DEBRE BERHAN UNIVERSITY

SPATIAL VARIATION AND DETERMINANT OF TIME TO FIRST BIRTH IN ETHIOPIA WOMEN AGE 15-49: SPATIAL AND PARAMETRIC FRAILTY MODELS

By

ZENEBE DESTA

A Thesis Submitted In Partial Fulfillment of the Requirements for the Degree of MSc in Statistics (Biostatistics)

DEPARTMENT OF STATISTICS, COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCE

FEBRUARY 2024

DEBRE BERHAN UNIVERSITY

SPATIAL VARIATION AND DETERMINANT OF TIME TO FIRST BIRTH IN ETHIOPIA WOMEN AGE 15-49

SPATIAL AND PARAMETRIC FRAILTY MODELS

By

ZENEBE DESTA

ADVISOR:- GEBREMEDHIN DESTA

Declaration

I, the undersigned, declare that the thesis entitled on **"Spatial Variation and Determinant of Time to First Birth in Ethiopia: Spatial and Parametric Frailty Models"** is my original work, and has not been presented for achieving any degree or diploma award in this university or any other institutions. All source of materials used for the thesis have been duly acknowledged. The thesis has been submitted in partial fulfillment for the requirements of Master of Science in Biostatistics, Debre Berhan University.

Name: Zenebe Desta Shenkut

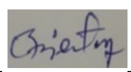
Signature: ________________

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

Gebremedhin Desta (assistant prof.)	____________________	______/____/ _______
Name of Advisor	Signature	Date

Approval Sheet

This is to certify that the thesis entitled on **“**Spatial Variation and Determinant of Time to First Birth in Ethiopia: Spatial and Parametric Frailty Models**”** submitted in partial fulfillment of the requirements for the degree of Master of Science in Statistics (Biostatistics) prepared by **Zenebe Desta** under my supervision and no part of the thesis has been submitted for any degree or diploma award in this university or any other institutions. Hence, the papers recommend that it can be accepted as fulfilling thesis requirements.

___________________________	______________________	______/____/ _______
Name of Chairman	Signature	Date
___________________________	______________________	______/____/ _______
Name of Internal Examiner	Signature	Date
___________________________		______/____/ _______
Name of External Examiner	Signature	Date

Acknowledgement

First and foremost, I would like to thank the almighty God, who helped and blessed me in all my activities, and to the Holy Virgin Mary. I only can say "Glory to the almighty God".

Secondly, I would like to express my deepest gratitude to my advisor; Gebremedhin Desta (Assistant Prof.) giving his support, valuable comments and advises by dissipates his golden time throughout the end. Really thank you very much for your involvement in my work and I sincerely thank all other lectures at Debre Berhan University, Statistics Department who helps me with valuable advises for my work.

I would like to thank also my intimate friends Mr. Mesfine Kiflu, Mr. Nigatu Getiye and Mr. Tayesew Shegaw for your valuable contribution by sharing your knowledge from the beginning of data management till the end of my thesis.

I would like to thank Dilla University for sponsoring and giving an opportunity to attend postgraduate program in order to update myself.

At last but not the least, I would like to thank all my best friends, my family members for their unreserved encouragement and supports ever time to be my accomplishment successful.

LIST OF ACRONYMS

AFT	Accelerated Failure Time
AIC	Akaike Information Criterion
ART	Assisted Reproductive Technology
CSA	Central Statistical Agency
EAs	Enumeration Areas
EDHS	Ethiopian Demographic and Health Survey
EMDHS	Ethiopia Mini Demographic and Health Survey
EPHI	Ethiopian Public Health Institute
ESPES	Enhancing Shared Prosperity through Equitable Services
FBI	First Birth Interval
GDHS	Ghana Demographic and Health Survey
ICPD	International Conference on Population and Development
IVF	In Vitro Fertilization
KM	Kaplan-Meier
LRT	Likelihood Ratio Test
MOFEC	Ministry of Finance Ethiopia
MOH	Ministry of Health
Pdf	probability density function
PH	Proportional Hazard
PHC	Population and Housing Census
SNNP	Southern Nations, Nationalities and Peoples
TFR	Total Fertility Rate

Table of Contents

LIST OF FIGURE

LIST OF TABLE

Abstract

The birth of the first child is the initial visible outcome of the fertility process, symbolizing a woman's transition into motherhood. Childbirth holds a significant transformative power, particularly for couples experiencing it for the first time. The main objective of this study is to find socioeconomic and demographic determinant and spatial variation of early first birth among reproductive women in Ethiopia and to estimate median age at first birth and spatial variation of prevalence of first birth and early birth in Ethiopia administrative areas. EDHS-2019 conducted by Ethiopian Central Statistical Agency (CSA) and other aid organization was used as a source of data for this study. The study included a total of 8885 women of age 15-49 years during the time of survey. Different survival model were used to explore factors associated with the age at first birth. Out of 8885 total reproductive age women included in the study, about 65.8% (5,846 women) had experienced their first birth (event), while the remaining 34.2% (3,039 women) had not undergone their first childbirth (censored). Among all models we have tested inverse Gaussian shared frailty model was the best model to fit the data as it had smallest AIC and BIC values. Some of the statistically significant factors affecting age at first birth were wealth index, education level, the use of contraceptive method and age of household head. Living in rural area, no formal education, wealth index were identified as predictors of early first births. Higher prevalence rates for the early age at first birth were found in Benishangul-Gumuz and Afar. In contrast, a lower prevalence rate was found in Hareri and Addis Ababa. Promoting education up to at least the primary education especially inrural area and Maximizing access to and utilization of contraceptives, and it is important to investigate factors associated with delayed first births among urban residents and those with primary, secondary and higher education levels

Key word: age at first birth, motherhood

1 CHAPTER ONE

1.1 Background

Childbirth holds a significant transformative power, particularly for couples experiencing it for the first time. Marriages often revolve around bringing new life into the world, making the age at first birth a crucial aspect that tends to shape the couple's happiness. In this article, we will explore the determinants of age at first birth and factors that influence age at first birth (Alazbih et al., 2023). Research on enhancing birth rates has gained substantial attention following the breakthrough of In Vitro Fertilization (IVF) in 1979. IVF comprises a series of procedures aimed at improving fertility, preventing genetic ailments, and assisting in conceiving a child. Authors in the mid-1980s took on the task of investigating the psychosocial impacts of infertility and IVF treatment. They found that infertility negatively affects emotional well-being, life satisfaction, and self-esteem. Whenever Assisted Reproductive Technology (ART) fails to enhance fertility, individuals frequently experience reduced life satisfaction, diminished self-confidence, and significant psychological distress (Iddrisu et al., 2020).

To study women's fertility behavior, scholars have utilized event histories linked to marriage and childbirth. The age at which a woman gives birth for the first time plays a pivotal role in her future life and holds a direct relationship with fertility. Without effective fertility control methods, the timing of a woman's initial childbirth has significant role on the number of children she bears throughout her reproductive years. Generally, women who give birth at a younger age have more children than those who give birth later in life (Alazbih et al., 2023). The birth of the first child is the initial visible outcome of the fertility process, symbolizing a woman's transition into motherhood. This event significantly shapes her future life and possesses a direct connection to fertility (Chernet et al., 2019).

In regions such as sub-Saharan Africa, where contraceptive usage is relatively low compared to more developed areas, early age at first birth tends to increase the number of children a woman will have. Even in areas where family planning is widespread, the timing of first births can impact the complete size of the family if contraception is utilized for spacing rather than limiting fertility (Ngalinda, 1998). In societies where births are confined to marriage, reproduction commences upon the onset of a valid marriage, and the time interval until the first birth is contingent on the demographic characteristics of women during the initial phases of married life (Mukhlesur et al., 2013).

Several Asian countries have managed to decrease fertility through government policies. Notably, China and Vietnam have experienced declines in their total fertility rate (TFR) due to stringent government policies discouraging early and arranged marriages (Lofstedt et al., 2005). However, a delayed fertility transition has been observed in certain regions, particularly in African countries like South Africa, Botswana, and Zimbabwe (Bongaarts, 2008; Moultrie et al., 2012). In comparison to the rest of the world, fertility rates in sub-Saharan Africa remain considerably higher. This region possesses a unique demographic scenario distinct from other world regions. While regions like Europe, South America, and Asia witnessed a decline in fertility rates during the 1950s and 1960s, sub-Saharan Africa stands out as the only region where the decline has been slower and occurred later (Ekane, 2013). According to Malmberg (2008), current fertility rates in sub-Saharan Africa mirror those of Asia and South America towards the end of the 1970s. Many countries in sub-Saharan Africa still experience relatively higher fertility rates.

The information presented so far indicates that sub-Saharan Africa is the only region worldwide that has not yet seen a significant decline in its fertility rates. According to the Nations (2014) report, out of 66 high-fertility countries (with over 3.2 children per woman), 45 are concentrated in sub-Saharan Africa. Ethiopia, with a population size of 94,351,001 and a doubling time of 29 years, is the second most populous country in sub-Saharan Africa after Nigeria (Ethiopia, 2013) along with the scarcity of resources. The country faces severe resource scarcity, and uncontrolled fertility has adversely impacted its socioeconomic, demographic, and environmental development. A weak infrastructure, low levels of education and health, poverty, war, famine, and low agricultural and industrial production have exacerbated the overpopulation problem (Ezra, 2001).

When we look back at the history of Ethiopia's population growth rate, there has been a steady increase since 1960. Ethiopia's population growth rate has steadily increased since 1960. The 1984 census revealed a growth rate of 2.3% for the 1960–70 period, 2.5% for the 1970–80 period, and 2.8% for the 1980–85 period. Population projections compiled in 1988 by the CSA estimated a growth rate of 2.83% for 1985–90 and 2.96% for 1990–95. According to the 2007

Ethiopian population census, the annual population growth rate between 1994 and 2007 stood at 2.6%. To formulate effective policies aimed at encouraging longer intervals between first births, it is crucial to study the impact of various socioeconomic and demographic factors on the time to first birth. This study examines the factors associated with time to first birth using parametric survival and spatial model.

1.2 Statement of Problem

High fertility in Ethiopia remains the dominant factor dictating the future size, growth, and composition of the population in the country. To reduce fertility and control the population growth of the country, the factors that influence fertility should be identified (Zhang, 2007). Experience of fertility transition countries also emphasizes the role of its determinant in fertility change (Bongaarts, 2011). The large population increases in many developing countries that followed the international epidemiological transition of the 1940s contributed to an increase in internal violent conflicts, including civil wars and violent protests(Acemoglu et al., 2020).

An early age at first birth often has a negative effect on further education and career-building, marital stability, asset accumulation, and on the woman's health. Girls who gave birth before the age of 18 were disproportionately affected by complicated maternal mortality and morbidity. Girls aged 10 - 14 years and girls aged 15 - 19 years were five times and twice respectively more likely to die in pregnancy or childbirth than women aged 20 – 24. For the formulation of effective policy to motivate people for longer time-to-first birth, it is necessary to study the effect of various socioeconomic and demographic factors that affect time to first birth and spatial distribution across region and zones.

Even though the issue of mother age at their first birth has wide ranging consequences on total fertility and health of mothers, not enough research has been conducted to identify its determinant factors and spatial variation. To identify factors influencing the timing of the first birth, use of appropriate statistical analysis method and use spatial model to show its variation among Ethiopian administrative region and zone plays a considerable role. Most studies have employed standard survival analysis methods like the Cox-proportional hazards (PH) model. However, the model has constantly been criticized for its restrictive assumption commonly referred to as the PH assumption and this research aimed to explore factors that affect time to first birth by using the parametric shared frailty model and spatial variation among region of Ethiopia. The Frailty term was added to account for the correlation that comes from the cluster, which accounts unobservable random effect and the paper have been, used spatial analysis to show both the spatial variation of age at first birth and early age at first birth. In general, the motivation behind this study is to address the following major research questions:

- ✓ What is the median age at first birth of Ethiopian women in Ethiopia based on EDHS 2019 data
- ✓ What are the key demographic and socioeconomic predictors of time to first birth in Ethiopia?
- ✓ To show the regional and zonal level of spatial variation of both early age at first birth and age at first birth in general?

1.3 Objective of the study

1.3.1 General Objective

- ❖ To estimate socioeconomic and demographic determinant and spatial variation of time to first birth among reproductive women in Ethiopia.

1.3.2 Specific Objectives

The specific objectives are:

- ✓ To estimate the median time to first birth for Ethiopian women.
- ✓ To estimate the survival time and compare the survival curves of time to first birth among different levels of covariates
- ✓ To assess the determinant factor associated with time to first birth for Ethiopian women.
- ✓ To show spatial variation of prevalence of first birth and early birth in Ethiopia administrative area.

1.4 Significance of the Study

The result of this study will provide information on the time-to-first birth among women in Ethiopia and its determinant factors. Specifically;

- ✓ To provide information about the covariates or risk factors of time to first birth for Ethiopian women based on recent data.
- ✓ Provide information to government and concerned bodies in setting policies and strategies.
- ✓ Use as a stepping stone for further studies related to time to first birth.

1.5 Scope of the study

In many studies, there are different types of constraint that depend on the availability of the resources and the area that the study is taken place. The scope of this study is limited to the dependent and independent variables which are concerned to assess the spatial distribution of time to first birth and identify associated factors in Ethiopia using EDHS 2019.

2 CHAPTER TWO

LITERATURE REVIEW

2.1 Time to First Birth

The Age at First Birth refers to the woman's age in years at the time of her first child's birth (1). The most important event in a woman's life is childbirth. It is a physiologically, mentally, socially, and culturally significant occurrence (2). The influence could be favorable in the short and long term, but it can also result in harmful and distressing occurrences (3). The first child's birth interval can be used as a fertility indicator. The first child is a significant event in the lives of women who are taking on more obligations. A woman's waiting period for her first child can influence her marriage's happiness or longevity. While delayed deliveries can cause disagreements, mistrust, and even marriage breakdowns, very early births, especially unexpected and unwanted ones, can do the same or much worse (Logubayom & Luguterah, 2013). According to Singh et al., (2006) among the various types of fertility data used, data on first birth interval have an upper hand over all other types of birth interval due to certain reasons. First, being the earliest and first event of the married life of a female, it hardly suffers from recall lapse; second, it is free from the inconsistent fluctuation of breastfeeding (G. Singh, 2007).

Fertility refers to the actual reproductive performance of women and it is the most important component of population dynamics and plays a major role in changing the size and structure of the population of a given area over time (Negussie, 2000). Many theoretical approaches have been developed to explain variations in fertility. The most common measure of fertility is the total fertility rate which is defined as the average number of births that a woman would have if she survived to the end of her childbearing age (Munda et al., 2012). Each marriage increases the likelihood of more children as women in the right age of childbearing and this leads to a high total fertility rate. Several indicators are used to measure fertility patterns, such as the first birth interval after marriage (Lloyd, 2005).

Risk Factors of Time to First Birth

The age at which individuals have their first child is influenced by a wide range of factors, including sociocultural, economic, educational, and demographic variables.

Socioeconomic Factors: Socioeconomic status plays a significant role in determining the age at which individuals have their first child. Research has consistently shown that women with higher levels of education and income tend to delay childbirth and have their first child at a later age(Kumar, 2006 & Islam 2009). This delay is often attributed to pursuing educational and career goals before starting a family.

Cultural and Social Norms: Cultural and social norms surrounding marriage, family, and fertility influence the age at first birth. In some societies, early marriage is common, which often leads to early childbearing. Conversely, in societies where women have greater autonomy and access to education and contraception, the average age at first birth tends to be higher(Zhenzhen, 2000).

Relationship Status: The timing of entering into a committed partnership or marriage also affects the age at first birth. Individuals who marry or form stable partnerships at an earlier age are more likely to have children earlier. On the other hand, individuals who delay or forgo marriage may delay childbearing as well.

Access to Reproductive Health Services: Access to reproductive health services, including contraception and family planning resources, can influence the timing of first births. Availability and affordability of modern contraceptives can enable individuals to delay childbearing until they are ready(Stokes & Hsieh, 1983).

Parental Background and Family History: Research has shown that individuals whose parents had early births are more likely to have their first child at an earlier age. It's important to note that the determinants of age at first birth can vary across countries, cultures, and socioeconomic contexts. Additionally, individual preferences, values, and personal circumstances also contribute to the decision-making process regarding childbearing. Understanding the determinants of age at first birth is crucial for policymakers and researchers to develop targeted interventions and policies that support reproductive health, education, and economic opportunities for individuals and promote informed family planning decisions

(Caldwell & Caldwell, 1987). While the former attributes variations in fertility behavior to socioeconomic and demographic differences among groups, the latter assigns a unique role to the availabilities of required materials like contraceptives as key to making variations in birth intervals. Time-to-first birth is associated with a couple's characteristics like age at marriage, education, occupation, and place of residence but with the influence of social norms. The age of women at first birth is an important determinant and it affects the growth of the population. Early childbearing increases the women’s reproductive span as compared to those similarly fecund women who bear children later. Education, residence urban/rural, age at first marriage, and marriage cohort played a significant role in the determination of marriage to the first birth interval. The college educated Taiwanese women had a longer timing than women who had completed only school education. Urban residents had a longer time than rural. The difference in time between urban and rural women was four months the family planning program had not attained the desired results and the prevalence rate in rural areas was lower than in urban.

Contraceptive use has shown an insignificant relationship with birth spacing (Stokes & Hsieh, 1983). Islam, (2009) also investigated the determinants of first birth in rural Bangladesh. Respondent's age, age of women at marriage, family income, and quality of care at the clinic were found as significant determinants. For the same country, using the Cox PH Model based on the Bangladesh Demographic and Health Survey (Haque et al., 2011), place of residence, region, women’s education, husbands' education, access to media, women's work status, wealth index, and contraceptive use were found to have a significant effect on time to first birth while religion and husbands' education were not. And Rabbi & Kabir, (2013) also used a Multivariate approach to determine significant factors of age at first birth. Accordingly, the age of women at marriage, and place of residence (urban as reference) were found to be negatively associated while mass media exposure, wealth index (poor as reference), and work status of women were positively associated with first birth interval.

Research conducted in Pakistan using the Cox Regression Model by considering the covariates: age of women at the first birth, age at marriage, ideal number of children (fertility intention), ideal number of boys (son preference), region (Punjab, Sindh, KPK, Baluchistan), education of both spouses, wealth index and occupation of both spouses was conducted to identify the potential covariates that affect time to first birth after marriage. This research revealed that women’s age at marriage, education (illiterate) and wealth index (poorer) contribute significantly to the first birth. However the ideal number of children (fertility intention), ideal number of boys (son preference), and education status of both spouses were insignificant. Shayan et al., (2014) in Iran investigated prognostic factors of first birth interval after marriage using Cox PH and Parametric Survival Models. The result showed that age at marriage, level of women’s education, and menstrual status had highly significant effects on the duration of birth interval after marriage but wealth index, both women's and husband's educational levels were not significant.

Hidayat et al., (2014) used Cox PH and Exponential distribution in modeling the first birth and associated factors in Indonesia. The result showed that place of residence, mother's education, and age at marriage significantly affects the age at first birth. However, knowledge about contraception and the work status of women was found to be insignificant. Nath et al., (2000) conducted a study on the effect of the status of women on the age at first birth in Indian Urban society. Education of women, work status, participation in family decisions, and age at marriage were taken as status variables along with socioeconomic variables (family income, family status, and caste system). Among these factors, the age of women at marriage, the education status of women, family income, and participation in family decisions were found to have a significant effect on the age at first birth. The caste system also had played insignificant contribution in the determination of time-to-first birth in India.

According to Obiyan et al., (2019), by using Semi-parametric Survival modeling for age at first birth in Nigeria, by using DHS data (2008), 31% of women who had their first child before age 15 years ended up having 7 or more children, only 8.3% of those who had first birth after age 25 years ended with this number. The corresponding mean numbers of children for these women are 6.62 and 2.62 respectively, a difference of 4.0. The risk of bearing the first child is lower for women who attend secondary and higher education than those who have no education. In addition, the working status of women and the use of contraceptive methods were found to be significant in determining first birth. A similar study in Nigeria using four parametric models whose various curves and estimates are compared with non-parametric values were considered, namely Inverse Gaussian, Log-logistic, Weibull, and Burr Type XII. The best model appears to be Inverse Gaussian based on the Akaike Information Criterion. In this study, the covariate, wealth index of the family, work status of women, education level of women and her partner, age at marriage of women, and place of residence were considered. The risk of giving her first birth for women living in rural, illiterate women, and women without jobs was higher than their respective counterparts. However, the education level of a husband had no contribution to the time of first birth after marriage. The mean and median waiting times for first birth after marriage by women in Nigeria are 28.8 and 20.0 months respectively (Simeon et al., 2014). Logubayom & Luguterah, (2013) used the Non-Parametric Survival Analysis technique and data from the 2008 Ghana Demographic and Health Survey (GDHS) to examine first birth interval after marriage. The study considered only women of childbearing age (15-49 years), who went into marriage without a child or a pregnancy. The result showed that region of residence, educational level of women, and wealth index of the family had significant effects while age at first marriage and age at first intercourse did not.

