BAYESIAN AND FREQUENTIST APPROACHES FOR FLEXIBLE PARAMETRIC HAZARD-BASED REGRESSION MODELS WITH GENERALIZED LOG-LOGISTIC BASELINE DISTRIBUTION: AN APPLICATION TO RIGHT-CENSORED ONCOLOGY DATA SETS

ABDISALAM HASSAN MUSE

DOCTOR OF PHILOSOPHY IN MATHEMATICS

(Statistics Option)

PAN AFRICAN UNIVERSITY

INSTITUTE FOR BASIC SCIENCES, TECHNOLOGY AND INNOVATION

(PAUSTI)

April 13, 2023

Bayesian and Frequentist Approaches for Flexible Parametric Hazard-Based Regression Models with Generalized Log-logistic Baseline Distribution: An Application to Right-Censored Oncology Data Sets

Abdisalam Hassan Muse

MS 400-0006/2019

A Thesis Submitted to Pan African University, Institute for Basic Sciences, Technology and Innovation (PAUSTI) in partial fulfillment of the requirement for the degree of Doctor of Philosophy in Mathematics (Statistics Option)

April 13, 2023

DECLARATION

This thesis is my original work, except where the acknowledgment is made in the text, and has not been submitted before to Pan African University or in any other university for a degree, diploma or other qualifications.

Signature.....Date.....13...../.....04...../.....2023

Abdisalam Hassan Muse

This thesis has been submitted with our approval as University supervisors

Signature.....Date.....13...../.....04...../.....2023

Prof. Dr. Samuel Musili Mwalili (PhD)

Department of Statistics and Actuarial Science,

Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi, Kenya.

Signature.....Date.....13...../.....04...../.....2023

Dr. Oscar Owino Ngesa (PhD)

Department of Mathematics, Statistics and Physical Sciences

Taita Taveta University, Voi, Kenya.

LIST OF PUBLICATIONS AND CONFERENCES TALKS

All but three of the chapters of this dissertation have been published in refereed journals:

List of Publications

1. **Muse, A. H.**, Mwalili, S. M., and Ngesa, O. (2021). On the log-logistic distribution and its generalizations: a survey. International Journal of Statistics and Probability, 10(3), 93 Muse et al. (2021b).

2. **Muse, A. H.,** Mwalili, S., Ngesa, O., Almalki, S. J., and Abd-Elmougod, G. A. (2021). Bayesian and classical inference for the generalized log-logistic distribution with applications to survival data. Computational Intelligence and Neuroscience, 2021. Muse et al. (2021a).

3. **Muse, A. H.**, Ngesa, O., Mwalili, S., Alshanbari, H. M., and El-Bagoury, A. A. H. (2022). A Flexible Bayesian Parametric Proportional Hazard Model: Simulation and Applications to Right-Censored Healthcare Data. Journal of Healthcare Engineering, 2022. (Muse et al., 2022g)

4. **Muse, A. H.**, Mwalili, S., Ngesa, O., Alshanbari, H. M., Khosa, S. K., and Hussam, E. (2022). Bayesian and frequentist approach for the generalized log-logistic accelerated failure time model with applications to larynx-cancer patients. Alexandria Engineering Journal, 61(10), 7953-7978. (Muse et al., 2022a).

5. **Muse, A. H.**, Chesneau, C., Ngesa, O., & Mwalili, S. (2022). Flexible Para-

metric Accelerated Hazard Model: Simulation and Application to Censored Lifetime Data with Crossing Survival Curves. Mathematical and Computational Applications, 27(6), 104., ((Muse et al., 2022)).

6. **Muse, A. H.**, Mwalili, S., Ngesa, O., Chesneau, C., Al-Bossly, A., & El-Morshedy, M. (2022). Bayesian and Frequentist Approaches for a Tractable Parametric General Class of Hazard-Based Regression Models: An Application to Oncology Data. Mathematics, 10(20), 3813. *Mathematics - MDPI*, ((Muse et al., 2022c)).

7. **Muse, A. H.**, Mwalili, S., Ngesa, O., Chesneau, C., Alshanbari, H. M., & El-Bagoury, A. A. H. (2022). Amoud Class for Hazard-Based and Odds-Based Regression Models: Application to Oncology Studies. Axioms, 11(11), 606.. *Axioms - MDPI*, ((Muse et al., 2022d)).

8. **Muse, A. H.**, Mwalili, S. M., and Ngesa, O. (2023). A Flexible Bayesian Parametric Proportional Odds Model with Generalized log-logistic baseline hazard. *Complexity - Hindawi and Wiley*, (**Under review**).

9. **Muse, A. H.**, Chesneau C., Mwalili, S. M., and Ngesa, O. (2023). A Tractable Bayesian Parametric Yang and Prentice Model with Generalized log-logistic baseline distribution: An Application to Right-censored Survival data with Crossover Survival Curves. *Concurrency and computation practice and experience- Wiley*, (To be submitted).

