484325	0.9866487
Education Level						
No Education(ref)						
Primary	0.052504	0.0248485	0.035	1.053906	1.0038087	1.1065044
Secondary	0.342978	0.0511985	0.000	1.409138	1.2745980	1.5578787
Higher	0.319903	0.0687533	0.000	1.376993	1.2033970	1.5756324
Religion						
Orthodox(ref)						
Catholic	0.188048	0.1224905	0.125	1.206891	0.9493017	1.5343783
Protestant	0.505299	0.3469592	0.016	1.657480	1.0996820	2.498215
Muslim	0.073825	0.0262870	0.005	1.076619	1.0225543	1.1335416
Others	0.079859	0.1099428	0.468	1.083135	0.8731709	1.3435882
Wealth Index						
Poorest(ref)						
Poorer	-0.011892	0.0279111	0.670	0.988178	0.9355723	1.0437425
Middle	-0.051367	0.0251057	0.041	0.949930	0.9043186	0.9978413
Richer	-0.003031	0.0306045	0.921	0.996973	0.9389294	1.0586055
Richest	-0.092001	0.0409779	0.025	0.912105	0.8417133	0.9883824
Head Education						
No Education(ref)						
Primary	0.004402	0.0227117	0.846	1.004412	0.9606817	1.0501321
Secondary	-0.191999	0.0684146	0.021	0.825308	0.7015438	0.9709057
Higher	-0.004652	0.0450729	0.918	0.995359	0.9111997	1.0872910

Work Status						
No(ref)						
Yes	0.0334404	0.0196139	0.088	1.034006	0.9950103	1.0745296
Access to media						
No(ref)						
Yes	-0.017308	0.0227976	0.448	0.982841	0.9398917	1.0277525
Head Occupation						
Agriculturalist(ref)						
Professional	-0.063056	0.044793	0.159	0.938891	0.8599769	1.0250456
Laborers	-0.009134	0.041322	0.825	0.990908	0.9138180	1.0745015
Business	-0.010505	0.031916	0.742	0.989550	0.9295464	1.0534279
Others	-0.054119	0.036391	0.137	0.947319	0.8821060	1.0173548
θ =0.0754, γ =0.145, τ = 0.0363						
Likelihood-ratio test of theta=0: chibar2(01) = 50.67 prob>=chibar2(01)=0.003						

Appendix VIIII: Description and categories of explanatory variables.

Variables	Categories
Women level of education	0 = No education
	1=Primary
	2=Secondary
	3=Higher
Place of residence for women	1=Rural
	2=Urban
Household wealth index	1=Poorest
	2=Poorer
	3=middle
	4=Richer
	5=Richest
Women's religion	1=Orthodox
	2=Catholic
	3=Protestant
	4=Muslim
	5=Others
Educational level of Head	0=No education
	1=Primary
	2=Secondary
	3=Higher
Access to media	0=No
	1=Yes

Occupational status of Head	1=Agriculturalist
	2=Professional
	3=Laborers
	4=Business
	5=Others
Working status of the respondent	0=Yes
	1=No