For the same country Obeng Gyimah, (2003), using regression analysis, also reported that women who had short first birth intervals tend to have a higher number of births than those whose first birth occur late regardless of their birth cohort. Gurmu & Etana, (2013), investigated Ethiopian Marriage to first birth interval Using Cox's proportional hazards model, which is significantly different for the age of women at marriage, region, education of women, and marriage cohort in Ethiopia. Of the nine regional states, the study showed that the Amhara region, where child marriage is commonly practiced, exhibits a longer interval between marriage and first birth. For the same country, Ethiopia (Nega & Woncheko, 2016), using the AFT model with 2011 EDHS, reported that the median survival time of first birth for rural women was 29 months. Nega & Woncheko, (2016) also used the AFT model to analyze the determinant of birth intervals in rural Ethiopia using the 2011 EDHS. They reported that the time to first birth was affected by region, educational level of the mother, and wealth index of the family. According to their finding, the estimated median time of first birth was 29 months.

3 CHAPTER THREE

Research Methodology

3.1 Data Source and Sampling design

The data for this study was extracted from the published reports Ethiopia Mini Demographic and Health Survey (EMDHS) which is the second Mini Demographic and Health Survey conducted in Ethiopia. The Ethiopian Public Health Institute (EPHI) implemented the survey at the request of the Ministry of Health (MoH). Data collection took place from March 21, 2019, to June 28, 2019. Financial support for the 2019 EMDHS was provided by the government of Ethiopia, the World Bank via MOFEC - Enhancing Shared Prosperity through Equitable Services (ESPES) and Promoting Basic Services Projects, the United Nations Children's Fund (UNICEF) and the United States Agency for International Development (USAID). ICF provides technical assistance through the DHS Program, which is funded by the United States Agency for International Development (USAID), and offers support and technical assistance for the implementation of population and health surveys in countries worldwide.

The sampling frame used for the 2019 EMDHS is a frame of all census enumeration areas (EAs) created for the upcoming 2019 Ethiopia Population and Housing Census (PHC), which will be conducted by the Central Statistical Agency (CSA). The census frame is a complete list of 149,093 EAs created for the 2019 PHC. An EA is a geographic area covering an average of 131 households. The sampling frame contains information about the EA location, type of residence (urban or rural), and estimated number of residential households. The 2019 EMDHS sample was stratified and selected in two stages. Each region was stratified into urban and rural areas, yielding 21 sampling strata. Samples of EAs were selected independently in each stratum in two stages. In the first stage, a total of 305 EAs (93 in urban areas and 212 in rural areas) were selected with probability proportional to EA size (based on the 2019 PHC frame) and with independent selection in each sampling stratum. In the second stage of selection, a fixed number of 30 households per cluster were selected with an equal probability of systematic selection from the newly created household listing. All women aged 15-49, who were either permanent residents of the selected households or visitors who slept in the household the night before the survey, were eligible to be interviewed (Ethiopian Public Health Institute Addis Ababa, 2019).

3.2 Variables in the study

The response (dependent) and predictor variables used in the model for the estimation of parameters are defined as follows.

3.2.1 The Response Variable

The response variable in this study was the age of the mother at her first birth. It is measured as the length of time from her date of birth until the age at which she gives her first birth which is measured in years and is assumed to be continuous. If a woman did not give birth until the date of the data collection, the data is said to be right censored. If, however, a woman has given birth until the date of the data collection, we say the event (first birth) has occurred. This kind of data, where the duration of time until some kind of event occurs is of interest, is called survival data or time to event data.

3.2.2 Explanatory Variables

The independent variables that were known to be associated with the timing of first birth and considered in this study are sex of household head, age of household head, wealth index, religion, region, education level, place of residence, contraceptive use and Marital status. These covariates are described together with their coding scheme in Table 3.1. Among these covariates only age of household head is continuous the rest of them are categorical.

Table 3. 1 Description of independent variables used in the analysis

Variables	Description	Categories
Region	Region of residence	1=Tigray, 2=Afar, 3= Amhara, 4= Oromia, 5=Somali 6= Benishangul, 7= SNNPR, 8= Gambela, 9= Harari, 10=Dire Dawa, 11= Addis Ababa
Residence	Place of residence	1= Urban , 2 =Rural
Women education	Women’s level of education	0= No education;1= Primary; 2= Secondary, 3 = Higher
Religion	Religion of respondent(women)	1= Orthodox, 2= Muslim,3= Catholic,4= Protestant,5=Other

Wealth index	Household wealth index	0= Lowest; 1= Second; 2=Middile,4=Fourth,5=Highest)
Age	Age of household head	Measure in year it continuous
Sex	Sex of household head	0= Female,1=Male
Marital status	Marital status of women	0= Unmarried/living alone ; 1= Married/union
Contraceptive	Use of Contraceptive	0 = Not use, 1=User

3.3 Method of Data Analysis

A survival model is a statistical model that is used to analyze data on the time to an event in this case time to first birth measured in year to achieve the objective. Mainly, non-parametric, semi-parametric, parametric survival (AFT and frailty) models were employed to estimate the survival time of age at first birth in Ethiopian women and spatial analysis were used to show the spatial variation of age at first birth across the region and zone level.

3.3.1 Survival data analysis

Survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs. The term survival analysis applies to techniques in which the data being analyzed is the time the process takes for a certain event of interest to occur. The use of survival analysis, as opposed to the use of different statistical methods is most important when there is censoring data (Aalen et al., 2008). It involves the modeling and analysis of data that have a principal endpoint, the time until an event occurs time to event data. By time we mean years months weeks or days from the beginning of the follow up of an individual until an event occurs. Survival data or time to event data, measures the time elapsed from a given origin to the occurrence of an event of interest. In survival analysis we usually refer to the time variable as survival time because it gives the time that an individual has survived over some follow-up period. We also use the term failure to define the occurrence of the event of interest even though the event may be a success such as recovery from therapy (Klein & Goel, 1992; Klein & Moeschberger, 2003).

Model Specification

In this study duration is a positive random variable, denoted by T, and measures the length of time from the start of their birth date to the date of event or to the end of the survey. Let the Greek letter delta (δ) denote a (0, 1) where 1 indicates that the event was observed and 0 indicates that the event was censored. The indicator of censoring, δ is defined as:

$$\delta = \begin{cases} 1 & if\ individual\ get\ first\ birth \\ 0 & if\ individual\ does\ not\ get\ first\ birth \end{cases}$$

The duration/survival time *T* time to specified event is a non-negative random variable from a homogeneous population and it is assumed that the length of this spell $t > 0$ is the realization of continuous random variable *T* with a cumulative distribution function and probability distribution function given by ***F(t)*** and ***f(t)*** respectively. From the cumulative distribution function $F(t)$ of *T*, which are $F(t) = P(T \leq t)$ and a probability density function $\mathbf{f}(\mathbf{t}) = \frac{dF(t)}{dt}$. The survivor function ***S(t)*** gives the probability that an individual survives longer than some specified time t is given by $S(t) = P(T > t)$ .The reverse cumulative distribution function of *T* provides a survival function **S(t)** and it is defined as $S(t) = 1 - F(t) = \int_t^{\infty} f(u)du$.

The survivor function ***S(t)*** and the failure function ***F(t)*** both satisfy the properties of probabilities. The survivor function is lies between zero and one and is strictly decreasing in ***t,*** ***S(t)*** is equal to one at the beginning of the spell and zero at infinity (Jenkins, 2005).

The hazard function ***h(t)*** is the instantaneous transition of motherhood at time ***t***, given that the women have survived up to that point in time t. This is the probability of women closing within a short interval, conditional on having survived up to the starting time of the interval. In addition, it is not a probability, but a rate, as it does not lie between 0 and 1, rather ranges between 0 and infinity. The hazard function, $h(t) > 0$ is given as (Lancaster, 1990).

$$h(t) = lim \lim_{\Delta t \to t} \frac{p(t \leq T < t + \Delta t / T \geq t)}{\Delta t}$$

Since *h(t)* is also equal to the negative of the derivative of ***lnS(t)***, we have the useful identity $s(t) = \exp\left(-\int_t^{\infty} h(u)du\right) = \exp(-H(t))$. The cumulative hazard function ***H(t)***, is the cumulative risk of failure occurring by time t or the sum of the risks an women faced going from duration 0 to t. where $\boldsymbol{H(t) = \int_0^t h(u)du = -\log[s(t)]}$.

3.3.1.1 Non-parametric Survival Methods

Survival data are conveniently summarized through estimates of the survival function and hazard function. Methods of estimating these functions from a sample of survival data are nonparametric or distribution free and it do not require specific assumptions to be made about the underlying distribution of the survival times. This method is used to present numerical or graphical summaries of the survival times for individuals in a particular group (Hanagal, 2019). The most widely used nonparametric estimate of the survival function for right censored data is undoubtedly the Kaplan-Meier (KM) estimator (Kaplan & Meier, 1958 and The log rank test, introduced by Mantel & Haenszel, 1959)

The Kaplan-Meier estimate of the survival function

The Kaplan-Meier (KM) estimator is the standard non-parametric estimator of the survival function used for estimating the survival probabilities from observed survival times both censored and uncensored (Collett, 2023; Klein & Moeschberger, 2003). Let $\boldsymbol{j}$ be the size of the risk set at j where the risk set denoted the women uncensored just before t(*j*). Let (j) be the number of observed events (j), *J* = 1...*r* Then the K-M estimator of t is defined by

$$\hat{S}(t) = \Pi_{j;t(j)} \left[1 - \frac{d(j)}{d(r)}\right]$$

This estimator is a step function that changes values only at the time of each event. The cumulative hazard function of the KM estimator can be estimated. If the largest observed time say t_n, corresponds to an event, then the KM estimated survival function reaches 0 (Kaplan & Meier, 1958). In this study the Kaplan-Meier Survival curve was fitted across different set of categorical covariates to compare the duration/survival status of women.

$\widehat{H}$(t) = −ln [$\hat{S}$(t)] where $\hat{S}_t$ is the KM estimator.

Median Survival Time

Median survival time is the time beyond which 50% of the individuals in the population under study are expected to survive and is given by that value t (50) which is such that S_t (50) = 0.5. Because the non-parametric estimates S(t) step functions, it will not usually be possible to realize an estimated survival time that makes the survival function exactly equal to 0.5. Instead, the estimated median survival time is defined to be the smallest observed survival time for which the value of the estimated survival function is less than 0.5. In mathematical terms, $\hat{t}(50) = \min \frac{t(i)}{\hat{S}(t_j)} \leq 0.50$ where t(i), is the observed survival time for the i^{th} individuals i = 1, 2, · · ·, n and t(j) is ordered recovery time, j = 1, 2, · · ·, r.

The Log–rank test

The log rank test, introduced by Mantel and Haenszel (1959), is a non-parametric statistical test for comparing two or more independent making any distributional assumption for two groups. The log-rank test statistic for comparing two groups is tested by constructing contingency tables for each unique event time, and compare observed with expected numbers of events. Let t_i be the i^{th} ordered event time in the two groups with $t_{(1)} < t_{(2)} < \ldots < t_{(k)}$ and d_{ij} is the number of subjects experiencing the event at time t_j in group i; r_{ij} is the number of individual women at risk at time $t_{(j)}$ in group i; d_j is the total number of women experiencing the event and r_j is the total number of women at risk. Under the null hypothesis, the expected number of subjects experiencing the event at time $t_{(j)}$ is given as $\boldsymbol{E(d_{1j}) = e_{1j} = \frac{d_j r_{1j}}{r_j}}$ and the variance e_{ij} is estimated by var(e_{ij}) = $\left(\boldsymbol{\frac{r_{1j} r_{2j} d_j (r_j - d_j)}{r^2{}_j (r_j - 1)}}\right)$ By constructing a contingency table for every observed event time, the resulting log rank test statistic is given by $\boldsymbol{Q_{logrank} = \frac{\left(\sum_{j=1}^{k}(d_{1j} - e_{1j})\right)^2}{\sum_{j=1}^{k} var(d_{1j})} \sim x^2(1)}$

The Log Rank test can also be extended to covariates with more than two categories which are approximated by $Q_L = \left[\sum_{j=1}^{m}(d_i - e_i)\right]^t \left[\sum_{j=1}^{m} var(d_i)\right]^{-1} \left[\sum_{j=1}^{m}(d_i - e_i)\right]$

In this study, the log-rank test of equality was conducted to show the survival experiences of women across the different groups.

3.3.1.2 Semi-parametric method

Cox Proportional Hazards Model

The proportional hazards model introduced by Cox (1972) is a regression model with event time as dependent variable. The Cox proportional hazard model has considerable flexibility and is the most widely used hazard regression model. It allows the inclusion of information about known covariates in models of survival data in an easy way and enables us to estimate the relationship between the hazard rate and covariates without having to make any assumptions about the shape of the baseline hazard function, referring the model as a semi-parametric model (Wienke, 2011).

Considering a set of covariates $X^t = X_1$,........, X_p the proportional hazards model writes down the hazard function of women i (i = 1, . . . , n) with covariates $x^t_i = (x_{i1}$,........,$x_{ip})$ as the product of a baseline hazard function common to all women, $h_0(t)$, and a multiplying factor depending on the value of X. Since this factor must be positive, it is common to use the exponential of a linear combination of these covariates. Thus, the relationship between the distribution of event time and the covariates can be described in terms of a model according to Cox (1972), in which the hazard rate at time t for an individual is $h_i(t) = \ h_0(t)\exp(\beta' X_i)$

Where, x_i is the p×1 observed vector of covariates for women i and β is the associated p×1 unobserved parameters vector. The model is also known as the proportional hazards model, since the hazard ratio for two individuals, j and i, defines as follows

$$\frac{\boldsymbol{h_j(t)}}{\boldsymbol{h_i(t)}} = \boldsymbol{exp\{\beta_1(x_{j1} - x_{i1}) + \beta_2(x_{j2} - x_{i2}) + \beta_k(x_{jk} - x_{ik})\}}$$

This is independent of time and all the time dependency is captured by the baseline hazard function which is common to all observations.

3.3.1.3 Parametric methods

Accelerated Failure Time (AFT) Models

The accelerated failure time model is an alternative if the proportional hazards assumption does not hold. In contrast to the proportional hazards model, the accelerated failure time model is best characterized in terms of the survival function (Duchateau and Janssen, 2008) and AFT models are less affected by the choice of probability distribution, the effect of the covariates is a multiplication of the expected survival time and it measured the direct effect of the predictor variables on the survival time instead of hazard. The common distributions of the AFT model include exponential, Weibull, log-logistic, lognormal and gamma distributions (Hanagal, 2011, Allison, 2010).

Weibull AFT model

Weibull model is the most widely used parametric survival model. It is a very flexible model that the hazard rate can be one of the three functions namely monotone increasing, constant, or decreasing function (Liu, 2012). Suppose that survival times are assumed to have a Weibull distribution with scale parameter **λ** and shape parameter **p**, then, the baseline hazard, survival and probability density function of the Weibull distribution are bellow table 3.2.

Exponential AFT model

The exponential distribution is the simplest parametric distribution function because its specification is based on a single parameter **λ** the baseline hazard, survival and probability density function of the exponential distribution are bellow table 3.2.

Log-normal AFT model

The lognormal distribution is another popular parametric function widely used in survival analysis. A random variable, T, is said to have a lognormal distribution, with parameters μ and σ, if log T has a normal distribution with mean μ and variance σ. The probability density, hazard and survival function of the Log-Normal distribution are given below table 3.2.

The log-logistic AFT model

Log-logistic model is an alternative to the foregoing Weibull distribution. This distribution is fairly flexible with two parameters, denoted by (θ, k). It is one of the parametric survival-time models in which the hazard is decreasing for $k \leq 1$ and hump-shaped for $k > 1$. If the baseline hazard function in the general accelerated failure time model is derived from a log-logistic distribution with parameters θ, k, then, a survival and hazard function given bellow table 3.2.

Table 0.1 Parametric distributions for baseline hazards

Distribution	f(t)	S(t)	h(t)	Parameter space
Exponential	$\lambda exp(-\lambda t)$	$\exp(-\lambda t)$	λ	$\lambda > 0$
Log-logistic	$\frac{\lambda \rho^{t^{p-1}}}{(1+\lambda t \rho)^2}$	$\frac{1}{1+\lambda t \rho}$	$\frac{\lambda \rho^{t^{p-1}}}{1+\lambda t \rho}$	$\lambda \in \mathcal{R}, \lambda > 0$

Weibull	$\lambda t\rho^{p-1}\exp(-\lambda t^{p})$	$\exp(-\lambda t^{p})$	$\lambda t\rho^{p-1}$	$\lambda, \rho > 0$
Log-normal	$\frac{1}{t\sigma\sqrt{2\pi}\exp(-\frac{\log(x-\mu)^2}{2\sigma}}$	$1-\Phi\left(\frac{log\ t}{\sigma}\right)$	$\frac{\frac{\Phi(\log t)}{\sigma}}{1-\Phi\left(\frac{log\ t}{\sigma}\right.}$	$\mu\in\mathcal{R}$, $\boldsymbol{\sigma}, \boldsymbol{t} > \mathbf{0}$

Baseline Survivor and Hazard Function

The survival time T is assumed to follow a distribution with density function f (t), then the survival function is given by **S(t) = P (T > t) =** $\int_{t}^{\infty} \boldsymbol{f(u)du}$.

The hazard function is a measure of the probability of failure during a very small interval, assuming that the individual has survived at the beginning of the interval. It is defined as:

$$\boldsymbol{S(t)} = \frac{\boldsymbol{f(t)}}{\boldsymbol{S(t)}} = \frac{\frac{\boldsymbol{d}}{\boldsymbol{dt}}\boldsymbol{S(t)}}{\boldsymbol{S(t)}}$$

The relationship between the survival and the hazard function is given by $\boldsymbol{S(t)} = \mathbf{exp}(-\int_{\mathbf{0}}^{\infty} \boldsymbol{h(u)du}$. Under the parametric approach, the baseline hazard function is defined as a parametric function and the vector of its parameters, say $\boldsymbol{\psi}$, is estimated together with the regression coefficients and the frailty parameters.

3.3.1.4 Modeling Frailty

The classical survival methods assume that up to the covariates included in the model, the population under study is homogeneous. However, in some applications it is more reasonable to consider the population as heterogeneous or as a collection of homogeneous clusters of individuals. Individuals may be exposed to different risk levels, and this even after controlling for known risk factors, or being grouped into clusters so that individuals from the same cluster share some common unobserved exposure. In fact, one cannot measure all risk factors related to the event of interest. This could be for economical or practical reasons or simply due to lack of knowledge about all factors affecting the event. The individuals who are more frail due to some unobserved factors will experience the event earlier while the individuals who are stronger will remain free of the event for a longer time, leading to more heterogeneity than what is taken into account by our model (Legrand, 2021).

Frailty models extend the Cox proportional hazards model by introducing unobserved frailties to the model. In this case, the hazard rate will not be just a function of covariates, but also a function of frailties. A frailty model is a random effects model that has a multiplicative effect on the hazard rates of all the members of the subgroups. In Univariate survival models, it can be used to model the heterogeneity among individuals, which is the influence of unobserved risk factors in a proportional hazards model. In multivariate survival models, the shared frailty model is used to model the dependence between the individuals in the group. In the multivariate case, unobserved frailty is common to a group of individuals (*Journal of the American Geriatrics Society* - *Wiley Online Library*, n.d.).

Shared Frailty Model

Many statistical models and methods proposed to model failure time data assume that the observations are statistically independent of each other. However, this does not hold in many applications. The shared frailty model is a conditional model in which frailty is common to all subjects in a cluster. The shared frailty model is responsible for creating dependence between event times. It is also known as a mixture model because the frailties in each cluster are assumed to be random. It assumes that given frailty, all event times in a cluster are independent. The shared frailty model was introduced by Clayton without using the notion of frailty and was extensively studied (Kleinbaum & Klein, 1996; Therneau et al., 2000).

Frailty models are the extensions of the proportional hazards model which is best known as the Cox model. The most popular model in survival analysis. Normally, in most clinical applications, survival analysis implicitly assumes a homogeneous population of individuals to be studied. This means that all individuals sampled in that study are subject in principle to the same risk (e.g., risk of death, risk of disease recurrence). In many applications, the study population cannot be assumed to be homogeneous but must be considered as a heterogeneous sample i.e., a mixture of individuals with different hazards. The shared frailty model assumes that individuals in a subgroup or pair share the same frailty u, but the frailty from group to group may differ. Conditional on the random term, called the frailty denoted by μ_i, the survival times in cluster I ($1 \leq i \leq n$) are assumed to be independent, the proportional hazard frailty model assumes

$$\boldsymbol{h_{(ij)}(t/x_i j, \mu_i) = \exp(\beta'^{x_{ij}} + \mu_i)\, h_o(t)}$$

Unobserved heterogeneity tests

In frailty models, θ is estimated to gain a sense of the degree of outcome heterogeneity between/among clusters. It indicates heterogeneity among clusters and a significant correlation among individuals within the same cluster when θ is substantial and significantly different from zero. The frailties, on the other hand, are exactly equal to one when θ is equal to zero, indicating that cluster effects are not present and that events are independent inside and among centers (Glidden and Vittinghoff, 2004). The models with and without frailty are compared using the likelihood ratio test. In other words, it is used to compare the null hypothesis $H_0 : \theta = 0$ to the alternative hypothesis $H_1 : \theta > 0$. The maximum likelihood method was used to estimate this heterogeneity parameter θ from the frailty models. Since the null hypothesis is at the boundary of the parameter space, the LR test statistic is not the usual x^2 but rather is a 50:50 mixture of chi-square distribution with 0 and 1 degree of freedom, denoted as $\tilde{x}^2{}_{01}$ was used, and thus requires careful consideration concerning the calculation of its p-value (Collet, 2015).