10. **Muse, A. H.**, Mwalili, S. , and Ngesa, O. (2023). Generating Survival Data to Simulate Accelerated Failure Time and Accelerated Hazard Models: A technique to Cancer Epidemiologist. *Computational mathematics and Data Sciecne-Elsevier*, (To be submitted).

11. **Muse, A. H.**, Mwalili, S., Ngesa, O., and Chesneau C. (2023). Accelerated

Odds Model: An application to Survival data with crossing survival curves. *Statistical Papers-Wiley*, (**Under review**).

12. **Muse, A. H.**, Chesneau, C., Mwalili, S., and Ngesa, O. (2023). On a general class for odds-based regression models: An application to oncology studies. *Annals of Data Science-Springer*, (**Under review**).

13. **Muse, A. H.**, Chesneau, C., Mwalili, S., and Ngesa, O. (2023). Super Model for Survival Regression Models: An Application to Censored Data Sets *Plose One*, (To be submitted).

14. **Muse, A. H.**, Mwalili, S., Ngesa, O., Chesneau, C. (2023). TUTORIAL IN BIOSTATISTICS: Bayesian Survival Regression Models using JAGS, BUGS, and Stan *Statistics in Medicine-Wiley*, (To be submitted).

R Software

The following R packages are published and are directly related to the work in this dissertation.

i. **Muse, A.H.,**, Mwalili, S., Ngesa, O., and Mutua, K.(2022). AHSurv: An R Package for flexible parametric accelerated hazards (AH) regression models. URL: https://cran.r-project.org/web/packages/AHSurv/index.html

ii. **Muse, A.H.,**, Mwalili, S., Ngesa, O., and Chesneau, C.(2022). AmoudSurv: An R Package for tractable parametric odds-based regression models.

URL: https://cran.r-project.org/web//packages/AmoudSurv/index.html

iii. **Muse, A.H.,**, Mwalili, S.,and Ngesa, O.,(2022). AdalSurv: An R Package for flexible parametric proportional hazards (PH) and Spatial Parametric PH regression models. (To be Submitted to R Journal)

Conference Presentations

I represented the School of Mathematics at the Pan African University, Institute for Basic Sciences, Technology and Innovation (PAUSTI) in the following conferences

1. **Muse, A. H.**, Mwalili, S. M., and Ngesa, O. (2021).Bayesian Inference for the Generalized Log-logistic Accelerated Failure Time model with Applications to Right-censored Data. *2nd International Conference of Mathematics and Its Applications in Science and Engineering (ICMASE),* Salamanca, Spain, 1st - 2nd July,2021 93.

2. **Muse, A. H.**, Mwalili, S. M., and Ngesa, O. (2021). A new generalized log-logistic distribution with increasing, decreasing, unimodal and bathtub-shaped hazard rates: properties and applications, in Proceedings of the Symmetry 2021 - The 3rd International Conference on Symmetry, 8–13 August 2021, MDPI: Basel, Switzerland, doi:10.3390/Symmetry2021-10765

3. **Muse, A. H.,** Mwalili, S. M., and Ngesa, O. (2021). Bayesian and Classical Inference for the Generalized Log-logistic Accelerated Hazard Model with Applications to Censored Survival Data with the Presence of Covariates, *Conference of the Southern Africa Mathematical Sciences Association - SAMSA2021* November 22 - 24, 2021, Botswana.

4. **Muse, A. H.**, Mwalili, S., and Ngesa, O. (2022). Amoud-G Family of Lifetime Distributions: Properties, Hazard-Based Regression Models and Applications to Survival Data. *2nd International Conference of Mathematics and Its Applications in Science and Engineering (ICMASE),* 4-7 July 2022, Technical University of Civil Engineering Bucharest, Romania, 78.

DEDICATION

This work is dedicated to my parents; Hassan Muse and Mariama Iman for their hard work in shaping my future towards intellectual inclination. To my lovely wife Hawa Rawda, for support and patience and to our beloved sons; Mu'ad and Mus'ab and my brothers and sisters and everyone supported me directly or indirectly.

ACKNOWLEDGEMENTS

First, all Praise and Gratitude be to Almightly Allah the lord of the Universe and all beings. Thank You for the bounties and guidance You bestow upon me.

With profound respect, I am highly indebted to express my hearfelt gratitude to my supervisors Prof. Dr. Samuel Musili Mwalili, and Dr. Oscar Owino Ngesa, for their meticulous, unrelenting, and professional guidance offered me in the preparation and completion of this document and for their support in diverse ways in the course of the exercise. I would like to convey my thanks to my review panel Prof. Dr. Thomas Achia (external examiner), Prof. Dr. George Orwa and Dr. Anthony Wanjoya (Internal Examiners) and Dr. Jane Aduda (Correction Supervisor) for providing valuabel feedback.

I further express my appreciation to the management and staff of the institute for being there for me when it matters most. My deepest gratitude goes to the African Union Commission for the financial support, and granting me the scholarship which ensures that I can undertake my studies here at the institute, without which this study would not have been succesful.

Much appreciation goes to my family, especially my beloved and supportive wife, Hawa, who is always by my side making my life worth living. To my little boys, Muad and Mus'ab, the most beautiful chapters of my life, my Mam and Dad for their support, tolerance, and encouragement in diverse ways, my brothers Dr. Abdifatah, Hamza, Kamal and my sisters Degan, Ramla, Anisa, Badriya, Hafsa, and Ubah.

Finally, many thanks to my colleagues and classmates.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AB	Average Bias
AFT	Accelerated Failure Time
AH	Accelerated Hazard
AIC	Akaike Information Criterion
AM	Amoud Model
AO	Accelerated Odds
BCAIC	Bozdogan's Consistent Akaike Information Criterion
BIC	Bayesian Information Criterion
BXII	Burr-XII Distributino
CAIC	Consistent Akaike Information Criterion
cdf	Cumulative Distribution Function
chf	Cumulative Hazard Function
DIC	Devaince Information Criterion
EW	Exponentiated Weibull
GG	Generalized Gamma
GH	General Hazard
GLL	Generalized Log-Logistic
GO	General Odds
GS	Gibbs Sampling
HQIC	Hannan-Quinn Information Criterion
K-S	Kolmogrov-Smirnov
LL	Log-Logistic
LN	Log-Normal

LOOIC	Leave-one-out cross-validation information criterion
McMC	Markov chain Monte Carlo
M-H	Metropolis-Hastings
MLE	Maximum Likelihood Estimation
MPHBR	Multi-Parameter Hazard-Based Regression
MPSRF	Multivariate Potential Scale Reduction Factor
MSE	Mean Squared Error
pdf	Probability Density Function
PGW	Power Generalized Weibull
PH	Proportional Hazard
PO	Proportional Odds
PSRF	Potential Scale Reduction Factor
sf	Survivor Function
SPHBR	Simple Parametric Hazard-Based Regression
RMSE	Root Mean Square Error
TTT	Total Time on Test
WAIC	Watanabe Information Criterion

ABSTRACT

In general, parametric hazard-based regression models can be motivated by allowing baseline distribution parameters to be affected by covariates. Furthermore, it is common practice to link covariates to a single parameter of interest; this method will be called a single parameter hazard-based regression (SPHBR) models. The role of the additional (covariate independent) parameters in these SPHBR models is frequently little more than to provide the model with enough generality to adjust to data. A more flexible technique is to regress these additional distributional parameters on covariates as well; this is referred to as multi-parameter hazard-based regression (MPHBR) models. The development of MPHBR models in the context of Bayesian survival analysis with particular interest towards application to right-censored oncology data is the main focus of this thesis.

Chapter 2 of this thesis examines the fundamentals of survival analysis as well as the methodologies employed to achieve the study's objective; these are common and can be ignored by readers with familiarity with the field. A flexible generalized log-logistic distribution that can incorporate both monotone and non-monotone hazard rates is developed in chapter 3 and examined using both Bayesian and classical inference methods. Using the baseline distribution proposed in chapter 3, chapter 4 presents a flexible parametric proportional hazard (PH) model. The tractability of the PH model is shown, and the implications of the method are discussed, including how to interpret covariate effects (via the hazard ratio), how to perform proportionality assumption checks on regression coefficients, and how to use Bayesian model selection techniques. To show how versatile the accelerated failure time (AFT) model is, chapter 5 presents an alternative parametric hazard-based regression model to the one presented in chapter 4. The parametric hazard-based regression models proposed in chapters 4 and 5 might not be accurate if crossovers exist in the

hazard or survival functions. The accelerated hazard (AH) model, a novel flexible hazard-based regression model that can take into account survival data with crossover survival curves, is proposed in Chapter 6 of this thesis as a solution to this issue. The parameters of the proposed AH model are estimated using Bayesian and frequentist approaches. The need to enable parametric hazard-based regression models—and really any parametric survival regression model—more interpretable motivates the presentation of a general class of parametric hazard-based regression models in Chapter 7. Covariate effects on the baseline hazard are straightforward to interpret due to the proposed general class. The class also has the benefit of allowing for both proportional and time-independent effects of some covariates on baseline hazard and non-proportional and time-dependent effects of other covariates in the same model, unlike PH, AFT, or AH. For estimating the model parameters, both the Bayesian and frequentist approaches are applied. In chapter 8, the Amoud Class, a novel class of survival regression models that includes all hazard-based and odds-based models as special cases and is more flexible in modelling survival data, is introduced. The main advantage of the class is that it can provide users with a quantitative tool for selecting which of the seven often employed methods for hazard-based and odds-based regression models is more appropriate for a certain set of data. For each of the models proposed in chapters 3, 4, 5, 6, and 7, several simulations are run using a variety of parameter settings and data generation scenarios in order to evaluate the efficacy of the model's estimators. To demonstrate the adaptability of the proposed survival regression models, applications of right-censored oncology data sets are explored. Finally, in Chapter 9, we conclude the thesis with a discussion and mention of some future works.