The Frailty Distributions

The frailty defined and denoted by $\mathbf{z_i}$ is an unobservable realization of a random variable **Z** with probability density function **f (.)** and the frailty distribution. Since **Zi** multiplies the hazard function **Z** has to be non-negative. The mean of **Z** is typically restricted to unity to separate the baseline hazard from the overall level of the random frailties.

The main difference between multivariate and univariate frailty models is the assumption of how frailty is distributed in the data. Shared (multivariate) frailty models assume that similar observations share frailty. i.e., that means the frailty distribution variability is related to a measure of dependence between clustered subjects, whereas it is rather interpreted as a measure of overdispersion which is caused either by misspecification or omitted covariates in the univariate case. For the study frailty distributions used gamma and the inverse Gaussian were considered. In both cases, a single heterogeneity parameter represents indexes of the degree of independence.

The Gamma Frailty Distribution

The gamma distribution has been widely applied as a mixture distribution for example (Collett, 2023). From a computational and analytical point of view, it fits very well with failure data. It is widely used due to its mathematical tractability. The density of a gamma-distributed random variable with a parameter is given by (Hanagal & Pandey, 2016).

$\boldsymbol{fZ(zi)} = \frac{\boldsymbol{zi^{\frac{1}{\theta}} exp\left(-\frac{zi}{\theta}\right)}}{\boldsymbol{\theta^{\frac{1}{\theta}}\, \Gamma\frac{1}{\theta}}}$, θ ≥ 0 where (.)s the gamma function, and its distribution is Γ (μ, θ) with μ is feed to 1 for identity and the variance θ with Laplace transformation $\boldsymbol{L(u) = \left(1 + \frac{\mu}{\theta}\right)^{-\theta}}$ the conditional survival function of the gamma frailty distribution is given by:

$$\boldsymbol{S_\theta\,(t) = \left[1 - \theta \ln\left(S(t)\right)\right]^{-\frac{1}{\theta}}}.$$

The conditional hazard function is given by:

$$\boldsymbol{h_\theta(t) = h(t)\left[1 - \theta \ln\left(S(t)\right)\right]^{-1}}$$

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

For the Gamma distribution, Kendall's Tau, which measures the association between any two event times from the same cluster in the multivariate case, $\boldsymbol{\tau = \frac{\theta}{\theta+2}, \theta \in (0, 1)}$

Inverse Gaussian frailty distribution

The inverse Gaussian (inverse normal) distribution was introduced as a frailty distribution alternative to the gamma distribution (Janssen, 2008). 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

$$f\ (z) = \frac{1}{\sqrt{2\pi\theta Z i^3}} \exp\left(-\left(\frac{1}{2\theta z}\right)(Z-1)^2\right)$$

The mean and the variance are 1 and θ, respectively with Laplace transform

$$L(s) = \exp(\frac{1}{\theta})(1 - \sqrt{1 + 2\theta s}), S \geq 0$$

For the inverse Gaussian frailty distribution, the conditional survival function is given by:

$$S_\theta(t) \;=\; \exp(\frac{1\,)}{\theta}(1 \;-\; [1 \;-\; 2\theta \ln S(t)]^{\frac{1}{2}})$$

The conditional hazard function is given by: $h_0(t) \;=\; h(t)[1 \;-\; 2\theta \ln S(t)]^{\frac{-1}{2}}$ here S (t) and with multivariate data, an Inverse Gaussian distributed frailty yields $\boldsymbol{\tau}$ given

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

Method of Parameter Estimation

Estimation of the frailty model can be parametric or semi-parametric. In the former case, a parametric density is assumed for the event times, resulting in a parametric baseline hazard function. Estimation is then conducted by maximizing the marginal log-likelihood (Casella & Berger, 2021). In the second case, the baseline hazard is left unspecified, and more complex techniques are available to approach that situation. Even though semi-parametric estimation offers more flexibility, parametric estimation is more powerful if the form of the baseline hazard is somehow known in advance. Frailty models account for the clustering present in grouped event time data.

For right-censored clustered survival data, the observation for subject j ∈ Ji = {1,…, ni} from cluster i ∈ I = {1,…,s} is the couple (y_{ij} , δ_{ij} where$y_{ij} = \min\left(t_{ij},\; c_{ij}\right)$is the minimum

between the survival time t_{ij} and the censoring time c_{ij}, and where $\delta_{ij} \;=\; I\left(t_{ij} \;\; \le c_{ij}\right)$ is the event indicator. When covariate information has been collected the observation will be $\left(y_{ij},\right.$ $\left.\delta_{ij}, X_{ij}\right)$ where X_{ij} denotes the vector of covariates for the ij^{th} observation. In the parametric setting, estimation is based on the marginal likelihood in which the frailties have been integrated by averaging the conditional likelihood to the frailty distribution.

Under assumptions of non-informative right-censoring and of independence between the censoring time and the survival time random variables, given the covariate information, the marginal log-likelihood of the observed data can be written as.

$l_{marg}\,(\boldsymbol{\psi}, \boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{Z}, \boldsymbol{X})$

$$= \sum_{i=1}^{s} 〚 \left\{ \left[\sum_{j=1}^{ni} \boldsymbol{\delta_{ij}} 〛 \left(\mathbf{log\,g}\left(\boldsymbol{h_0}\left(\boldsymbol{y_{ij}}\right)\right) + \boldsymbol{X_{ij}^T \beta} \right) \right] + \mathbf{log\,g}\left[(-\mathbf{1})^{di} \boldsymbol{L^d} \left(\left[\sum_{j=1}^{ni} \boldsymbol{H_0}(\boldsymbol{y_{ij}}) \mathbf{ex\,p}(\boldsymbol{X_{ij}^T}) \right] \right) \right] \right\}$$

Where $d_i = \sum_{j=1}^{ni} \boldsymbol{\delta_{ij}}$ is the number of events in the i^{th} cluster, and $L^q(.)$ the qth derivative of the Laplace transforms of the frailty distribution is defined as:

$L(s) = E[\exp(-Zs)] = \int_0^{\infty}(-Z_i s)\, f(Zi) dzi, \quad S \geq 0$, where ψ represents a vector of parameters of the baseline hazard function, β the vector of regression coefficients and θ the variance of the random effect.

The estimates of h_0, β, θ are obtained by maximizing the marginal log-likelihood of the above. This can be done if one can compute higher-order derivatives $L^q(.)$ of the Laplace transform up to $q = \max\{d1, d2, d3, \ldots, ds\}$. Symbolic differentiation is performed in R but is impractical here; mainly because this is very time-consuming (Wasserman, 2021).

Prediction of Frailties

Besides parameter estimates, the prediction of frailties is sometimes desirable. The frailty term Zi can be predicted by $Zi = E\left(\frac{Z}{zi}, \varphi, \beta, \theta\right)$, with Zi the data of the i^{th} cluster. This conditional expectation can be achieved.

$$\hat{Z}i = E\left(\frac{Z}{zi}, \varphi, \beta, \theta\right) = -\frac{\boldsymbol{l^{(di+1)}}\left(\sum_{j=1}^{ni} \boldsymbol{H_0}(\boldsymbol{y_{ij}}) \mathbf{ex\,p}\left(\boldsymbol{X_{ij}^T}\right)\right)}{\boldsymbol{L^{di}}\left(\sum_{j=1}^{ni} \boldsymbol{H_0}\left(\boldsymbol{y_{ij}}\right) \mathbf{ex\,p}(\boldsymbol{X_{ij}^T})\right)}$$

3.3.2 Spatial Analysis

Spatial analysis is a statistical method that is useful to specify geographical areas with high or low rates of disease/death occurrence and variability over the region or country(health & 2016, 2016).

In this study, the spatial models were used for the age at first birth dataset which are spatially arranged. Such spatial arrangement of the cluster can be modeled in several ways, including geostatistical approaches, where we use the exact geographical locations latitude and longitude of the cluster, and lattice approaches, where we use only the positions of the cluster relative to each other. The methodology was applied to the analysis of the age at first birth dataset. This information hopefully will enable us to implement the geostatistical model. Moreover, this offers the possibility of using spatial models to discern spatial patterns of age at first birth, and how these patterns distribute over the study area (Early age at firs birth and age at first birth). ArcGIS software, were utilized, to develop maps of the resulting fitted statistical model, which in turn are useful for detecting any spatial correlation. Further, different spatial covariance structures were explored in real datasets on age at first birth giving due attention to the various implementation issues of the models. Comparative analysis among the computing models was performed using different ways, including Deviance Information Criterion, as well as maps of the fitted clusters of community-level effects.

3.3.2.1 Spatial Autocorrelation Analysis

Spatial autocorrelation can be defined as a particular relationship between the spatial proximity among observational units and the numeric similarity among their values; positive spatial autocorrelation refers to situations in which the nearer the observational units, the more similar their values and vice versa for its negative counterpart. Moreover, when neighboring locations share a certain amount of information then we can say that the spatial autocorrelation or spatial dependence exists(Lee, 2016). Spatial dependence prevails in many geographical datasets and exemplifies Tobler's First Law of Geography i.e., things closer together are more similar than things further apart(*Moellering and Tobler 1972 - Google Scholar*, n.d.).

Global Moran's I is the more popular test statistic for spatial autocorrelation(Lee, 2016). Moran's I is a statistics that was used to measure whether age at first birth patterns are dispersed, clustered, or randomly distributed in the study area (Medicine & 1996, n.d.) by taking the entire data set and producing a single output value which ranges from −1 to +1. Moran's I values close to −1 indicate spatial distribution is dispersed, whereas Moran's I values close to +1 indicate spatial distribution is clustered, and an I value of 0 means distributed randomly(Moran, 1950). A statistically significant Moran's I ($p<0.05$) leads to rejection of the null hypothesis (age at first birth is randomly distributed) and indicates the presence of significant spatial autocorrelation or dependence(Anselin & Getis, 1992).

Global Moran's I examine whether a spatial correlation exists or not over an entire cluster, and it is calculated as follows(Overmars et al., 2003):

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i-\bar{x})^2(x_j-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \qquad i \neq j \quad \ldots\ldots\ldots\ldots\ldots(3.1)$$

Where n is the number of observations of the whole cluster, x_i and x_j are the observations at locations of i and j, $\bar{x}$ is the mean of x, and w_{ij}, an element of spatial weights matrix *w*, is the spatial weight between locations of i and j.

3.3.2.2 Hotspot Analysis of age at first birth and early age at first

Local Getis-Ord Gi* statistics(Ord et al., 1995) are important to identify the hot and cold spot areas using GPS latitude and longitude coordinate readings that were taken at the nearest enumeration areas or EDHS 2019 clusters. An age at first birth hotspot refers to the occurrence of high prevalence rates of age at first birth clustered together on the map, whereas cold spot refers to the occurrence of low prevalence rates of age at first birth clustered together on the map(Ord et al., 1995).

Anselin local Moran's I cluster map was used to determine the local level of clustering of the age at first birth. Furthermore, local Moran's index measures the correlation of clusters whether positively correlated (High-High and Low-Low) clusters or negatively correlated clusters (High-Low and Low-high) that mean an outlier cluster. And also clusters of high values (High-High) and clusters of low values (low -low) were identified. The level of clustering of the age at first birth within the area was identified using Z-score. High and low level of clustering of age at first birth was indicated a positive and negative Z-scores respectively. Bernoulli model was employed to test for the presence of statistically significant spatial clusters of age at first birth (Kulldorff, 2006). Age at first birth was taken as cases and those who are censored as controls to fit the Bernoulli model. The number of cases in each location has a Bernoulli distribution and the model required data for cases, controls, and geographical coordinates.

The default maximum spatial cluster size of <50% of the population was used as an upper limit, which allowed both small and large clusters to be detected, while clusters that contained more than the maximum limit were ignored.

Z-score were computed to determine the statistical significance of clustering, and the p-value is used to determine if the number of observed age at first birth within the potential cluster is significant or not (Kulldorff, 2006). The null hypothesis of no clusters is rejected when the p-value ≤ 0.05.

3.3.2.3 Spatial Interpolation

It is very expensive and laborious to collect reliable data in all areas of the country to know the burden of a certain event. Therefore, part of a certain area can be predicted using observed data by a method called interpolation. The spatial interpolation techniques were used to predict age at first birth for un-sampled areas in the country based on sampled enumeration area measurements. There are various deterministic (i.e., create surfaces from measured points) and geostatistical (i.e., utilize the statistical properties of the measured points) interpolation methods. Among all of the methods, geostatistical interpolation techniques such as Ordinary Kriging and Empirical Bayesian Kriging were considered the best methods since they incorporate spatial autocorrelation and statistically optimize the weight(Bhunia et al., 2018). In this study, Ordinary Kriging interpolation method was used to predict age at first birth in unobserved areas of Ethiopia.

3.3.2.4 Ordinary Kriging (OK) Method

The ordinary kriging method incorporates statistical properties of the measured data (spatial autocorrelation). The kriging approach uses the semivariogram to explain the spatial continuity(Hoogland et al., 2010).

The semivariogram measures the strength of the statistical correlation as a function of distance(Bhunia et al., 2018). The range is the distance at which the spatial correlation vanishes, and the sill corresponds to the maximum variability in the absence of spatial dependence(Bhunia et al., 2018). The coefficient of determination (R^2) were employed to determine the goodness of fit(*(Robertson 2008) - Google Scholar*, n.d.).

3.4 Model Selection and Diagnostics checking

3.4.1 Comparison 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. 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 the Akaike Information Criterion (AIC). A data-driven model selection method such as an adapted version of Akaike's information criterion AIC is used to find the truncation points of the series. In some circumstances, it might be useful to easily obtain AIC value for a series of candidate models (Breiman, 2017; Hoerl & Kennard, 1970). In this study, we used the AIC criteria to compare various candidates of parametric frailty models.

The model with the smallest AIC value is considered a better t. For comparing models that are non-nested type, Akaike's information criterion (AIC) is defined as:

$$AIC = -2\log(L) + 2(k + c + 1)$$

Where k is the number of covariates, c *is* the number of model-specific distributional parameter. The preferred model is the one with the lowest values of the AIC. In addition to these criteria, a likelihood ratio test (LRT) will be used to compare models that are nested type, particularly the effect of the random effects.

Evaluation of the Baseline Parameters

The graphical methods can be used to check if a parametric distribution fits the observed data or not. Appropriateness of assumed distributions baseline hazard function is evaluated as follows: The appropriateness of the model with the exponential baseline can graphically be evaluated by plotting

- ❖ $\left(-\log\left(\hat{s}(t)\right)\right)$ versus t where S(t) is the Kaplan-Meier survival estimate. This plot should be linear, because for the exponential distribution, $S(t) = \exp(\lambda t)$, and hence, $-\log\left(S(t)\right) = \lambda t$ is linear with time.

- A model with the Weibull baseline has a property that the log(-log(S(t)) is linear with the log of time, where $S(t) = exp(\lambda t^{p})$. Hence, $log\left(-log\left(S(t)\right)\right) = log(\lambda) +\backslash$ $\rho\log(t)$.property allows a graphical evaluation of the appropriateness of a Weibull model by plotting $\log\left(-\log\left(\hat{S}(t)\right)\right)$ versus log(t) where $\hat{S}(t)$ is Kaplan-Meier survival estimate [86].
- log-normal baseline plot of $\Phi^{-1}\left\{1 - exp\left(-H(t)\right)\right\} = \Phi^{-1}\{1 - \hat{s}(t)\}$ versus log time (t) should be linear if the log-normal distribution is appropriate.
- The log-failure odd versus log time of the log-logistic model is linear. The failure odds of the log-logistic survival model can be computed as:

$$\frac{1 - S(t)}{S(t)} = \frac{\frac{\lambda t^{p}}{1 + \lambda t^{p}}}{\frac{1}{1 + \lambda t^{p}}} = \lambda t^{p}$$

Therefore, the log-failure odds can be written as:

$$\frac{\log\left(1 - S(t)\right)}{S(t)} = \log(\lambda t^{p}) = \log(\rho) + \rho\log(t)$$

Therefore, the appropriateness of the model with the log-logistic baseline can graphically be evaluated by plotting $\log((1 - \left(\frac{\hat{S}(t)}{\hat{S}(t)}\right))$ versus log time where $\hat{S}(t)$.

3.5 Model Diagnostics checking

The Cox Snell Residuals

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 & Moeschberger, 2003; Ko, 2000). The first such residual is the Cox-Snell residual which provides a check of the overall fit of the model. The Cox-Snell residual, r_j, is defined by:

$$\widehat{H}\left(\frac{Tj}{Hj}\right)$$

Where $\widehat{H}$ is the cumulative hazard function of the fitted model. If the model fits the data, then the rj's should have a standard $(\lambda = 1)$ exponential distribution, so that a hazard plot of rj versus the Nelson Aalen estimator of the cumulative hazard of the rj_0s should be a straight line with slope 1. The three baseline hazard functions will be considered in this thesis.

Table 0.2Cox Snell residuals

Exponential	$\lambda \hat{t} i \exp(\widehat{\beta},\ Xj)$
Log-logistic	$\ln\left(\frac{1}{1+\lambda \hat{t} i^{\rho} \exp(\widehat{\beta}, Xj)}\right)$
Weibull	$\lambda \hat{t} i^{\rho} \exp(\widehat{\beta}, Xj)$
Log-normal	$\ln\left[1-\Phi \frac{(lnt_j-\ \hat{\mu}-\hat{\gamma}^t Z_j\ )}{\hat{\delta}}\right]$

4 CHAPTER FOUR

RESULT AND DISCUSSION

4.1 Descriptive Statistics

Descriptive statistics were employed to gain insights into the distribution of variables in our study. The primary focus was on the response variable, which represents the survival time measured in years from the date of birth to the date of the first childbirth. This variable is continuous, and the survival event is coded as 1 if the event occurred (i.e., first childbirth) and 0 if censored. The study included a total of 8,885 women aged 15-49 during the data collection period.

Of this total, 65.8% (5,846 women) had experienced their first birth(event), while the remaining 34.2% (3,039 women) had not undergone their first childbirth(censored). The mean age at first birth was calculated as 18 years, with a 95% confidence interval ranging from 15 to 21 years. This suggests that the average age at first birth is relatively high in Ethiopia.

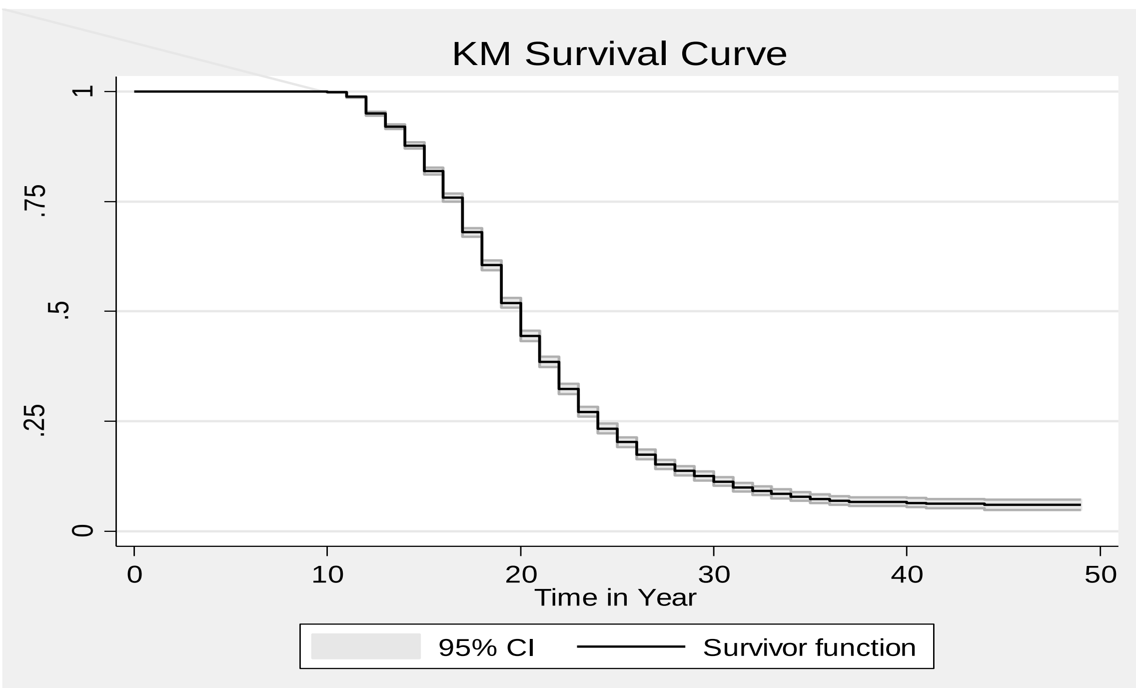
Analyzing the regional distribution, Oromia 12%, SNNPR and Amhara had an 11% experience rate, while Addis Ababa recorded the rate at 9.2%. other region Gambella 8.1%, Tigray 8.2%, Somalia 8.4%, Benishangul gumuz 7.2%, Harari 8.6%, Dire dawa 9.1% and Afar 7.2% of the total. In terms of rural versus urban areas, approximately 72% of events occurred in rural regions and 28% in Urban area of Ethiopia. Education-wise, 54% of women respondents had no formal education, while 31% had completed primary education. Regarding wealth indexes, 42% were classified as higher income, and 43% fell below the middle-income category. When considering the religious affiliation of women, approximately 43% were Muslim, and 35% identified as Orthodox. Of the total respondents, 86% were in a union or married, while 14% were unmarried or living alone. In terms of contraceptive use, 68% of women utilized either modern or traditional methods, while the remaining 32% did not use any contraception. Examining the household head's gender, 75% were male, and the remaining 25% were female.