CHAPTER 1

INTRODUCTION

1.1 Overview

Survival analysis is defined as a collection of statistical techniques used to study the time between the start of an observation and the occurrence of an event of interest in a population. Survival analysis is used in a variety of fields to examine data pertaining to the time it takes for an 'event' to occur. It should be noted that the wheels of survival analysis have been reinvented several times in various scientific disciplines, with terminology varying from one discipline to the next. In the fields of biostatistics, epidemiology, and biomedical sciences, it is commonly referred to as time-to-event analysis or life time data analysis. Duration analysis or transition analysis is used in economics; reliability analysis is used by engineers; and event history analysis is used by demographers, sociologists, and political scientists. The response variable in a time-to-event analysis is the time until the event, which is also known as failure time, survival time, or event time.

The event of interest or the outcome variable in medical research may be pain relief, disease incidence, patient death, recurrence of symptoms, relapse from remission, incubation time of certain diseases, remission duration of certain diseases in clinical trials, such as Hepatitis B, AIDS, and so on, and in industry, the failure time of certain manufactured products (Lawless, 2011; Collett, 2015; Bogaerts et al., 2017; Cox and Oakes, 2018; Cordeiro et al., 2020; Legrand, 2021) .

According to Feng et al. (2012) the following are the key requirements of time-to-event analysis:

i. An agreed scale for the measurement of time (for example: time since diagno-

sis, calendar time);

ii. An unambiguous origin for the measurement of 'time', and;

iii. A precise definition of 'response' or occurrence of the event of interest.

The major difference between the survival analysis and the other types of statistical analysis is that survival time data contain both complete and incomplete observations; which means that the time to the occurrence of the event may not be necessarily observed for all objects within the study period. In the above examples, we may not observe for all subjects the events of death (the patient survives indefinitely), disease incidence or relief from pain. The occurrence of the event within the study time leads to a complete information in the sense that the time-to-event is observed. Whereas nonoccurrence of the event during the observation period leads to an incomplete information (partial information) which means that the exact time-to-event is unknown.

This unavoidable feature in time-to-event analysis which happens when the exact survival time is partially known and unknown in the remainder is known as censoring (Collett, 2015). Survival analysis distinguishes from other branches of statistical analysis by censoring or truncation. Truncation is defined as a late entry of an individual in the study who is then followed until the event occurs, while censoring is the dropout of a participant or lost to follow-up. Both truncation and censoring are common examples of incomplete observations arising due to the non-occurrence of the event within the study period, information contains only partial information about the random variable of interest (Legrand, 2021). These two common traits of survival data analysis also give rise to the requirement for special statistical methods to properly handle time-to-event data.

1.2 Hazard-Based Regression Models

Statistical analysis is often required to prepare summary of data for prediction. One way to achieve this is to search for a theoretical model which adequately fits the observed data and identify the covariates which are significantly associated with the response. Models are central to all statistical researches. Statistical modeling is a family of probability distributions. Probability distributions make easier characterization of the variability and uncertainty prevailing in a data set by identifying the patterns of variation. Statistical probability distributions are the basis of statistical techniques in both theory and practice. They form the foundation of every parametric statistical method including inference, modeling, survival analysis, reliability analysis among others (Lawless, 2011). For example, statistical distributions not only summarize the observations into a concise mathematical form containing a few parameters but also provide means to analyze the basic structure that govern the data generating mechanism. In contrast, statistical probability distributions have been applied in the engineering sciences to model the life cycle of a machine. In the medical sciences, statistical distributions have been applied to study duration to the recurrence of cancer after surgical removal.

Survival models are based on the statistical distributions. The family of distributions may be non-parametric, semi-parametric, or parametric. To analyze survival data, different survival analysis methods such as empirical survival function, life table (actuarial method), Kaplan and Meier (K-M) or Product Limit (PL) estimator methods, long rank test, Cox Proportional Hazards (PH) model, parametric hazards models such as exponential, Weibull, Gompertz; parametric survival models such as log-logistic, log-Normal, gamma, etc have been generally used.

Censored survival data are analyzed by hazard-based regression models which require some assumptions on the way covariates affect the hazard rate function. There

are four different ways that the covariates affect the hazard rate function; namely:

i. Proportional Hazard (PH) Model

ii. Accelerated Failure Time (AFT) model

iii. Accelerated Hazard (AH) Model

iv. General Class of Hazard Structure (GH) Form

1.3 Linking the effects of Covariates and Probability Distributions

There are explanatory variables or covariates in many studies that can be related to the lifetime of patients, individuals, equipment, and so on. For example, age, gender, and so on could all be explanatory variables for the time it takes cancer patients to die. In engineering, covariates can include the type of material used, the year of manufacture, and so on. This study focuses on the methodology of flexible Bayesian modeling of covariate effects on the hazard rate function. In survival analysis, making inferences for covariate effects and baseline hazards from survival data is a standard problem.

The Cox proportional hazard (PH) model is well-known for investigating the relationship between covariates and the hazard rate function (Cox, 1972). The key assumption of the Cox proportional hazard model is that the hazard of any individual is a fixed proportion of the hazard of any other individual. This implies that the hazard ratio is determined solely by the covariates and not by time. Through a logarithmic link function and a linear predictor, the baseline hazard function is combined with hazard multipliers that are dependent on covariate values. The proportional hazard model is commonly used in time-to-event analysis.

In the context of Cox proportional hazard models, five types of assumptions are made in survival analysis.

i. The first is the semi-parametric assumption.

ii. The second assumption is the baseline hazard's shape

iii. The third is the assumption of proportional hazard.

iv. The fourth assumption is that it is not appropriate to model survival data with crossing survival curves.

v. The fifth assumption is the nature of the relationship between the values of the covariates and the hazards.

Each of these assumptions might not be appropriate and hence could be relaxed, allowing us to provide a more flexible parametric hazard-based regression models:

i. The first assumption can relaxed by using a parametric approach. Parametric survival models may lead to more precise estimates compared to semi-parametric and non-parametric methods if the distributional assumption is correct and valid (Collett, 2015).

ii. The second assumption can relaxed by developing and employing a modified flexible baseline hazard distribution that can accommodate all basic shapes of the hazard rate, including uni-modal, bathtub, increasing, decreasing, and constant hazard rates (Khosa, 2019).

iii. The third assumption can relaxed by using a non-proportional hazard models, i.e., by developing flexible parametric hazard-based regression models named accelerated failure time models (Lawless, 2011).

iv. The fourth assumption can relaxed by using a non-proportional hazard models, i.e., by developing flexible parametric accelerated hazard (AH) models that can capture survival data with crossing survival curves (Chen et al., 2003).

v. The fifth assumption can relaxed by using general class of parametric hazard-based regression models that contain most of the common hazard-based regression models (Rubio et al., 2019).

This thesis is primarily concerned with relaxing all of the assumptions mentioned above. Using the generalized log-logistic distribution as a baseline hazard, we discuss four ways for relaxing the form of the relationship between the covariate values and the hazards.

1.4 Statement of the Problem

The Cox-PH model, a semi-parametric model, has dominated the analysis of survival data over the past fifty years. While Cox's original paper discussed extensions to eliminate the PH assumption (Cox, 1972), much work has been done to increase the flexibility and versatility of hazard-based regression models using tractable functions for both the baseline hazard and the inclusion of time-dependent parameters, primarily using modified distributions, fractional polynomials or splines (Ciampi and Etezadi-Amoli, 1985; Chen and Jewell, 2001; Rubio et al., 2019; Zhang et al., 2019). This raises six issues that must be addressed.

The first issue concerns the alternative approaches that should be employed when the semi-parametric assumption of the Cox-PH model are invalid. The second concerns the kind of distribution to employ when account for multiple hazard rates within the proportional hazard assumption is satisfied. The third concerns the alternative parametric hazard-based regression model that should be used when the PH assumption is in valid. The fourth issue is how to fit crossover survival curves if the PH model fails, using other parametric hazard-based regression models. The development of a general class that can combine all three typical parametric hazard-based regression models is the topic of the fifth issue. Last but not least, the issue of how to merge hazard-based regression models with other well-liked survival re-

gression models, such as odds-based regression models including proportional odds (PO), accelerated odds (AO), and general odds (GO) models, into a single nested model, is still open.

As a result, this study investigates the solutions to these six issues and develops a flexible parametric hazard-based regression models by employing a flexible baseline known as the generalized log-logistic (GLL) distribution, which accommodates all the basic shapes of hazard rate function (constant, increasing, decreasing, unimodal, and bathtub), and various common lifetime distributions are its sub-models including (log-logistic, Burr-XII, Weibull and exponential distributions).

1.5 Motivation Statement

In recent years, several new probability distributions have been extended in the literature for describing real-life problems in survival and reliability analysis. In this regard, Khan and Khosa (2016) developed a generalized log-logistic PH model. However, the inferential procedures and the mathematical properties of the distribution have not received attention so far. Therefore, in this study we extend their work by considering the Bayesian and classical inference for the GLL distribution and extending the proposed distribution to all common types of parametric hazard-based regression models to analyze right-censored cancer survival data with the presence of covaraiates.