Table 4. 1 descriptive analysis table

		Survival Event		
Variable	**Categories**	**Overall (8885)**	**Censored (3039)**	**Event (5846)**
Region	Addis Ababa	818 (9.2%)	444 (15%)	374 (6.4%)
	Afar	641 (7.2%)	141 (4.6%)	500 (8.6%)
	Amhara	948 (11%)	305 (10%)	643 (11%)
	Benishangul	747 (8.4%)	224 (7.4%)	523 (8.9%)
	Dire Dawa	812 (9.1%)	343 (11%)	469 (8.0%)
	Gambela	723 (8.1%)	194 (6.4%)	529 (9.0%)
	Harari	763 (8.6%)	287 (9.4%)	476 (8.1%)
	Oromia	1,052 (12%)	344 (11%)	708 (12%)
	SNNPR	1,008 (11%)	318 (10%)	690 (12%)
	Somali	640 (7.2%)	210 (6.9%)	430 (7.4%)
	Tigray	733 (8.2%)	229 (7.5%)	504 (8.6%)
Residence of residence	Rural	5,934 (67%)	1,736 (57%)	4,198 (72%)
	Urban	2,951 (33%)	1,303 (43%)	1,648 (28%)
Education level	Higher	751 (8.5%)	421 (14%)	330 (5.6%)
	No Education	3,640 (41%)	464 (15%)	3,176 (54%)
	Primary	3,345 (38%)	1,534 (50%)	1,811 (31%)
	Secondary	1,149 (13%)	620 (20%)	529 (9.0%)
Media Access	Yes	7,455 (84%)	1,609 (53%)	5,846 (100%)
	No	1,430 (16%)	1,430 (47%)	0 (0%)
Wealth Index	Fourth	1,344 (15%)	455 (15%)	889 (15%)
	Highest	2,901 (33%)	1,320 (43%)	1,581 (27%)
	Lowest	2,031 (23%)	516 (17%)	1,515 (26%)
	Middle	1,268 (14%)	375 (12%)	893 (15%)
	Second	1,341 (15%)	373 (12%)	968 (17%)
Contraceptive use	Yes	2,117 (24%)	224 (7.4%)	1,893 (32%)
	No	6,768 (76%)	2,815 (93%)	3,953 (68%)
Religion	Catholic	78 (0.9%)	25 (0.8%)	53 (0.9%)
	Muslim	3,635 (41%)	1,135 (37%)	2,500 (43%)
	Orthodox	3,374 (38%)	1,311 (43%)	2,063 (35%)
	Other	87 (1.0%)	22 (0.7%)	65 (1.1%)
	Protestant	1,711 (19%)	546 (18%)	1,165 (20%)
Marital status	Yes	5,613 (63%)	596 (20%)	5,017 (86%)
	No	3,272 (37%)	2,443 (80%)	829 (14%)
Sex of HH	Male	6,380 (72%)	2,024 (67%)	4,356 (75%)
	Female	2,505 (28%)	1,015 (33%)	1,490 (25%)
Age of Household Head		40 (30, 50)	45 (33, 57)	38 (30, 46)
Age at First Birth		18.0 (16, 21)	18.0 (16, 22)	18.0 (15, 21)

4.2 The Nonparametric Estimation Results

In Figure 4.1, the Kaplan-Meier survival estimate illustrates a steady decline in the survival probability of women from the age of 10 onwards, with the likelihood of surviving to 18 years approximately at 75% and decreasing to around 25% after age 25. This dynamic portrayal indicates the evolving distribution of survival times within the studied population. Concurrently, Figure 4.1 captures the variation in hazard rates, revealing an escalating risk of the event (likely first childbirth) from the 10th to the 35th year, followed by a deceleration. Notably, the peak hazard rate is observed between the ages of 15 and 30. Together, these figures provide an age related dynamics, shedding light on critical periods of heightened risk and the changing survival probabilities among women. All other categorical variable KM curve in appendix shows the group difference clearly, this may an indication the variable may be affecting the age at first birth.



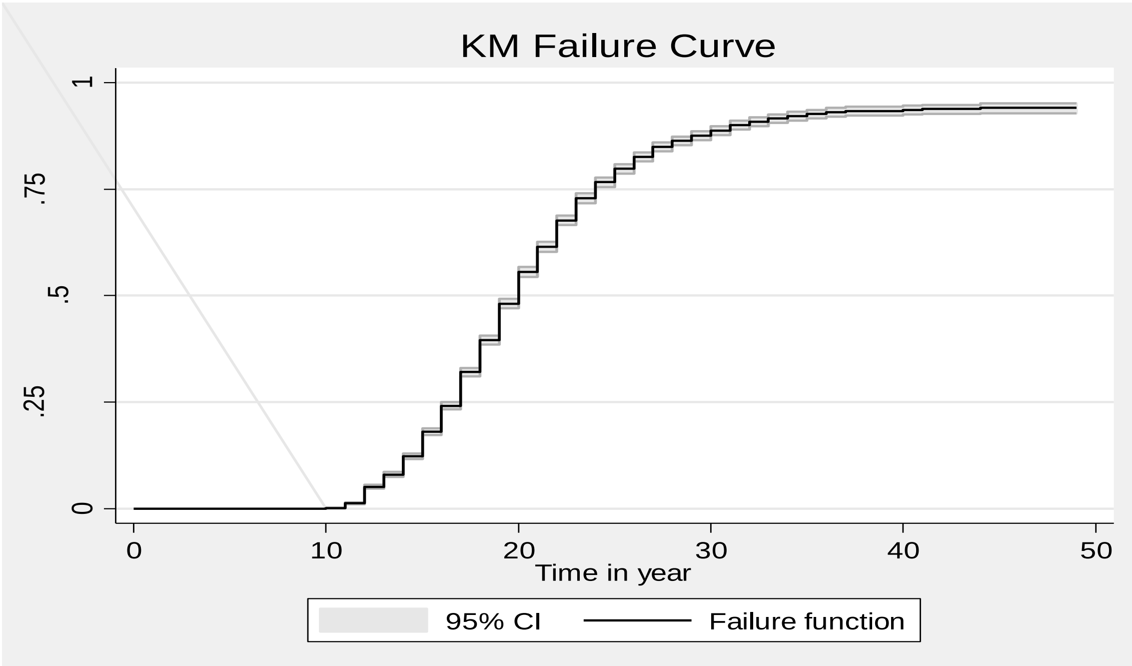


Figure 0.1 KM survival and failure curve

4.2.1 Median survival time

The median survival time for age at first childbirth in Ethiopian women is 20 years as shown in figure below. This is the time 50% of women survive until age 20 and 50% of women experience the event of first birth at 20 years.

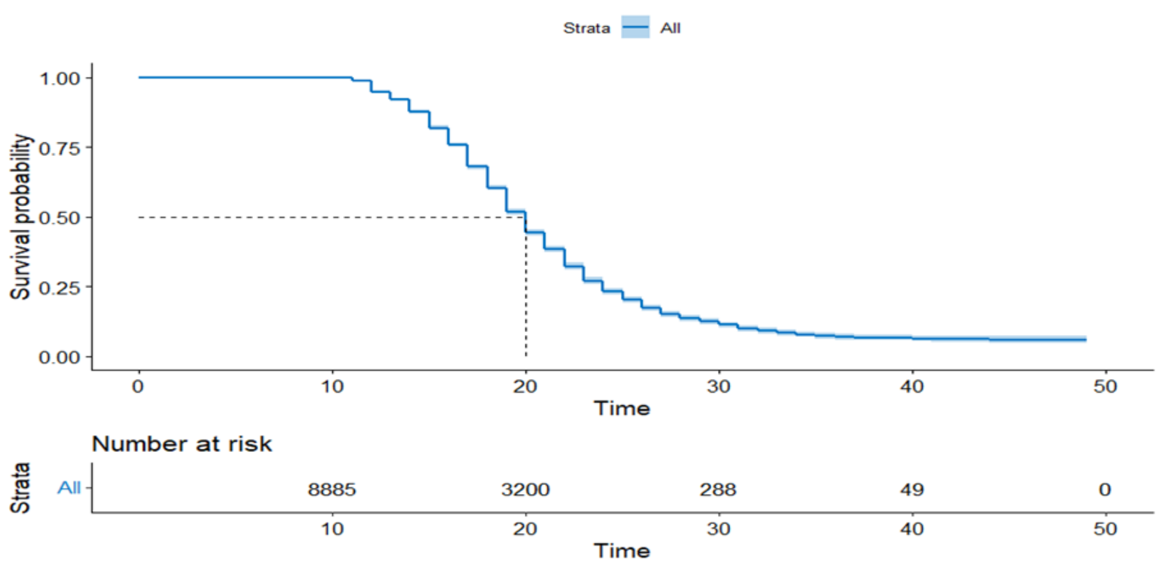


Figure 0.2 Median survival curve

4.3 Log rank test for each categorical variable

The log-rank test was applied to assess the significance of differences among the estimated Kaplan-Meier (KM) survival curves across various categories of women covariates. The results, as presented in Table 4.2, demonstrated a high level of significance, indicating substantial variations in survival experiences. The log-rank test revealed statistically significant differences in survival probabilities for covariates such as Region, Residence, Wealth Index, Use of Contraceptive, Marital Status, Education Level, Religion, Media access and Sex of Household Head. This implies that all categorical variables exhibited distinct survival experiences, emphasizing the diverse impact of these factors on survival probabilities. The statistical significance underscores the importance of considering these covariates when examining survival dynamics among women, highlighting the nuanced interplay between demographic characteristics and survival outcomes.

Table 4. 2 Results of the log-rank test for each categorical variable

Covariates	Chi-square	Df	P-value
Region	636.07	10	0.0000
Residence	546.19	1	0.0000
Level of Education	1024.01	3	0.0000
Wealth Index	680.05	4	0.0000
Contraceptive use	230.20	1	0.0000
Marital status	1478.00	1	0.0000
Religion	143.38	4	0.0000
Sex of HH	125.79	1	0.0000

4.4 Checking the PH Assumption

Before progressing to further analyses with the Accelerated Failure Time (AFT) model, a meticulous examination of the goodness of fit for the Cox Proportional Hazards (PH) model was undertaken. The scaled Schoenfeld residuals were employed to scrutinize the assumption of proportional hazards, revealing non-proportional hazards as evidenced by the residuals. This signifies that the hazard ratios for covariates change over time, indicating a temporal variation in their impact on survival. Additionally, a global test in table 4.3 assessing the assumption of proportionality in the Cox-PH model yielded a highly significant result (p-value=0.0000), further substantiating the departure from proportional hazards. Given these outcomes, careful consideration and potential alternative modeling approaches, such as the AFT model, are warranted to better capture the dynamic nature of covariate effects on survival outcomes.

Table 4. 3 ***Goodness of fit testing for PH assumption***

	chi2	Df	Prob>chi2
global test	**348.57**	**26**	**0.0000**

4.5 Accelerated Failure Time Models

4.5.1 Uni-variable AFT model

Prior to proceeding directly to multivariable analysis, uni-variable analyses were conducted for each covariate with a p-value less than 0.25, employing various AFT models, including Exponential, Weibull, Log-logistic, and Log-normal distributions. The AFT models revealed that covariates such as region, residence, level of education, wealth index, marital status, use of contraceptive methods, sex of household head, age of household head and religion were found to be significant and access to media were insignificant at a 25% significance level (Table 4.4). These significant variables emerged as candidate predictors for further analysis in the study. Subsequently, multivariable AFT models were fitted, including all the significant factors identified in the uni-variable analysis at a 5% significance level. Model comparison was performed using these significant covariates for each AFT model, and the best models were selected based on the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).

Table 4. 4 ***Uni-variable AFT models result***

	Distribution							
	Exponential		**Weibull**		**Lognormal**		**Loglogistic**	
Covariate	**LR X^2**	**P**	**LR X^2**	**P**	**LR X^2**	**P**	**LR X^2**	**P**
Residence	157.92	0.000	714.97	0.000	560.56	0.000	570.19	0.000
Education	573.13	0.000	1073.84	0.000	1175.27	0.000	1257.16	0.000
Wealth	203.07	0.000	832.01	0.000	708.79	0.000	728.96	0.000

Contraceptive	217.15	0.000	319.93	0.000	132.92	0.000	116.40	0.000
Marital status	1532.48	0.000	1727.58	0.000	1345.58	0.000	1390.42	0.000
Sex of HH	43.01	0.000	211.09	0.000	104.96	00100	93.26	0.000
Age of HH	131.52	0.000	440.52	0.000	199.30	0.000	191.91	0.000
Religion	40.59	0.000	165.13	0.000	149.42	0.000	154.95	0.000

4.5.2 multivariable AFT models

In the multivariable analysis, AFT models were employed, specifically using the exponential, Weibull, log-normal, and log-logistic distributions. These models were fitted by incorporating all the factors identified as significant in the univariate analysis at a 25% level of significance. The selection of covariates at this threshold aimed to capture a broader range of potential predictors that could influence the duration times of women age at first childbirth. The model comparison process involved evaluating the performance of each AFT model based on the significant covariates. Two widely used model selection criteria, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), were employed to assess the goodness of fit and balance between model complexity and explanatory power. The AIC and BIC are statistical measures that consider both the likelihood of the model and the number of parameters, helping to identify the most parsimonious and effective model. The AFT model with the lowest AIC and BIC values was considered the best-fitting model, indicating its superiority in explaining the observed data while penalizing for model complexity. This rigorous model comparison process aimed to identify the most appropriate AFT model for understanding the duration times of women age at first childbirth based on the selected covariates. Based on table 4.5 log-logistic AFT model is most apropriate model.

Table 4. 5 AIC and BIC of AFT model

Distribution	DF	AIC	BIC
Exponential	17	15407.04	15527.6
Weibull	18	4745.988	4873.646
Lognormal	18	3300.232	3427.89
Log-logistic	**18**	**3230.841**	**3358.5**

Table 4. 6 ***multivariable log logistic AFT models***

Covariate	Coef.	Std. Err.	Z	P>z	[95% Conf. Interval]	
Residence						
Rural	-.029381	.0106226	-2.77	0.006	-.050201	-.008561
Education level						
Primary	.0777227	.0071904	10.81	0.000	.0636297	.0918157
Secondary	.1765601	.011261	15.68	0.000	.1544889	.1986313
Higher Education	.2980647	.0138259	21.56	0.000	.2709664	.3251631
Wealth						
Second	-.0347888	.0099194	-3.51	0.000	-.0542304	-.0153472
Middle	-.0459861	.0102982	-4.47	0.000	-.0661702	-.025802
Fourth	-.0460747	.0103438	-4.45	0.000	-.0663483	-.0258012
Highest	.0154597	.0127186	1.22	0.224	-.0094683	.0403877
Contraceptive use						
Yes	-.0274992	.0071101	-3.87	0.000	-.0414347	-.0135636
Marital status						
Yes	-.2150108	.0089535	-24.01	0.000	-.2325593	-.1974623
Sex of HH						
Male	.0418918	.0078574	5.33	0.000	.0264915	.0572921
Age of HH	.0019313	.0002313	8.35	0.000	.001478	.0023845
Religion						
Muslim	-.00976	.0072944	-1.34	0.181	-.0240569	.0045368
Catholic	-.1392125	.0330927	-4.21	0.000	-.2040731	-.0743519
Protestant	-.0701375	.0086647	-8.09	0.000	-.08712	-.0531551
Other	.0610208	.0319953	-1.91	0.056	-.1237305	.0016889
_cons	3.026257	.0181339	166.88	0.000	2.990715	3.061798

4.6 Frailty Survival Model

In this study, a comprehensive multivariable survival analysis was conducted, utilizing various baseline hazard distributions, namely exponential, Weibull, log-normal, and log-logistic, coupled with Gamma and Inverse Gaussian frailty distributions. The inclusion of covariates such as residence, level of education, wealth index, marital status, use of contraceptive methods, sex of the household head, age of the household head, and religion aimed to capture the diverse factors influencing the duration of women's age at first birth in Ethiopia. To account for unobserved heterogeneity between clusters, the study employed shared frailty models and assessed the significance of frailty parameters using likelihood ratio tests. The results indicated that the estimated frailty parameters were statistically significant at a 5% level of significance across all shared frailty models, suggesting a meaningful contribution of the frailty component in addressing unobservable district or geographical heterogeneity in the duration of women's age at first birth.

Furthermore, a detailed comparison of information criteria revealed that the estimated frailty models outperformed all AFT models, as evidenced by lower values of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) refer to Table 4.7. This reduction in AIC and BIC values underscores the appropriateness of frailty models over AFT models, emphasizing the importance of considering clustering effects to better capture the unobservable district or geographical heterogeneity in the duration of women's age at first birth in Ethiopia. The findings highlight the nuanced and complex nature of the underlying factors affecting the timing of first births and emphasize the significance of accounting for unobserved heterogeneity in survival analyses.

Table 4. 7 ***Comparison of gamma and inverse Gaussian shared frailty model with*** **different baseline** ***distribution***

	Frailty distribution			
Baseline distribution	**Gamma**		**Inverse-gaussian**	
	AIC	**BIC**	**AIC**	**BIC**
Exponential	15407.04	15527.6u	15409.04	15536.69
Weibull	4444.132	4578.882	4433.414	4568.164

Lognormal	3114.718	3249.469	3104.967	32393717
Loglogistic	3050.865	3185.615	**3039.127**	**3173.877**

Among the fitted frailty models, the log-logistic with Inverse-Gaussian shared frailty model demonstrated the most favorable fit for the duration of women's age at first birth. This conclusion is drawn from the observation that this particular model yielded the lowest values for both the AIC and BIC measuring 3039.127 and 3173.877, respectively (refer to Table 4.7). The minimized AIC and BIC values signify superior goodness of fit and model parsimony, making the log-logistic with Inverse-Gaussian shared frailty model the most appropriate choice for capturing the nuanced heterogeneity in the timing of women's age at first.

4.6.1 Multivariable Analysis of Log-logistic Gamma Shared Frailty Model

The discussion presented in Table 4.8 provides an analysis of the estimated log-logistic Inverse-Gaussian shared frailty survival model for determining the duration of women's age at first birth in Ethiopia. The covariates considered in the model include residence, level of education, wealth index, marital status, use of contraceptive methods, sex of the household head, age of the household head, and religion. The inclusion of these covariates aims to capture the diverse factors that influence the duration of age at first birth among Ethiopian women.

The estimated random effect, represented by the frailty term, was found to be significant. This suggests that there are time-to-failure differences across clusters, indicating the presence of unobservable heterogeneity in the data. The shared frailty model assumes an Inverse-Gaussian distribution for the frailty component, with a mean of 1 and a variance equal to theta (θ). In this case, the estimated variability of unobserved heterogeneity in the clusters is $\theta = 0.1059069$. The significant p-value obtained from the likelihood ratio test for the hypothesis $\theta = 0$ indicates that the inclusion of the frailty component is important in the model, as it contributes significantly to explaining the observed data.

The acceleration factor (ϕ) is used to determine the effect of covariates on the duration of age at first birth. A ϕ value greater than 1 indicates a prolongation of the event's duration, while confidence intervals of the acceleration factor excluding 1 at a significance level of 5% suggest that the corresponding covariates are significant factors in determining the survival time of age at first birth among Ethiopian women.

Form table 4.8 the covariate education level, the confidence interval for the covariate includes 1, but the p-value (P=0.000) indicates that primary education, secondary education, and higher education levels significantly prolong the time to first birth in Ethiopian women, as the corresponding acceleration factors (ϕ) are less than 1.

The place of residence, as indicated by the p-value of 0.03, show a significant difference in terms of the duration of age at first birth this indicate the residence rural are accelerate the age at first birth as compare to urban residence. The wealth index analysis revealed that the second, middle and fourth wealth index categories significantly prolonged the survival time of age at first birth among women but the highest or richest wealth index does not show significant difference with the reference categories that is the lowest wealth index.

The use of contraceptive methods was found to significantly prolong the survival time of age at first birth in Ethiopian women. Additionally, women who were married or in a union were significantly different from the reference group of those living alone or unmarried, and they accelerated the age at first birth, as indicated by the negative sign of the acceleration factor (ϕ), which suggests a decrease in the expected survival time.

Based on Religion a women belonging to being Catholic, Protestant and other are significantly different from the reference group orthodox and the survival time of age at first birth is short as compared from Orthodox and the religion Muslim does not show significant difference from religion Orthodox. Lastly, the age of the husband was found to be significant, with an acceleration factor (ϕ) less than 1. This indicates that increases age of husband by a unit, the age at first birth decreases by 0.001672 times, leading to a prolongation of the duration.