To summarize, the following facts compelled us to work on Bayesian parametric hazard-based regression models using GLL baseline distribution:

i. Parametric survival models may lead to more precise estimates compared to semi-parametric and non-parametric methods if the distributional assumption is correct and valid (Collett, 2015).

ii. Bayesian Inference does not depend on asymptotic approximation for statisti-

cal inference (Khan and Basharat, 2021).

iii. Due to the availability of software, Bayesian implementation for complex models is relatively easier and simpler than frequentist methods (Alvares et al., 2021).

iv. LL distribution has a similar shape to log-normal distribution, but has heavier tails, its cdf has an explicit closed-form, which is very useful for analyzing censored data in survival analysis (Bennett, 1983b).

v. The LL distribution is closed under both proportionality odds (PO) and multiplication of failure time (AFT) frameworks. It is not a PH model. However, when the log-logistic distribution is generalized, it has the appealing feature of being a member of all parametric hazard-based regression models because its failure rate function is quite flexible (Singh, 1998).

vi. The GLL distribution reflects the structure of the heavy tails and the skewness and generally shows some improvement over the log-logistic distribution and other competitive models (Khan and Khosa, 2016).

Considering the above points, we now present the main and specific objectives of this study.

1.6 Objectives

1.6.1 General objective

The main objective of this study is to develop flexible parametric hazard-based regression models with generalized log-logistic baseline distribution for right-censored survival data using both Bayesian and frequentist appproaches.

1.6.2 Specific objectives

The specific objectives of this study are:

i. To develop parametric proportional hazard regression model with generalized log-logistic baseline distribution.

ii. To develop parametric accelerated failure time regression model with generalized log-logistic baseline distribution.

iii. To develop parametric accelerated hazard regression model with generalized log-logistic baseline distribution.

iv. To extend the developed models to general class of parametric hazard-based regression models with generalized log-logistic baseline distribution.

v. To assess the performance of the estimators for the developed models parameters using simulation.

vi. To develop a general class for hazard-based and odds-based regression models.

vii. To demonstrate the application of the developed models using real-life right-censored survival data sets.

viii. To develop R packages to model the flexible parametric survival regression models considered in this study.

1.7 Significance of the study

The aim of this study, which focused on the Bayesian and frequentist approaches for the hazard-based regression models using a flexible baseline generalized log-logistic distribution, is to contribute towards enhancement of teaching and learning of Survival models in engineering, social science, economics, and medical research centers.

The findings of the study are therefore significant to the academic and industrial world. This study sheds light on the developing and generalization techniques for statistical models while paying attention to the parametric hazard-based regression models with the application of right-censored survival data sets. The findings obtained in this study makes the generalized log-logistic distribution applicable in cases where life testing experiments are faced with censored data. Additionally, the study contributes and augments the usefulness of survival data analysis.

The understanding of the relationship between covariates and lifetime distributions significantly improves the power, sensitivity, and efficiency of hypothesis tests associated with survival data sets. As a result, it is critical to develop new Bayesian parametric hazard-based regression models for the application of modified baseline distributions that can incorporate different hazard rate shapes. Hence, using a flexible baseline distribution, new parametric hazard-based and odds-based regression models were developed in this study.

1.8 The Scope of the Study

The academic scope of this study is limited to the parametric hazard-based regression models in right-censored time-to-event data with generalized log-logistic baseline hazard using frequentist and Bayesian approaches.

1.9 Research Mindmap

This section summarizes the research mind map in Figure 1.1 that connects the problems tackled in each chapter and why we followed this order.

Figure 1.1: The Research Mindmap

1.10 Thesis Outline

The development of flexible Bayesian parametric hazard-based regression models for time-to-event analysis is the subject of this thesis. The thesis is divided into chapters, each of which represents a full research paper that has been published or submitted in a peer-reviewed journal. Each paper was written as a separate article that can be read independently of the rest of the thesis but draws completely separate conclusions that are related to the overall research objectives. This thesis is divided into nine chapters, which are described in detail below.

Chapter 1: This chapter offers an introduction to the study, presenting the parametric hazard-based regression models, giving a description for the link between covariates and probability distributions, the problem statement, the study's motivation, the study's general and specific objectives, the significance of the study, the scope of this work and the research overview and contributions.

Chapter 2: This chapter presents the basic concepts and methods for Bayesian survival analysis and provides a very general introduction to survival analysis and an overview of the main probabilistic functions, such as the hazard rate, cumulative hazard rate, survival, pdf, cdf, reverse hazard rate, and odds functions. We also emphasize the characteristics for the survival data that differ it from the other statistical methods. We describe the most usual classical distributions used in the survival analysis. Then, we discussed the survival regression models that we didn't covered this thesis. Additionally, We emphasized the concept of likelihood function and we described how the censoring influence its construction. We also provide a brief description of the most widely used computing tools to account for the inference process. Then, we introduce the model selection criterions that we will applied through out this work. Finally, we present the total time on test and how it used for the description of the hazard rate shapes.

Chapter 3: This chapter presents the generalized Log-logistic Distribution. This distribution serves as an extension to some of the common lifetime distributions applied in survival analysis including: log-logistic, Burr XII, Weibull, and exponential distributions. Some of its properties including, quantile function, moments, skewness, kurtosis, residual life function, and reversed residual life functions are discussed. The model parameters are estimated using both Classical (via maximum likelihood estimation), and Bayesian (via non-informative priors) techniques. An extensive simulation study is demonstrated by using Monte Carlo technique. A real-life data set related to Bladder cancer patients is illustrated. The results show that the distribution can provide a better fit than several existing lifetime models and its sub-models (Muse et al., 2021a).

The rest of the thesis focuses on the study of parametric hazard-based regression models using GLL baseline hazard.

Chapter 4: This chapter presents a flexible Bayesian parametric proportional hazard model with the baseline distribution as GLL. The model parameters are estimated by using Bayesian approach. An extensive simulation study was carried out to assess the performance of the proposed model's estimators with finite sample sizes. The proposed inference procedure is illustrated using a right-censored cancer data set (Muse et al., 2022g).

Chapter 5: In this chapter, we presented a flexible parametric accelerated failure time model with the baseline distribution as GLL. The model parameters are estimated using both classical (via MLE) and Bayesian (assuming non-informative priors) approaches. Extensive simulation studies are conducted to assess the performance of the proposed model's estimators. The proposed AFT model is applied to a right-censored cancer data set (Muse et al., 2022a).

Chapter 6: This chapter shares the structure and methodology with chapter 6, but applies a different parametric hazard-based regression model named accelerated hazard model. This model can handle survival data with crossover survival curves

that the above two mentioned models cannot incorporate (Muse et al., 2022).

Chapter 7: This chapter is devoted to a general class of hazard-based regression models for right-censored survival data. We assume that the baseline distribution of lifetime variate is GLL. The MLE for the parameters of the proposed model are derived. Both Bayesian and Classical inferences are used to estimate the parameters of the proposed model. An extensive simulation studies are conducted to assess the performance of the proposed model's estimators. Two right-censored cancer data sets are applied to the developed model (Muse et al., 2022c).

Chapter 8: In this chapter, We investigate a universal class named Amoud Class for survival regression models with special cases that include PH, PO, AFT,AO, AH, GO and GH models, as well as combinations of these. Based on a versatile parametric distribution (generalized log-logistic) for the baseline hazard, we introduced a technique for applying these various hazard-based and odds-based regression models. The proposed model has good inferential features, and it performs well when different information criterions and frequentist likelihood ratio tests were used to select hazard-based in this study and the odds-based models. The proposed model's utility is demonstrated by an application on two-censored lifetime data sets (Muse et al., 2022d).

Chapter 9: This is the last chapter of this thesis. It presents some concluding remarks and suggests different issues for further research.

The final part of the thesis includes the usual section with all the bibliographic references mentioned in the document as well as the Appendices.

CHAPTER 2

BASIC CONCEPTS AND LITERATURE REVIEW

2.1 Introduction

This chapter reviews the basic concepts of survival analysis, including the probabilistic functions, the common features of survival data that make standard statistical inference methods inappropriate, and the classical distributions used in parametric survival models. The other topics discussed and reviewed include the basic methods in modelling censored survival data that were used to achieve the study's objectives. Methods for discrete analyses are not presented here because the primary focus is on continuous models for survival data. The framework described in this chapter is the foundation for the methodological extensions and applications presented in subsequent chapters.

2.2 Survival Analysis and Probabilistic Functions

Let $T \geq 0$ be a non-negative random variable, denoting the survival time. The distribution of survival times is characterized by any of ten probabilistic functions: the cumulative distribution function (cdf), the probability density function (pdf), the survival function (sf), the hazard rate function (hrf), the cumulative hazard function (chf), the reverse hazard function (rhf), the odds function, the derivative of the odds function, the mean residual life function (mrl), and the vitality function. In this section, ten probabilistic functions for a continuous random variable T are presented and reviewed.

2.2.1 Cumulative Distribution Function

The cumulative distribution function (cdf) also known as failure function or lifetime distribution function is the probability that the event will occur before time t and is expressed as:

$$F(t) = \Pr(T \leq t), t \geq 0 \tag{2.1}$$

F(t) can take the value from 0 to 1 and is a monotone increasing function of t. This function is also named the cumulative incidence function, since it summarises the cumulative probability of death occuring before time t (Collett, 2015).

2.2.2 Survival Function

Definition 2.2.1 *The survival function is defined as the probability that an individual survives longer than time t.*

$$S(t) = P(T \geq t), t \geq 0. \tag{2.2}$$

Survival function is also known as the reliability function, or complementary cumulative distribution function (ccdf). $S(t)$ is a monotone, left continuous, and non-increasing function, with $S(0) = 1$, and $\lim_{t\to\infty} S(t) = 0$. The graph of $S(t)$ as a function of t is called survival curve. Note that $S(t) = 1 - F(t)$, where F(t) is the cumulative distribution function (cdf) of T.