In summary, the estimated log-logistic Inverse-Gaussian shared frailty survival model in Table 4.8 provides valuable insights into the factors influencing the duration of age at first birth among Ethiopian women. The results highlight the significance of various covariates such as residence, education level, wealth index, contraceptive use, marital status, sex of household head, religion and age of the husband in determining the survival time. These findings contribute to a better understanding of the factors influencing reproductive patterns in Ethiopia and can inform policy and interventions aimed at improving maternal and child health

Table 4. 8 ***Loglogistic inverse Gaussian shared frailty model***

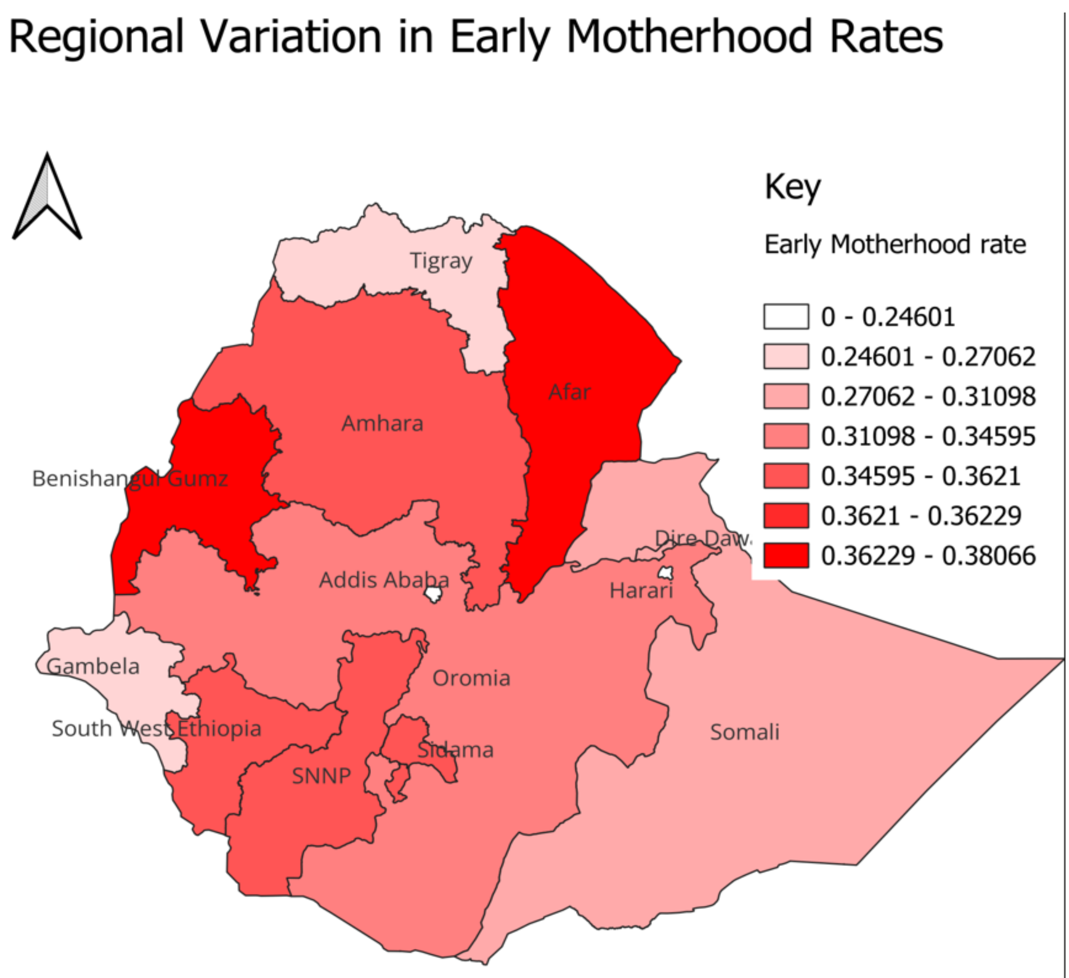
Covariate	Coef.	Std. Err.	Z	P>z	[95% CI]	
Residence						
Rural	-.0285303	.013109	-2.18	0.030	-.0542234	-.0028372
Education Level						
Primary	.0859654	.0071307	12.06	0.000	.0719894	.0999414
Secondary	.1811671	.0111376	16.27	0.000	.1593378	.2029963
Higher Education	.3000476	.013963	21.49	0.000	.2726806	.3274145
Wealth index						
Second	-.0289162	.0102035	-2.83	0.005	-.0489146	-.0089177
Middle	-.0411416	.0107807	-3.82	0.000	-.0622714	-.0200118
Fourth	-.0462436	.0111911	-4.13	0.000	-.0681776	-.0243095
Highest	-.0060885	.014203	-0.43	0.668	-.033926	.0217489
Contraceptive use						
Yes	.0276735	.0070827	-3.91	0.000	-.0415552	-.0137917
Marital status						
Yes	-.2046676	.0087883	-23.29	0.000	-.2218922	-.1874429
Sex Household head						
Male	.0389457	.0080063	4.86	0.000	.0232537	.0546377
Age HH	.001672	.0002303	7.26	0.000	.0012206	.0021233
Religion						
Muslim	-.0137906	.0087576	-1.57	0.115	-.0309552	.003374
Catholic	-.0942737	.0327589	-2.88	0.004	-.15848	-.0300674
Protestant	-.065897	.0102358	-6.44	0.000	-.0859587	-.0458353
Other	-.0873169	.0334091	-2.61	0.009	-.1527975	-.0218362
_cons	3.015694	.0200882	150.12	0.000	2.976322	3.055066
/ln_gam	-1.98968	.0127978	-155.47	0.000	-2.014764	-1.964597
/ln_the	-2.245195	.1417766	-15.84	0.000	-2.523072	-1.967318
Gamma	.1367391	.00175			.1333519	.1402124
Theta	.1059069	.0150151			.0802128	.1398314
LR test of theta =0: chibar2(01) = 193.71				Prob >= chibar2 = 0.000		

4.7 Spatial Analysis

Spatial analysis is a statistical method that is useful to specify geographical areas with high or low prevalence rates of age at first birth and variability over the region/zone. In this study, the researcher described the characteristics of the cluster, spatial distribution, spatial autocorrelation, hotspot analysis, and spatial interpolation for the age at first birth of Ethiopian women.

4.7.1 Regional Spatial Distribution for early age at first birth and age at first birth

A total 13 region employed for the both age at first birth and early age at first birth. Based on Figure 4.3, the bright red color indicates region with a high prevalence rate for age at first birth and early age at first birth whereas the white color indicates region with a low prevalence rate for the age at first birth and early age at first birth.



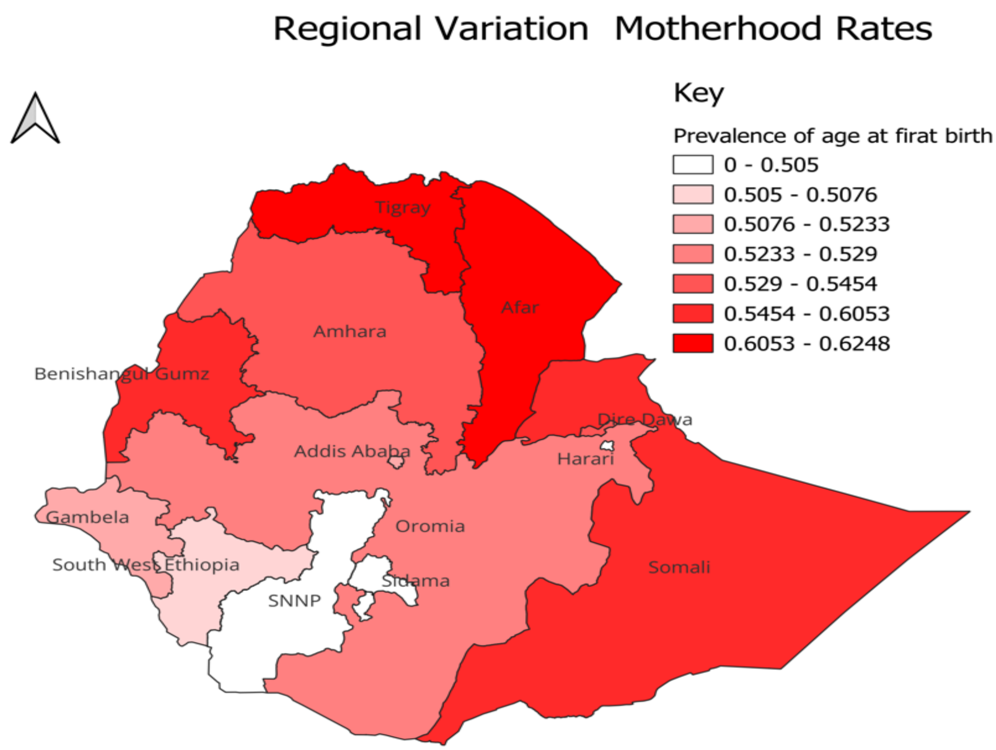


Figure 0.3 Regional Distribution for early age at first birth and age at first birth

4.7.2 Spatial Distribution for early age at first birth and age at first birth in administrative zone

A total of 305 clusters were considered for the spatial analysis for the age at first birth. 305 clusters are available in statistical analysis with non-zero latitude and longitude values. Each point on the map represents one enumeration area (cluster) with prevalence rate for age at first birth in each cluster. Based on Figure 4.4, the grey color indicates areas with a high prevalence rate for age at first birth whereas the black color indicates enumeration areas with a low prevalence rate for the age at first birth. The red color indicates areas with a high prevalence rate for early age at first birth whereas the black color indicates enumeration areas with a low prevalence rate for the early age at first birth a higher age at first birth prevalence rate was found in northeast and higher early age at first birth prevalence rate was found in northwest Ethiopia.

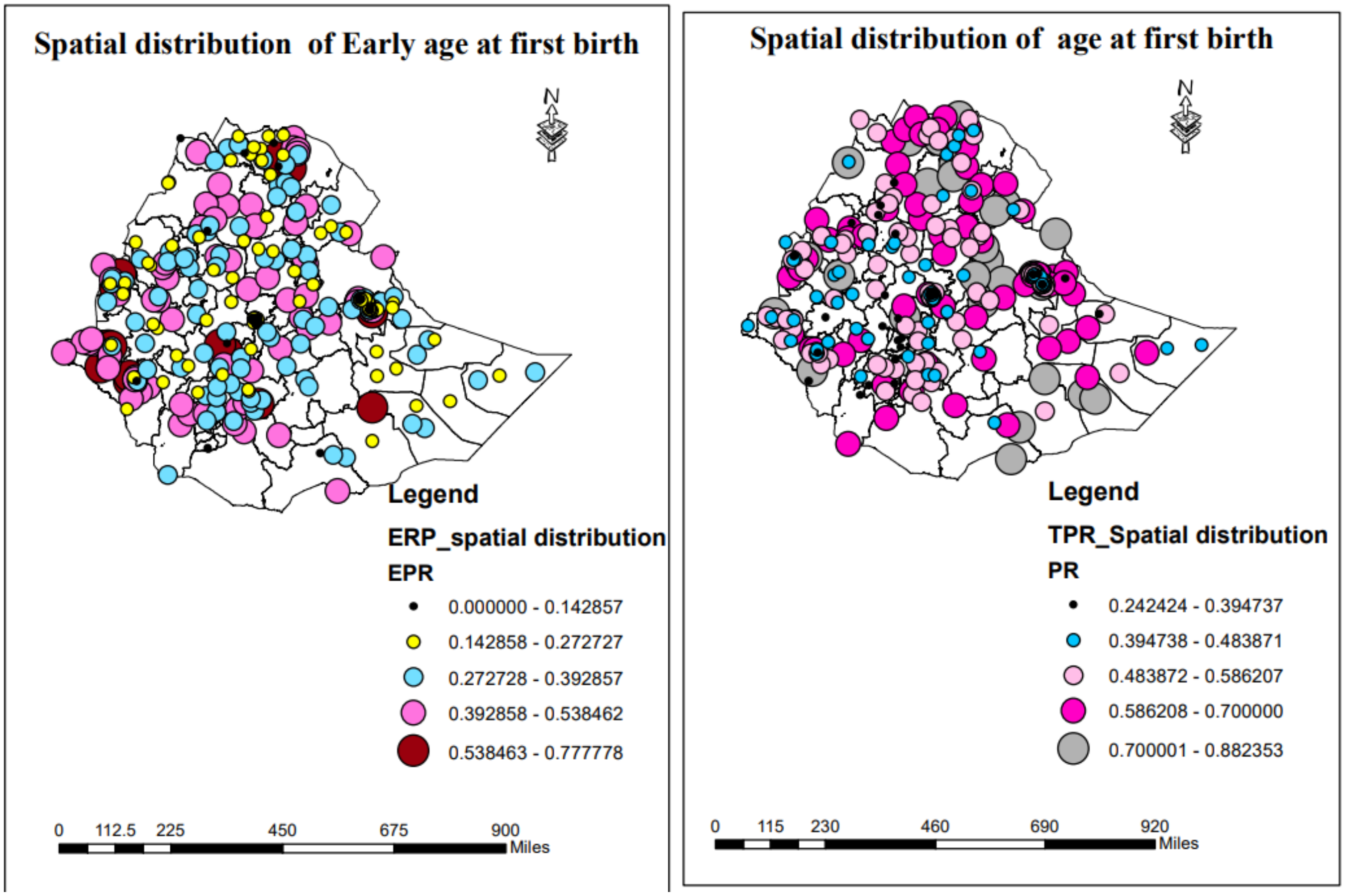
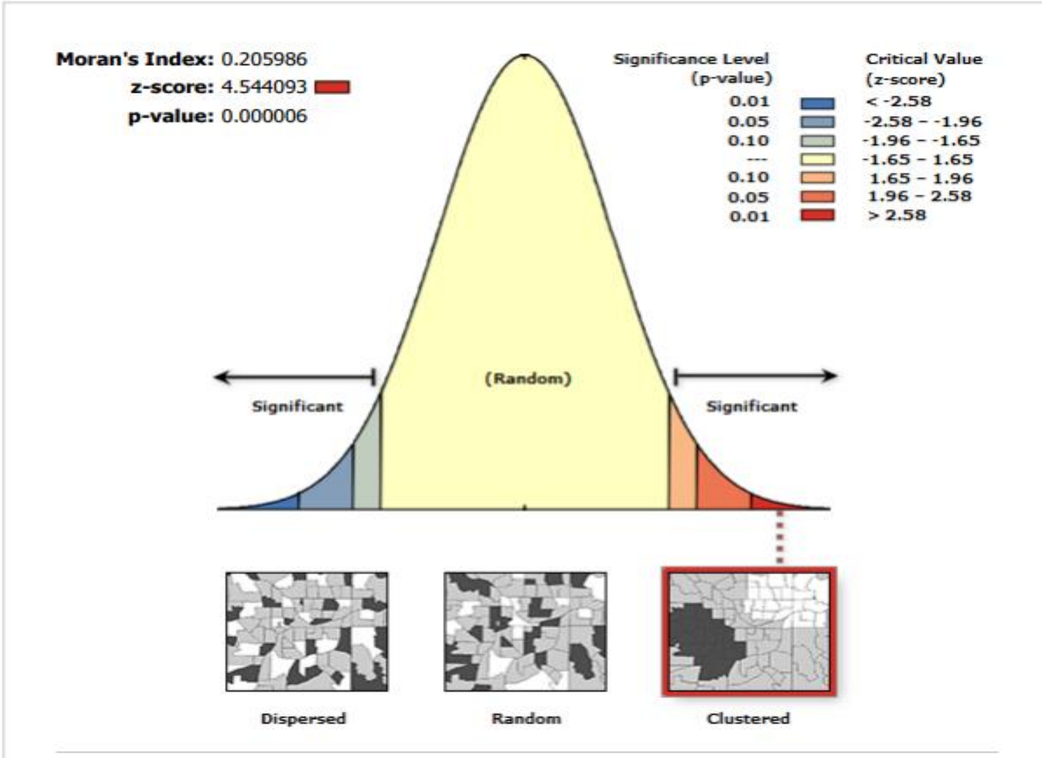


Figure 0.4 Zonal Distribution for early age at first birth and age at first birth

4.7.3 Spatial Autocorrelation for the age at first birth

The clustered patterns (on the right sides) showed that age at first birth prevalence rate were observed and the hierarchical nature of the observation. The z-value showed that there is a clustered pattern with the probability of a chance less than 1%. The bright red and blue colors to the end of the right and left tails of the bell shaped respectively indicated that there was an increased significance level (Figure 4.5). the Moran's I index measures the spatial autocorrelation, meaning that Moran's Index allows to detect the prevalence of age at first birth in a given cluster may be similar to those neighboring clusters. The estimated Moran's I index is 0.205986 and 0.661727 with lowest p-value which is less than 0.05. This implies that there is significant evidence of unexplained spatial autocorrelation or clustered pattern in the prevalence of age at first birth and early age at first birth.



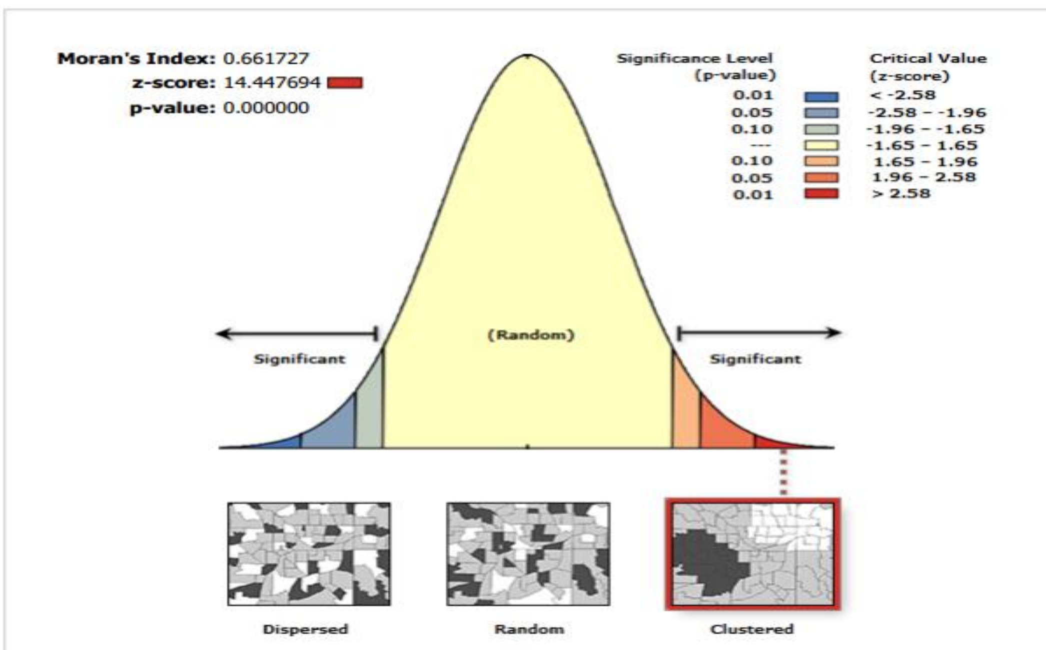


Figure 0.5Spatial Autocorrelation for the age at first birth earl age at first birth, EDHS, 2019

4.7.4 Hot Spot Analysis for the age at first birth

Hot spot analysis was performed to identify high-risk and low-risk areas for the early age at first birth and age at first birth in Ethiopia. The red and green color figure 4.6 left below indicates early age at first birth and age at first birth significant high risky (hotspot) areas and it was found in the northeast and west part of Ethiopia respectively. Whereas figure 4.6 right below the blue and red color indicates low risky (cold spot) areas for early age at first birth and age at first birth respectively

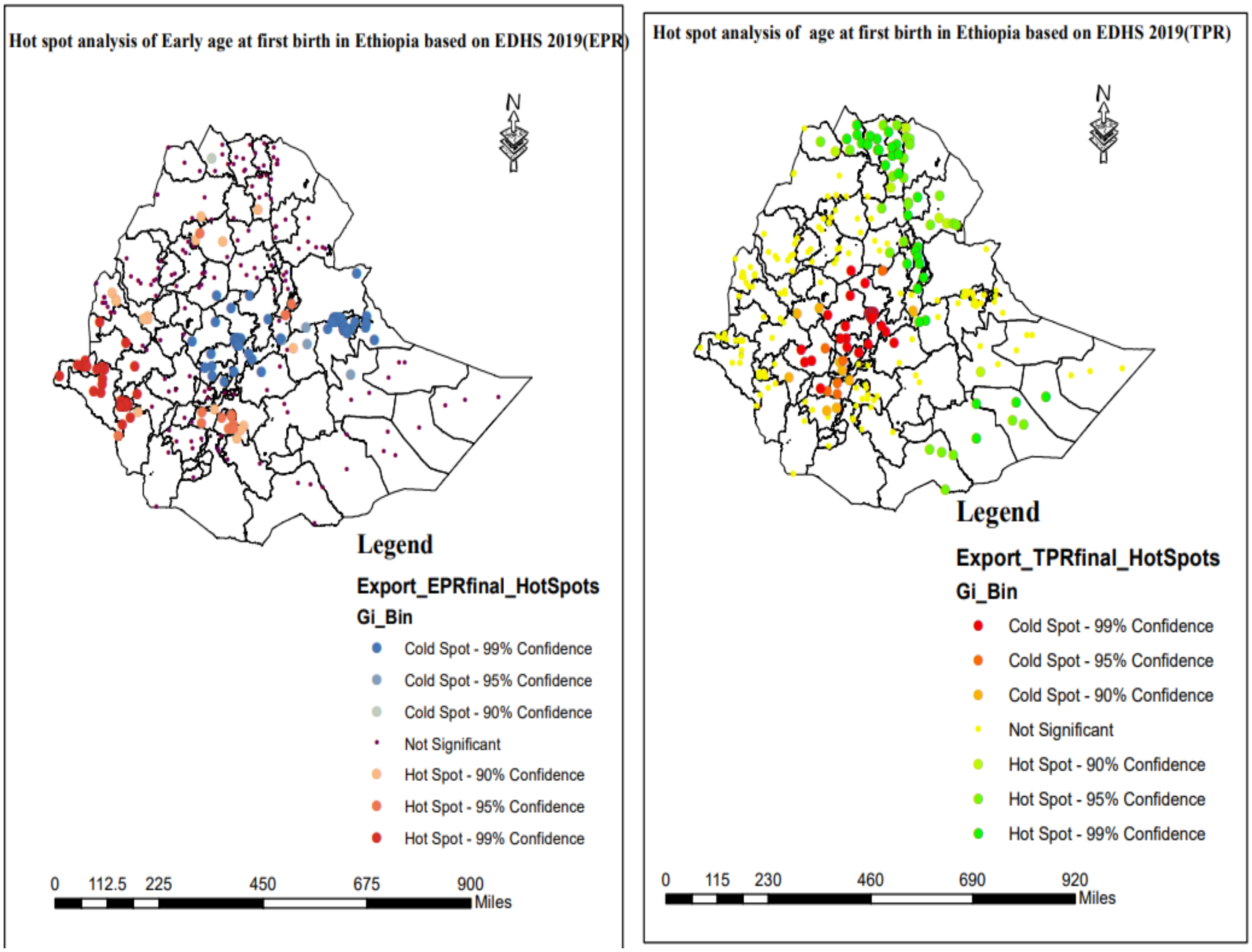


Figure 0.6Hot Spot analysis for early age at first birth and age at first birth in Ethiopia, 2019

4.7.5 Spatial Interpolation

Spatial interpolation for prediction of early age at first birth and age at first birth: the spatial kriging interpolation is an analysis predicted high and low risk regions for the early age at first birth and age at first birth. Therefore, using the ordinary kriging interpolation, the bright green color on the map figure 4.7 left below indicates the predicted highest early age at first birth and blue black color on the map indicates the predicted highest age at first birth prevalence rates in Amhara region, Addis Ababa, and central part of Tigray region whereas the blue ramp color figure 4.7 right below indicates the predicted lowest early age at first birth and age at first birth prevalence rates in Afar, Somalia, Gambela, and south east of Oromia region(Figure 4.7).

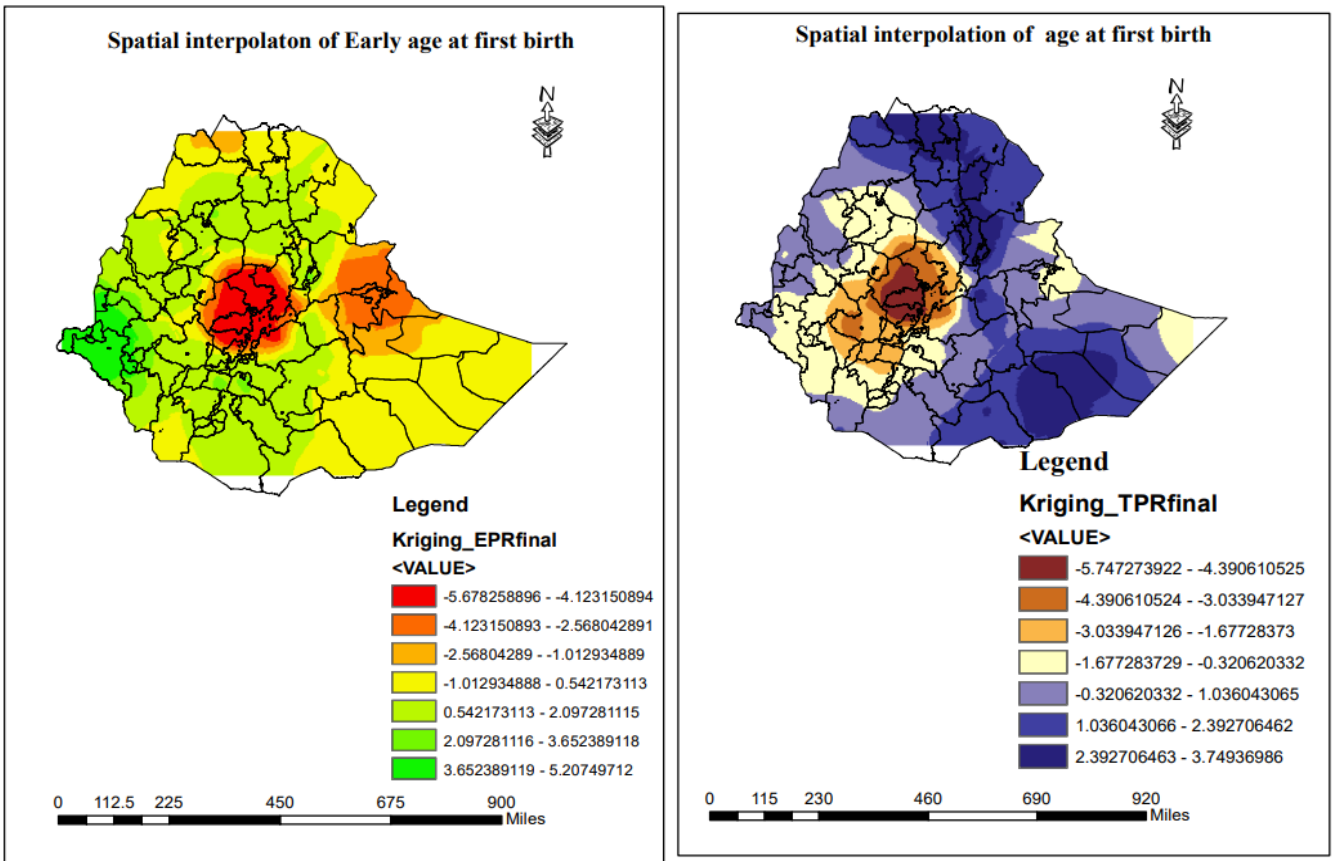


Figure 0.7 Spatial Interpolation of early age at first birth and age at first birth in Ethiopia, EDHS, 2019

4.8 Model Diagnostics

4.8.1 Checking adequacy of parametric baselines using graphical methods

After the model has been fitted, it is desirable to determine whether a fitted parametric model adequately describes the data or not. Therefore, the appropriateness of model with Weibull baseline can be graphically evaluated by plotting log (-log(S(t)) versus log(time), the Log-logistic baseline by plotting log($\hat{S}(t)$ /1−$\hat{S}(t)$) versus log(time) and the Log-normal baseline by plotting ($\phi-1[1 - S(t)]$ against log (t). If the plot is linear, the given baseline distribution is appropriate for the given dataset. Accordingly, their respective plots are given in figure 4.8 below and the plot for the Loglogistic baseline distribution make straight line better than Exponential, Weibull and Lognormal baseline distribution. This evidence also strengthens the decision made by AIC value that Log-normal baseline distribution is appropriate for the given dataset.

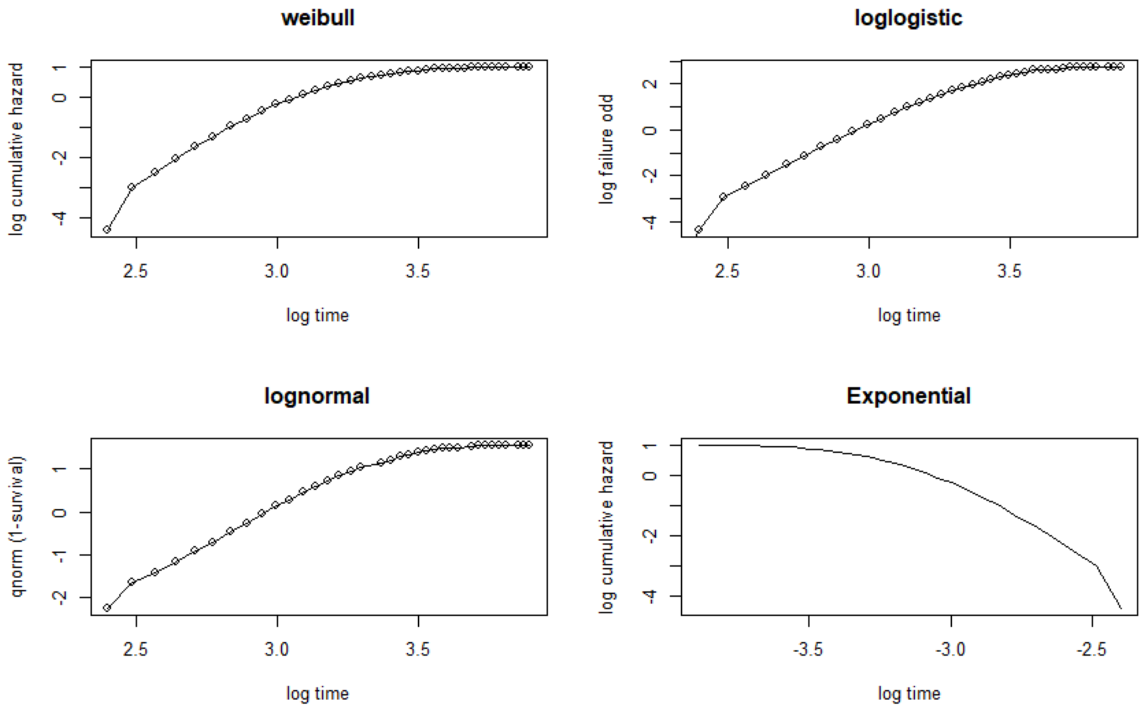


Figure 0.8 baseline hazard graph

4.8.2 Cox Snell Residual plots

The cox-Snell residual plot is one way of investigating which model is well fitted in the data.

The Cox- Snell residuals plot of the cumulative hazard function with the fitted exponential, Weibull, log-logistic, and lognormal with maximum likelihood estimation and looking log-logistic residual plot is close to the reference line in the age at first birth in Ethiopian women based on EDHS 2019 dataset that is shown in the figure below. We see that the hazard function follows the 45 degree line closely except for very large values of time. It is very common for models with censored data to have some wiggling at large values of time and it is not something which should cause much concern. Overall we would conclude that the final model fits the data very well.

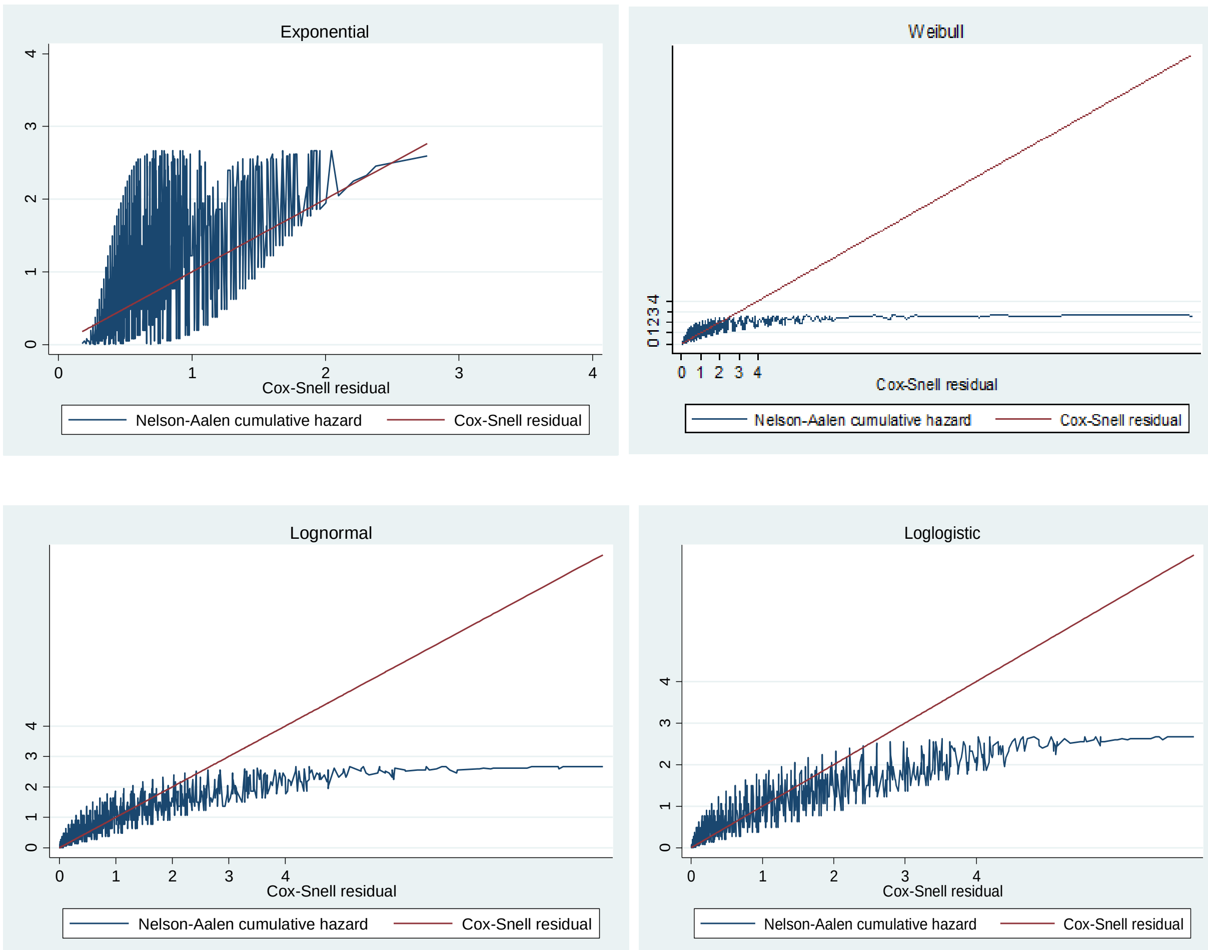


Figure 0.9 cox snell residual graph

4.9 Discussion

The objective of this research was to investigate how the timing of first birth among women of reproductive age in Ethiopia is distributed across space and identify the factors that influence it. The study utilized a parametric shared frailty analysis method to examine the determinants and used spatial analysis to understand the spatial distribution of age at first birth across the space. The results of this study enhance our understanding of the factors that impact reproductive patterns in Ethiopia and can be used to guide policy-making and interventions aimed at improving maternal and child health outcomes.

The study found that the median age at first birth in Ethiopia was 20 years, indicating a high fertility rate. This finding aligns with similar research conducted in Martial Island, Ghana, and Nigeria, where the median age at first birth ranged from 19.91 years in Ghana to 20.2 years in Martial Island (Carneiro & Ferreira Da Silva, n.d.; M. Kunnuji et al., n.d.; Zelterman, 1992) . In contrast, the findings of our study showed that the median age at first birth in Ethiopia was significantly lower when compared to the median ages (30 years or higher) observed in the majority of developed countries. (*Fertility*, 2016; *The Changing Face of Motherhood*, n.d.). One possible explanation for this disparity could be the increased opportunities for girls to pursue education during their adolescent years, as well as a growing number of women seeking economic independence through employment. These factors may contribute to delaying the timing of first births among women in developed countries, resulting in higher median ages compared to Ethiopia (*The Changing Face of Motherhood*, n.d.)

Indeed, the availability of education, awareness about the potential consequences of early childbirth, and access to contraceptives in developed countries can greatly support women in delaying their first births. These factors contribute to a better understanding of family planning options and enable women to make informed decisions about their reproductive health. By having the means to delay their first birth, women in developed countries can prioritize their education, career, and overall well-being before starting a family. (*Fertility*, 2016).

The analysis of the estimated log-logistic Inverse-Gaussian shared frailty survival model in the study offers valuable insights into the factors that influence the duration of age at first birth among Ethiopian women. Another possible reason for the similarity in lower median ages at first birth between Ethiopia and other countries could be the limited educational opportunities available to girls, especially in rural areas where the majority of the population resides. The lack of access to quality education may hinder girls from pursuing higher levels of education, which in turn can lead to earlier first births. This disparity in educational opportunities could contribute to the lower median age at first birth observed in both Ethiopia and other countries facing similar challenges. (Gupta & Mahy, 2003), which forces them to get married at an early age, to get social and financial support (Chernet et al., 2019; UNICEF., 2005). Women in developed countries have reproductive rights and the ability to make informed decisions about their reproductive health, leading to higher median ages at first birth. They have access to comprehensive healthcare services and family planning resources (*Fertility*, 2016). In contrast, our findings showed a higher median age at first birth compared to the findings observed in Bangladesh (Dewau et al., 2021a).

The variation in findings between the studies may be due to differences in study design and the specific study period, which was based on the 2011 Bangladesh Demographic and Health Survey (BDHS). The results underscore the importance of factors such as education level, wealth index, contraceptive use, marital status, and age of the husband in influencing the timing of first birth. Young housewives, often with lower education and limited awareness of the negative consequences of early childbearing, tend to be economically dependent on their spouses and have limited decision-making power, which hinders their ability to delay childbirth to older ages (*Early Marriage: Child Spouses*, n.d.). The study revealed an inverse association between women's education and early motherhood, which aligns with findings from a study conducted in Degua Tembien District, Tigray, Northern Ethiopia. This consistency in findings further supports the notion that higher levels of education are linked to a reduced likelihood of early motherhood (kidan Ayele et al., 2018), Bangladeshi (Mohammad Fazle Rabbi & Imrul Kabir, 2013)and results elsewhere (Beguy et al., 1932; Mohammad Fazle Rabbi & Imrul Kabir, 2013; *The Changing Face of Motherhood*, n.d.; Valadan et al., 2011).

Promoting secondary education for adolescent girls is crucial for effectively delaying childbirth and supporting reproductive health (Gupta & Mahy, 2003; Mohammad Fazle Rabbi & Imrul Kabir, 2013). The inverse association between educational attainment and early motherhood may be attributed to factors such as increased enrollment and retention of girls in secondary education, which reduces the likelihood of early marriage and sexual experience, while also promoting awareness of reproductive health issues (*Age at First Child: Does Education Delay Fertility Timing? The Case of Kenya by Celine Ferre :: SSRN*, n.d.; Dewau et al., 2021b; Gupta & Mahy, 2003) In contrast, women with lower levels of education often lack sufficient knowledge about the high-risk periods of pregnancy, have limited awareness of family planning methods, and may not fully comprehend the potential health costs associated with early childbearing for both mothers and children (Angeles et al., 2005; Dewau et al., 2021b). he study found that residing in regions such as Oromia, SNNP (Southern Nations, Nationalities, and Peoples), Eastern pastoralist, and western semi-pastoralist regions significantly increased the likelihood of early first birth compared to residing in urban regions (such as Addis Ababa, Dire Dawa, and Harare), even after controlling for other factors and accounting for clustering effects. This finding aligns with a previous study conducted in Ethiopia, indicating consistency in the regional patterns of early motherhood (Mekonnen et al., 2009) and Ghana (Anuwoje Ida, 2015). The observed regional differences in the likelihood of early first birth may be attributed to factors such as lower proportions of educated women and limited access to contraceptives and reproductive health services in rural regions.

These disparities in educational and healthcare resources could contribute to higher rates of early motherhood in these areas compared to urban regions (Gunasekera, 1998). Additionally, women in rural regions may have limited decision-making power concerning their reproductive health and the timing of their first birth. Factors such as traditional gender roles and cultural norms might restrict their agency in making choices related to childbirth, potentially contributing to higher rates of early motherhood in these areas.

The study found a contrasting result where a wealth index(second, middle and higher) was associated with an increased likelihood of early first birth compared to the poorest group, which may be attributed to factors such as early marriage and limited access to reproductive health services and education, particularly in rural areas (Dewau et al., 2021a) In many rural areas, wealth serves as a prerequisite for marriage, commonly known as "macha," where families pool resources for the new couple. As a result, girls from wealthier families tend to marry at an early age and become mothers during adolescence. This cultural practice contributes to the association between higher wealth and early motherhood observed in the study. The study demonstrated that women who used contraceptives had a delayed first birth compared to non-users, consistent with findings in Northeast Ethiopia (Habitu et al., 2018), East Asia (Mekonnen et al., 2009) and studies (Goli et al., 2015; Patton et al., 2009; *The Changing Face of Motherhood*, n.d.; Valadan et al., 2011; Zelterman, 1992). The observed delay in unintended pregnancies and births among sexually active women who appropriately utilized contraception may explain the association found.

However, the study findings should be interpreted with caution due to limitations such as self-report bias, potential under-reporting of certain events, limited predictor variables, and time-varying factors. Despite these limitations, the study provides valuable insights into the significant drivers of early motherhood among reproductive age women in Ethiopia, with the strength of using nationally representative data for generalizability to similar developing countries. The statistical analysis employed a suitable model for the data. In this study, residence was identified as a predictor of the timing of the first birth. Specifically, women residing in rural areas had a higher likelihood of experiencing early first births compared to those living in urban areas. This finding is consistent with similar studies conducted in Nigeria, suggesting a shared pattern across different settings (M. O. N. Kunnuji et al., 2018) and Bangladesh (Haque et al., 2011). The higher prevalence of cultural practices like early marriage and abduction in rural areas of sub-Saharan Africa (SSA) compared to urban areas could explain the increased likelihood of early first births in rural regions (Malmberg, 2008). Additionally, women in rural areas are often characterized by lower levels of education, as well as having parents with lower levels of education. This lack of education contributes to poor awareness of the consequences of early childbirth and a higher unmet need for contraceptives, further exacerbating the likelihood of early first births in rural areas (Bongaarts, 2011).

5 CHAPTER FIVE

Conclusion, Recommendation and limitation

5.1 Conclusion

The study found that the median age at first birth was 20 years, which is at the lower end of the optimum age range for first births (20-29 years). Living in rural area, no formal education, wealth index (second middle and fourth), living with together or married and being followers of the Catholic, Protestant, regions (Afar and Benshangul-gumuz) and other religion were identified as predictors of early first births. Conversely, primary, secondary and higher education levels, residing in urban regions, age of household head, sex of household being male and contraceptive use were factors associated with delaying the age at first birth.

5.2 Recommendation

Early childbirth, often resulting from lower or no education and sexual behavior, poses health risks to young mothers and their children and hinders educational and career prospects. Recommendations for policy maker include:-

- promoting education up to at least the primary education especially in rural area,
- Maximizing access to and utilization of contraceptives, and
- Further exploring factors influencing early first births among Catholic, Protestant and other religion followers.
- Region Afar and Benshangul-gumuz have high prevalence of early age at first birth so these regions need special attention to prevent early age at first birth.

Additionally, it is important to investigate factors associated with delayed first births among urban residents and those with primary, secondary and higher education levels.

5.3 Limitation of the thesis

The study findings should be interpreted in light with numerous limitations. There was the possibility, that other variables not included in the analysis significantly affect time to first birth like early age at first sex, early age at first marriage, parental education and parental economic statusThis study's primary strength is the use of nationally representative and population-based data to forecast the timing of first births among reproductive-age women in Ethiopia.

6 Reference

Aalen, O., Borgan, O., & Gjessing, H. (2008). *Survival and event history analysis: a process point of view*. Springer Science & Business Media.

Acemoglu, D., Fergusson, L., & Johnson, S. (2020). Population and Conflict. *Review of Economic Studies*, *87*(4), 1565–1604. https://doi.org/10.1093/restud/rdz042

Age at First Child: Does Education Delay Fertility Timing? The Case of Kenya by Celine Ferre :: SSRN. (n.d.). Retrieved December 26, 2023, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1344718

Alazbih, N. M., Kaya, A. H., Mengistu, M. Y., & Gelaye, K. A. (2023). Determinants of time to first marriage and birth intervals among women of child bearing age in Dabat Health and demographic surveillance system site, Northwest Ethiopia. *PLoS ONE*, *18*(2 February), 1–17. https://doi.org/10.1371/journal.pone.0281997

Angeles, G., Guilkey, D. K., & Mroz, T. A. (2005). The effects of education and family planning programs on fertility in Indonesia. *Economic Development and Cultural Change*, *54*(1), 165–201. https://doi.org/10.1086/431261

Anselin, L., & Getis, A. (1992). Spatial statistical analysis and geographic information systems. *The Annals of Regional Science*, *26*(1), 19–33. https://doi.org/10.1007/BF01581478/METRICS

Anuwoje Ida, L. (2015). The Statistical Distribution and Determinants of Mother's Age at First Birth. *American Journal of Theoretical and Applied Statistics*, *4*(2), 41. https://doi.org/10.11648/J.AJTAS.20150402.11

Beguy, D., Ndugwa, R., science, C. K.-J. of biosocial, & 2013, undefined. (1932). Entry into motherhood among adolescent girls in two informal settlements in Nairobi, Kenya. *Cambridge.OrgD Beguy, R Ndugwa, CW KabiruJournal of Biosocial Science, 2013•cambridge.Org*, *45*(6), 721–742. https://doi.org/10.1017/S0021932013000199

Bhunia, G. S., Shit, P. K., & Maiti, R. (2018). Comparison of GIS-based interpolation methods for spatial distribution of soil organic carbon (SOC). *Journal of the Saudi Society of Agricultural Sciences*, *17*(2), 114–126. https://doi.org/10.1016/J.JSSAS.2016.02.001

Bongaarts, J. (2008). Fertility transitions in developing countries: Progress or stagnation? *Studies in Family Planning*, *39*(2), 105–110.

Bongaarts, J. (2011). Can family planning programs reduce high desired family size in sub-Saharan Africa? *International Perspectives on Sexual and Reproductive Health*, *37*(4), 209–216.

Breiman, L. (2017). *Classification and regression trees*. Routledge.

Caldwell, J. C., & Caldwell, P. (1987). The cultural context of high fertility in sub-Saharan Africa. *Population and Development Review*, 409–437.

Carneiro, J., & Ferreira Da Silva, J. (n.d.). *Determinants of Export Performance: a Study of Large Brazilian Manufacturing Firms* (Issue 2). http://www.anpad.org.br/bar

Casella, G., & Berger, R. L. (2021). *Statistical inference*. Cengage Learning.

Chernet, A. G., Shebeshi, D. S., & Banbeta, A. (2019). Determinant of time-to-first birth interval after marriage among Ethiopian women. *BMC Women's Health*, *19*(1), 1–6. https://doi.org/10.1186/s12905-019-0858-3

Collett, D. (2023). *Modelling survival data in medical research*. CRC press.

Dewau, R., Mekonnen, F. A., & Seretew, W. S. (2021a). Time to first birth and its predictors among reproductive-age women in Ethiopia: inverse Weibull gamma shared frailty model. *BMC Women's Health*, *21*(1). https://doi.org/10.1186/S12905-021-01254-Z

Dewau, R., Mekonnen, F. A., & Seretew, W. S. (2021b). Time to first birth and its predictors among reproductive-age women in Ethiopia: inverse Weibull gamma shared frailty model. *BMC Women's Health*, *21*(1). https://doi.org/10.1186/S12905-021-01254-Z

Early Marriage: Child Spouses. (n.d.). Retrieved December 25, 2023, from https://www.unicef-irc.org/publications/291-early-marriage-child-spouses.html

Early Marriage A Harmful Traditional Practice A Statistical Exploration 2005 - Google Books. (n.d.). Retrieved December 25, 2023, from https://books.google.com.et/books?hl=en&lr=&id=FOn-h6oSVQwC&oi=fnd&pg=PA1&dq=The+United+Nations+Children's+Fund+(UNICEF)+2005.+Early+marriage+a+harmful+traditional+practice+a+statistical+exploration+early+marriage+a+harmful+traditional+practice+a+statistical+exploration+contents%3B+2005.&ots=SxFC33srRo&sig=1YW6OzLDsX2tJk22tBf8f1ho3bg&redir_esc=y#v=onepage&q&f=false

Ekane, D. (2013). *Fertility trends in sub Saharan Africa*.

Ethiopia, C. S. A. (2013). Population Projection of Ethiopia for All Regions at Wereda Level from 2014–2017. *Agency CS, Editor. Addis Ababa*.

Ethiopian Public Health Institute Addis Ababa. (2019). Ethiopia Mini Demographic and Health Survey. In *FEDERAL DEMOCRATIC REPUBLIC OF ETHIOPIA Ethiopia* (Issue July).

Ezra, M. (2001). Demographic responses to environmental stress in the drought-and famine-prone areas of northern Ethiopia. *International Journal of Population Geography*, *7*(4), 259–279.

Goli, S., Rammohan, A., journal, D. S.-M. and child health, & 2015, undefined. (2015). The effect of early marriages and early childbearing on women's nutritional status in India. *SpringerS Goli, A Rammohan, D SinghMaternal and Child Health Journal, 2015•Springer*, *19*(8), 1864–1880. https://doi.org/10.1007/s10995-015-1700-7

Gunasekera, H. R. (1998). Demographic and Health Survey--1993. In *Sri Lanka journal of population studies* (Vol. 1, Issue 1).

Gupta, N., & Mahy, M. (2003). Adolescent childbearing in sub-Saharan Africa: Can increased schooling alone raise ages at first birth? *Demographic Research*, *8*, 93–106. https://doi.org/10.4054/DEMRES.2003.8.4

Gurmu, E., & Etana, D. (2013). Socio-economic and Demographic Determinants of Children's Primary School Enrolment in Ethiopia. *Eastern Africa Social Science Research Review*, *29*(1), 1–30.

Habitu, Y., Yalew, A., & Bisetegn, T. (2018). *Research Article Prevalence and Factors Associated with Teenage Pregnancy, Northeast Ethiopia, 2017: A Cross-Sectional Study*. https://www.academia.edu/download/69610255/1714527.pdf

Hanagal, D. D., & Pandey, A. (2016). Gamma shared frailty model based on reversed hazard rate. *Communications in Statistics-Theory and Methods*, *45*(7), 2071–2088.

Haque, M., Islam, T. M., Tareque, M. I., & Mostofa, M. (2011). Women empowerment or autonomy: A comparative view in Bangladesh context. *Bangladesh E-Journal of Sociology*, *8*(2), 17–30.

health, S. B.-A. review of public, & 2016, undefined. (2016). Spatial data analysis. *Annualreviews.Org*, *37*, 47–60. https://doi.org/10.1146/annurev-publhealth-032315-021711

Hidayat, R., Sumarno, H., & Nugrahani, E. H. (2014). Survival analysis in modeling the birth interval of the first child in Indonesia. *Open Journal of Statistics*, *4*(03), 198.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67.

Hoogland, T., Heuvelink, G., Wetlands, M. K.-, & 2010, undefined. (2010). Mapping water-table depths over time to assess desiccation of groundwater-dependent ecosystems in the Netherlands. *Springer*, *30*(1), 137–147. https://doi.org/10.1007/s13157-009-0011-4

Iddrisu, A. K., Bukari, F. K., Opoku-Ameyaw, K., Afriyie, G. O., & Tawiah, K. (2020). Factors That Determine the Likelihood of Giving Birth to the First Child within 10 Months after Marriage. *Journal of Pregnancy*, *2020*. https://doi.org/10.1155/2020/4675907

Islam, S. (2009). Differential determinants of birth spacing since marriage to first live birth in rural Bangladesh. *Pertanika J Soc Sci Hum*, *17*(1), 1–6.

Janssen, L. D. P. (2008). *The frailty model: Springer Science+ Business Media*. LLC.

Journal of the American Geriatrics Society - Wiley Online Library. (n.d.). Retrieved September 30, 2023, from https://agsjournals.onlinelibrary.wiley.com/journal/15325415

kidan Ayele, B. G., Gebregzabher, T. G., Hailu, T. T., & Assefa, B. A. (2018). Determinants of teenage pregnancy in degua tembien district, Tigray, Northern Ethiopia: A community-based case-control study. *PLoS ONE*, *13*(7). https://doi.org/10.1371/JOURNAL.PONE.0200898

Klein, J. P., & Goel, P. (1992). *Survival analysis: state of the art*.

Klein, J. P., & Moeschberger, M. L. (2003). *Survival analysis: techniques for censored and truncated data* (Vol. 1230). Springer.

Kleinbaum, D. G., & Klein, M. (1996). *Survival analysis a self-learning text*. Springer.

Ko, D. (2000). Applied survival analysis: regression modeling of time to event data-David W. Hosmer Jr. and Stanley Lemeshow, Wiley, New York, 1999, pp. 386+ xiii price $84.95 ISBN 0-471-15410-5. *Journal of Statistical Planning and Inference*, *1*(91), 173–175.

Kulldorff, M. (2006). Tests of spatial randomness adjusted for an inhomogeneity: A general framework. *Journal of the American Statistical Association*, *101*(475), 1289–1305. https://doi.org/10.1198/016214506000000618

Kumar, G. A., & Danabalan, M. (2006). Determinants of delayed first birth. *Indian Journal of Community Medicine*, *31*(4), 272.

Kunnuji, M., … I. E.-R., & 2018, undefined. (n.d.). A survival analysis of the timing of onset of childbearing among young females in Nigeria: are predictors the same across regions? *Reproductive-Health-Journal MON Kunnuji, I Eshiet, CCP NnoromReproductive Health, 2018•reproductive-Health-Journal …*. Retrieved December 25, 2023, from https://reproductive-health-journal.biomedcentral.com/articles/10.1186/s12978-018-0623-3

Kunnuji, M. O. N., Eshiet, I., & Nnorom, C. C. P. (2018). A survival analysis of the timing of onset of childbearing among young females in Nigeria: Are predictors the same across regions? *Reproductive Health*, *15*(1). https://doi.org/10.1186/S12978-018-0623-3

Lee, S.-I. (2016). Correlation and Spatial Autocorrelation. *Encyclopedia of GIS*, 1–8. https://doi.org/10.1007/978-3-319-23519-6_1524-1

Lloyd, C. B. (2005). *Growing Up Global: The Changing Transitions to Adulthood in Developing Countries [Executive Summary]*.

Lofstedt, P., Ghilagaber, G., Shusheng, L., & Johansson, A. (2005). Changes in marriage age and first birth interval in Huaning County, Yunnan Province, PRC China. *Southeast Asian Journal of Tropical Medicine and Public Health*, *36*(5), 1329.

Logubayom, I. A., & Luguterah, A. (2013). Survival analysis of time to first birth after marriage. *Survival*, *3*(12).
Malmberg, B. (2008). *Demography and the development potential of sub-Saharan Africa*. Nordiska Afrikainstitutet.
Medicine, T. W.-S. in, & 1996, undefined. (n.d.). THE SPATIAL AUTOCORRELATION COEFFICIENT MORAN'S I UNDER HETEROSCEDASTICITY. *Wiley Online Library*. Retrieved December 26, 2023, from https://onlinelibrary.wiley.com/doi/abs/10.1002/(SICI)1097-0258(19960415)15:7/9%3C887::AID-SIM257%3E3.0.CO;2-E
Mekonnen, Y., Telake, D. S., & Wolde, E. (2009). Adolescent childbearing trends and sub-national variations in Ethiopia: a pooled analysis of data from six surveys. *BMC Pregnancy and Childbirth*, *18*(1). https://doi.org/10.1186/S12884-018-1917-8
Moellering and Tobler 1972 - Google Scholar. (n.d.). Retrieved December 26, 2023, from https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Moellering+and+Tobler+1972&btnG=
Mohammad Fazle Rabbi, A., & Imrul Kabir, M. H. M. (2013). Factors Influencing Age at First Birth of Bangladeshi Women- A Multivariate Approach. *American Journal of Public Health Research*, *1*(7), 191–195. https://doi.org/10.12691/AJPHR-1-7-8
Moran, P. A. P. (1950). A Test for the Serial Independence of Residuals. *Biometrika*, *37*(1/2), 178. https://doi.org/10.2307/2332162
Moultrie, T. A., Sayi, T. S., & Timæus, I. M. (2012). Birth intervals, postponement, and fertility decline in Africa: A new type of transition? *Population Studies*, *66*(3), 241–258.
Mukhlesur, R., Ameen, A., & Mohammad, A. (2013). Analysis of marriage to first birth interval in Bangladesh. *Int J Manag Sustain*, *2*, 208–219.
Munda, M., Rotolo, F., & Legrand, C. (2012). parfm: Parametric frailty models in R. *Journal of Statistical Software*, *51*, 1–20.
Nath, D. C., Leonetti, D. L., & Steele, M. S. (2000). Analysis of birth intervals in a non-contracepting Indian population: An evolutionary ecological approach. *Journal of Biosocial Science*, *32*(3), 343–354.
Nations, U. (2014). World fertility report 2013: Fertility at the extremes. In *Department of Economic and Social Affairs, Population Division*. United Nations New York.
Nega, W., & Woncheko, E. (2016). The determinants of birth interval in rural Ethiopia. *The Determinants of Birth Interval in Rural Ethiopia*.
Negussie, A. A. (2000). *Proximate Determinants of Birth Interval Length in Amhara Region: The Case of Fagita Lekoma Woreda, Awi-Zone*. Addis Ababa University.
Ngalinda, I. (1998). *Age at first birth, fertility, and contraception in Tanzania*.
Obeng Gyimah, S. (2003). A cohort analysis of the timing of first birth and fertility in Ghana. *Population Research and Policy Review*, *22*, 251–266.
Obiyan, M., Akinlo, A., & Ogunjuyigbe, P. (2019). Maternal socioeconomic status and fertility behaviour in Nigeria: Evidence from a cross sectional nationally representative survey. *European Scientific Journal*.
Ord, J., analysis, A. G.-G., & 1995, undefined. (1995). Local spatial autocorrelation statistics: distributional issues and an application. *Wiley Online LibraryJK Ord, A GetisGeographical Analysis, 1995•Wiley Online Library*, *27*(4), 286–306. https://doi.org/10.1111/j.1538-4632.1995.tb00912.x
Overmars, K. P., De Koning, G. H. J., & Veldkamp, A. (2003). Spatial autocorrelation in multi-

scale land use models. *Ecological Modelling*, *164*(2–3), 257–270. https://doi.org/10.1016/S0304-3800(03)00070-X

Patton, G. C., Coffey, C., Sawyer, S. M., Viner, R. M., Haller, D. M., Bose, K., Vos, T., Ferguson, J., & Mathers, C. D. (2009). Global patterns of mortality in young people: a systematic analysis of population health data. *The Lancet*, *374*(9693), 881–892. https://doi.org/10.1016/S0140-6736(09)60741-8

Rabbi, A. M. F., & Kabir, M. (2013). Factors influencing age at first birth of Bangladeshi women-A multivariate approach. *American Journal of Public Health Research*, *1*(7), 191–195.

Shayan, Z., Ayatollahi, S. M. T., Zare, N., & Moradi, F. (2014). Prognostic factors of first birth interval using the parametric survival models. *Iranian Journal of Reproductive Medicine*, *12*(2), 125.

Simeon, A. A., Khalid, Z. M., & Malaysia, B. (2014). Survival modeling of first birth interval after marriage. *Life Science Journal*, *11*(7).

Singh, G. (2007). Estimation of the duration of postpartum amenorrhea in the presence of some censored data. *The Journal of Obstetrics and Gynecology of India*, *57*(1), 53–57.

Singh, O. P., Singh, V. K., & Pathak, D. S. (2006). A Probability distribution for first birth interval. *Canadian Studies in Population [ARCHIVES]*, *33*(1), 69–81.

Stokes, C. S., & Hsieh, Y.-S. (1983). Female employment and reproductive behavior in Taiwan, 1980. *Demography*, *20*, 313–331.

The Changing Face of Motherhood. (n.d.).

Therneau, T. M., Grambsch, P. M., Therneau, T. M., & Grambsch, P. M. (2000). *The cox model*. Springer.

UNICEF. (2005). *Early marriage*. 40. https://books.google.com/books/about/Early_Marriage_A_Harmful_Traditional_Pra.html?id=FOn-h6oSVQwC

Valadan, M., Davari, T., & Sepahi, A. (2011). *Pregnancy outcomes in women of advanced maternal age*. https://www.sid.ir/EN/VEWSSID/J_pdf/1001420110204.pdf

Wasserman, L. (2021). *647 All of Statistics-A Concise Course in Statistical Inference*.

Zelterman, D. (1992). A Statistical Distribution with an Unbounded Hazard Function and Its Application to a Theory from Demography. *Biometrics*, *48*(3). https://doi.org/10.2307/2532346

Zhang, H. (2007). *Fertility Intentions and Subsequent Behaviors Among Thai Married Women: The Kanchanaburi Demographic Surveillance System, 2000-2004*. Mahidol University.

Zhenzhen, Z. (2000). Social–demographic influence on first birth interval in China, 1980–1992. *Journal of Biosocial Science*, *32*(3), 315–327.

APPENDIX

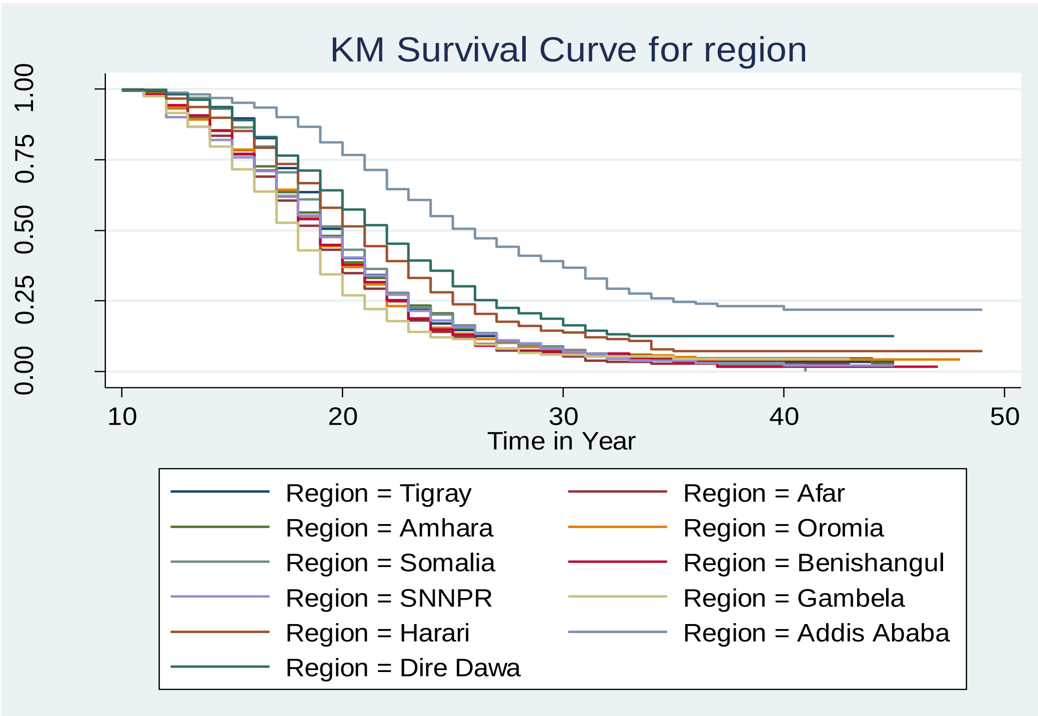


Figure 0.1 KM of Region

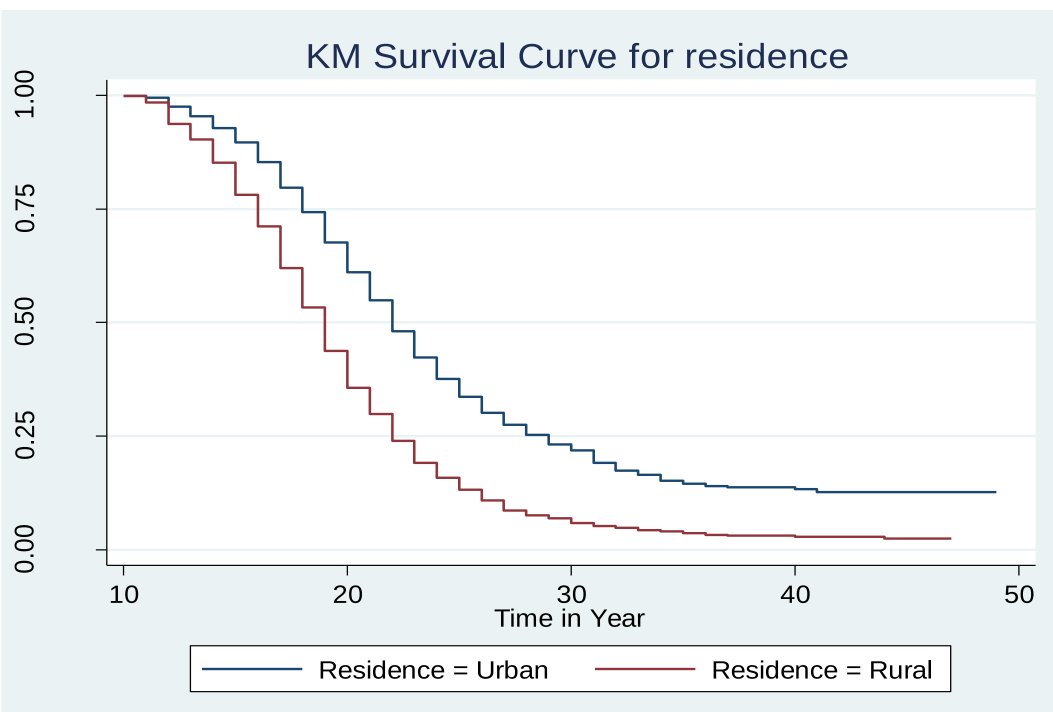


Figure 0.2 KM of Residence

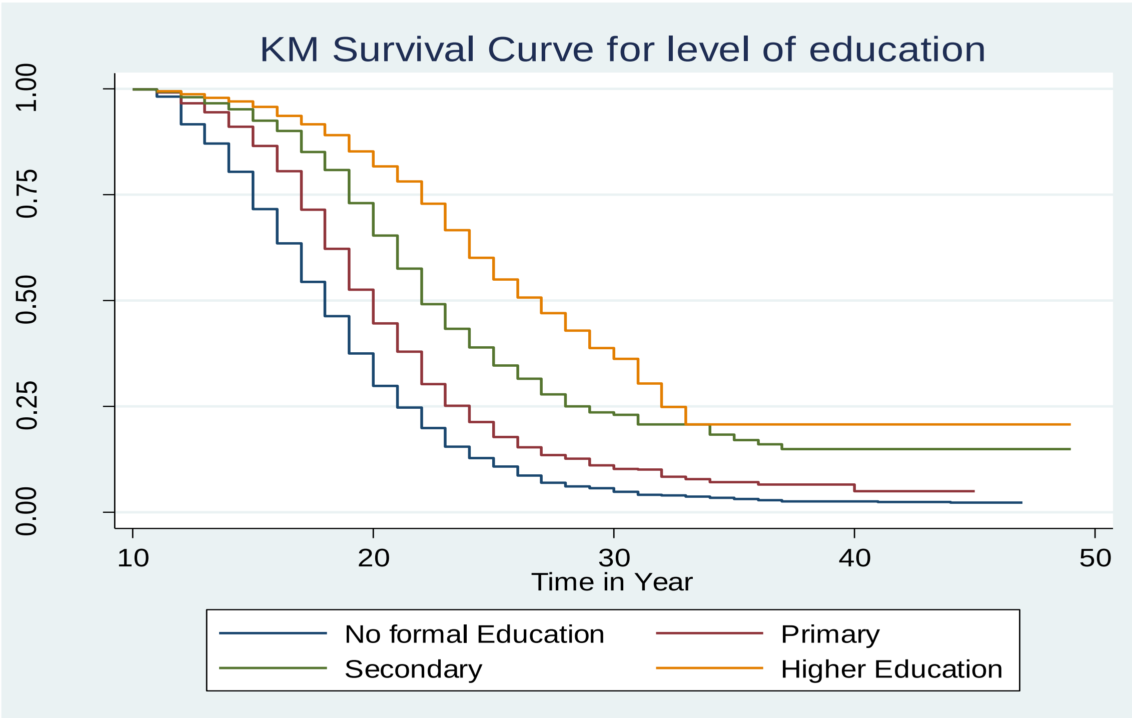


Figure 0.3 KM curve for Education level

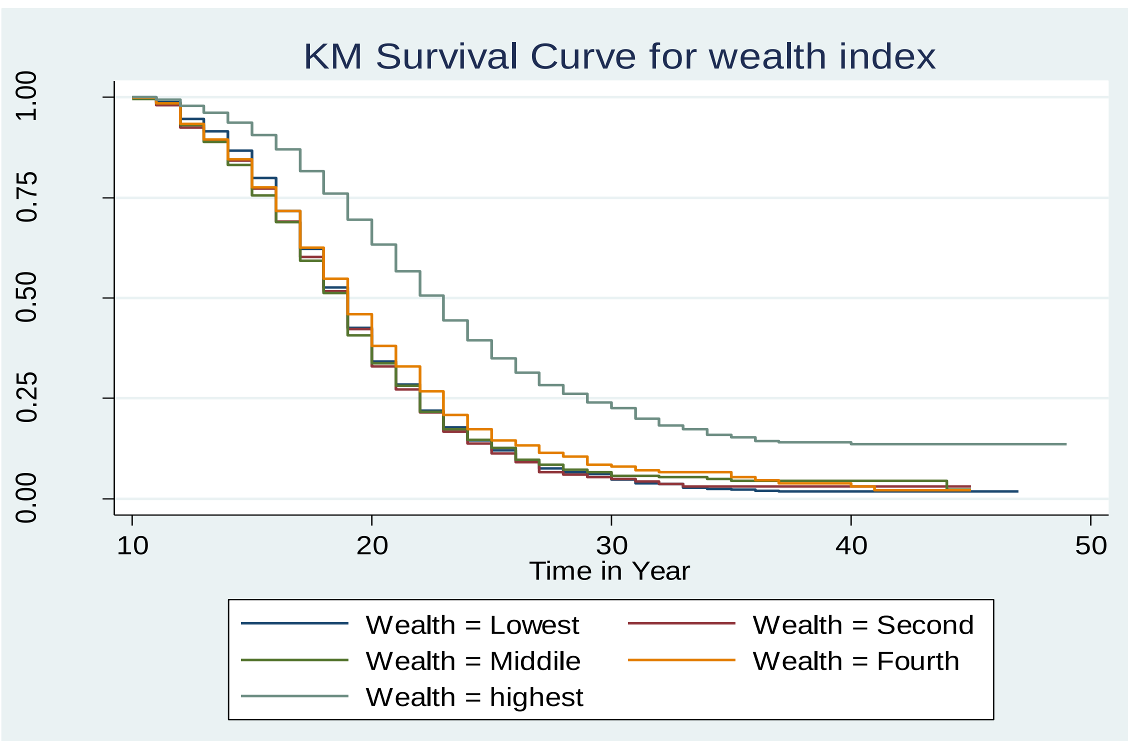


Figure 0.4 KM of wealth index

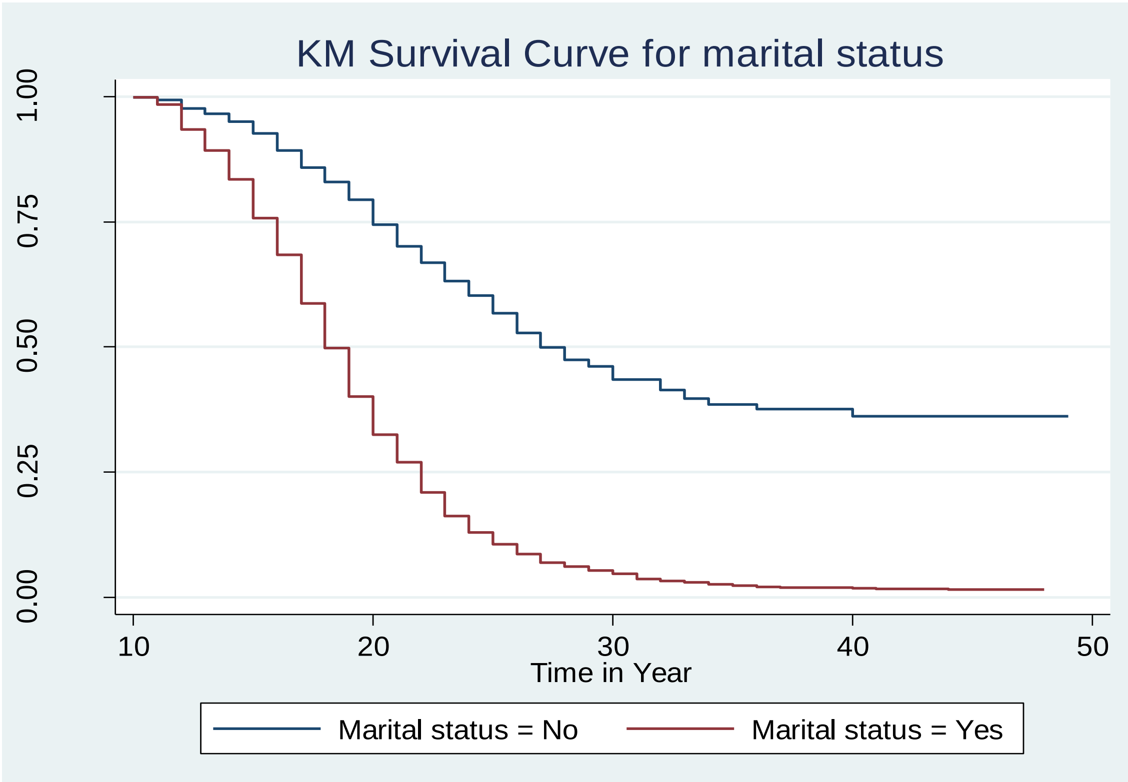


Figure 0.5 KM of Marital status

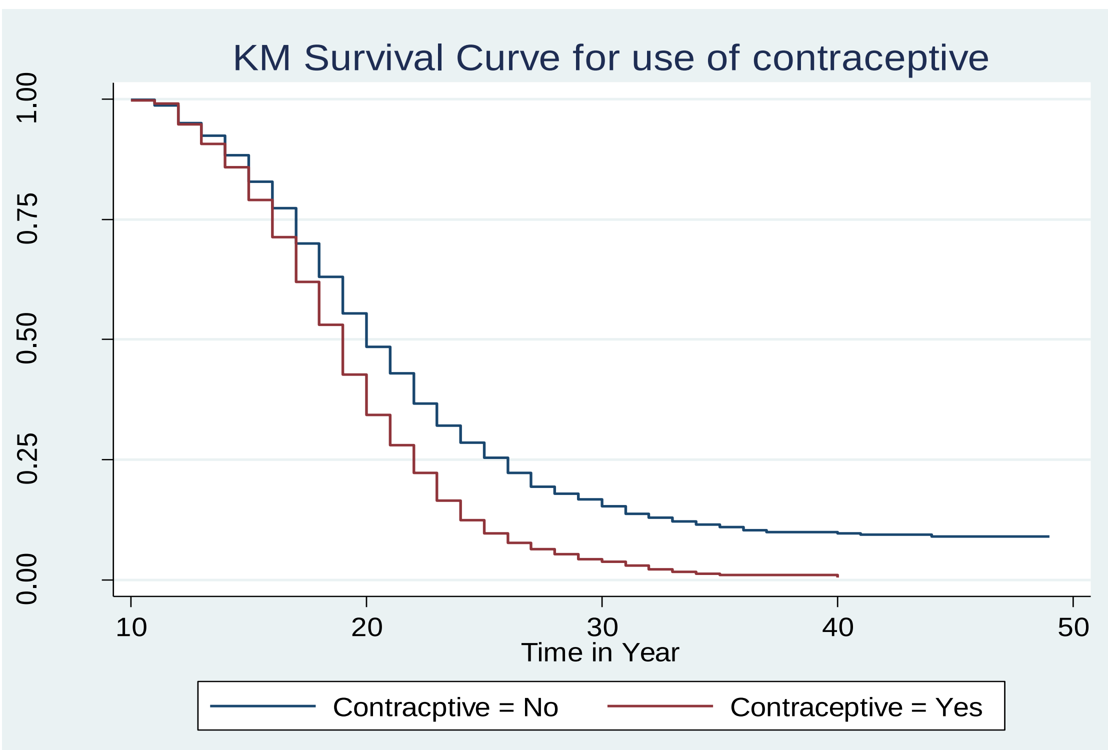


Figure 0.6 KM of Use of contraceptive method

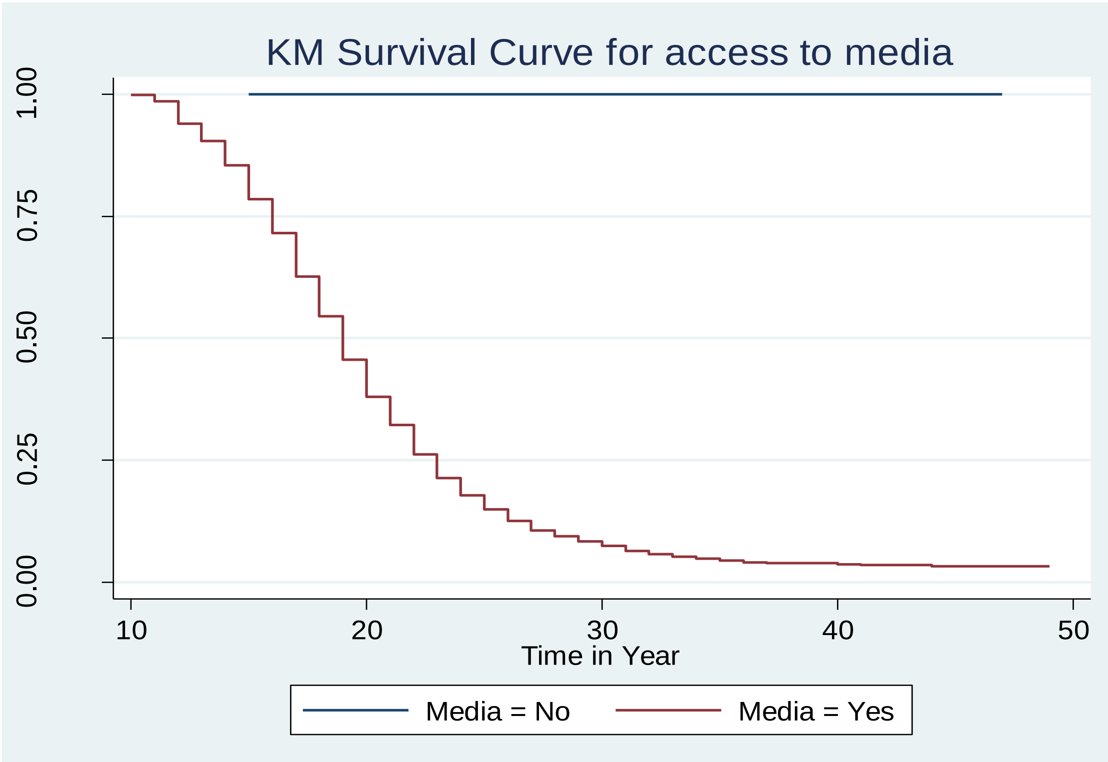


Figure 0.7 KM of women media access

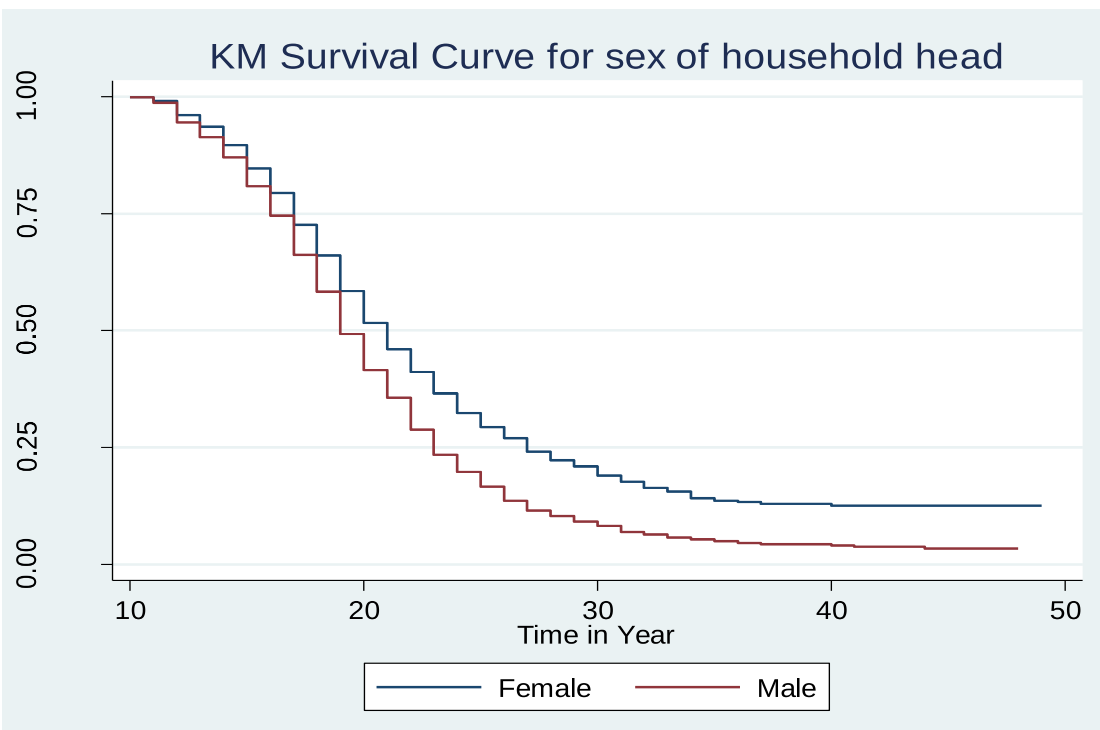


Figure 0.8 KM Sex of household head

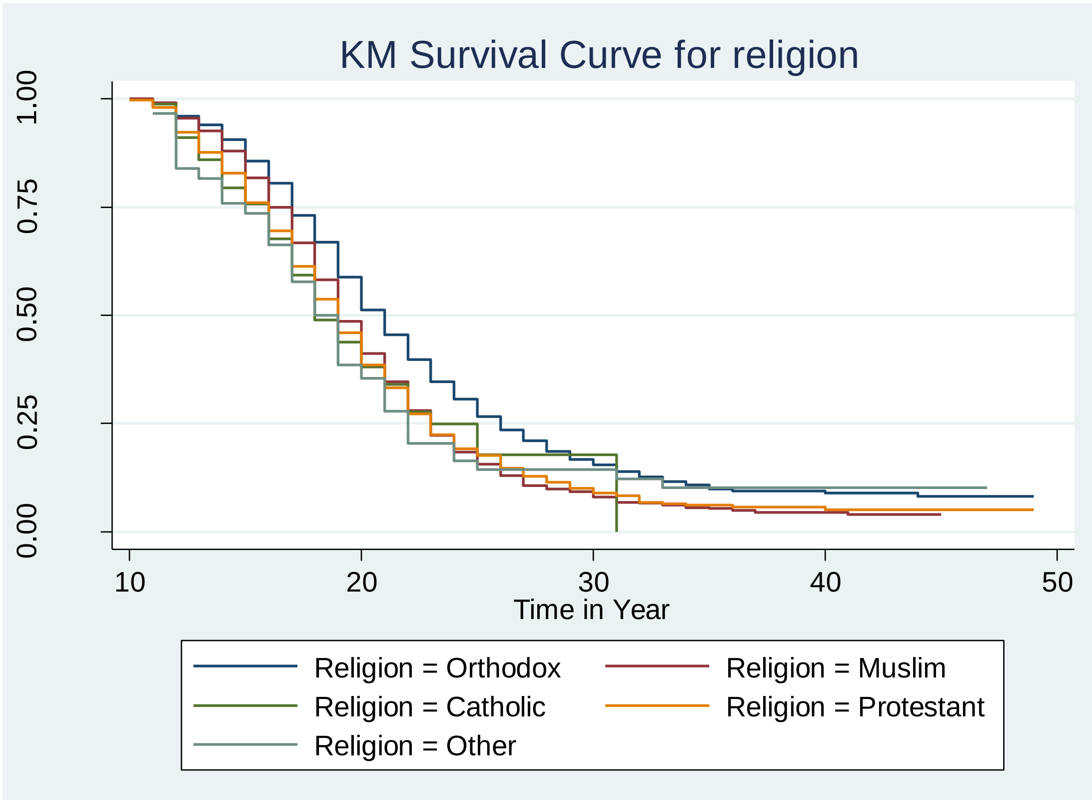


Figure 0.9 KM of Religion