2.2.3 Probability Density Function

The probability density function (pdf) also known as the failure density or the density function is the derivative of the cdf and is written as:

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt} = \lim_{\Delta t\to 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t}, \tag{2.3}$$

which is the rate of increase of $1 - S(t)$, so that

$$S(t) = \int_t^\infty f(s)ds. \tag{2.4}$$

2.2.4 Hazard Rate Function

Definition 2.2.2 *The hazard rate function (hrf), denoted $h(t)$ is defined as the conditional probability of failure rate at t given that the individual has survived up to time t, that is*

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} \quad t \geq 0. \tag{2.5}$$

The hrf gives the instantaneous rate per unit time for the event to occur at t, given that the individual survives up to time t. There is clearly defined relationship between $S(t)$ and $h(t)$, which is given by the formula

$$h(t) = \frac{f(t)}{S(t)}. \tag{2.6}$$

Note that $h(t)\Delta t$ is the approximate probability of the event to occur in $[t + \Delta t]$, given survival up to time t. The hrf goes by several aliases:

i. In vital statistics and in the life sciences it is called as the age-specific death rate.

ii. In actuarial science it is called as the force of mortality or force fo decrement.

iii. In the engineering sciences it is called as the failure rate.

iv. In point process and extreme value theory it is called as the rate function or intensity function.

v. In economics its reciprocal is known as MILL'S ratio.

The hrf is also commonly known as the hazard rate or failure rate (Lawless, 2011). It is simple to demonstrate that building linkages between the pdf, sf, and cdf is possible when employing the hrf specified in Equation (2.6).

$$h(t) = \frac{f(t)}{S(t)} = \frac{\frac{d}{dt}F(t)}{S(t)} = \frac{\frac{d}{dt}(1 - S(t))}{S(t)} = \frac{-\frac{d}{dt}S(t)}{S(t)} = \frac{-S'(t)}{S(t)}. \tag{2.7}$$

Furthermore, Equation (2.7) suggests that

$$h(t) = -\frac{d}{dt}\log S(t), \tag{2.8}$$

so that

$$\log S(t) = -\int_0^t h(u)du + C. \tag{2.9}$$

The boundary condition $S(0) = 1$ implies $C = 0$, and hence

$$S(t) = \exp\left[-\int_0^t h(u)du\right], \tag{2.10}$$

so that

$$F(t) = 1 - \exp\left[-\int_0^t h(u)du\right]. \tag{2.11}$$

Combining Equations (2.7) and (2.10), we get

$$f(t) = h(t)S(t) = h(t)\exp\left[-\int_0^t h(u)du\right] \quad t \geq 0. \tag{2.12}$$

In another way, differentiating Equation (2.11) yields $f(t)$ in terms of $h(t)$:

$$\begin{aligned} f(t) = \frac{dF(t)}{dx} &= \frac{d\left[1 - \exp\left(-\int_0^t h(u)du\right)\right]}{dx} \\ &= h(t)\exp\left[-\int_0^t h(u)du\right] \quad t \geq 0. \end{aligned} \tag{2.13}$$

The hrf may increase, decrease, constant, or indicate a more complicated process.

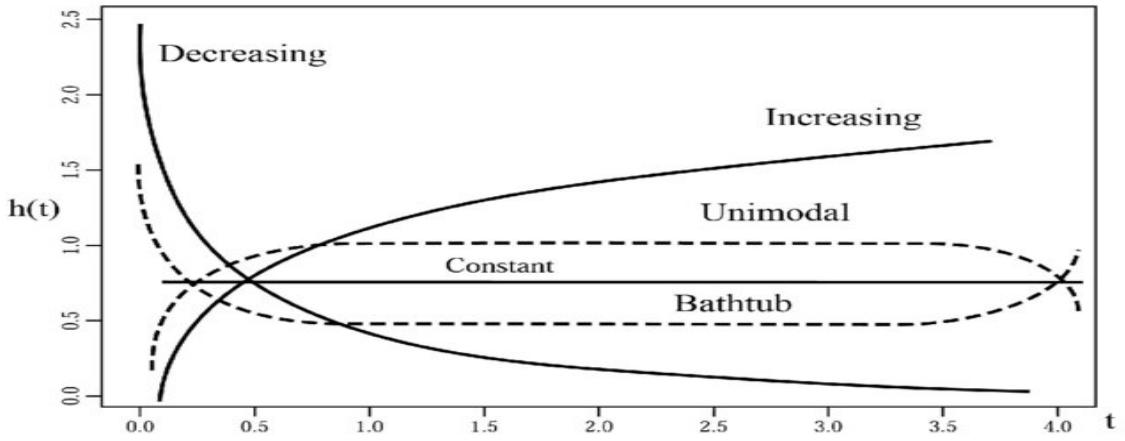

Figure 2.1: Different shapes for the baseline hazard rate function (Ramos, 2018)

For example, bathtub shape (when it has both decreasing and increasing shapes), and unimodal (or reverse bathtub) shapes. The possible different shapes for the hazard rate are summarized in Figure 2.1.

2.2.5 Cumulative Hazard Function

The cumulative hazard function, represented by H(t), is defined as

$$H(t) = \int_0^t h(u)du, \tag{2.14}$$

and satisfies three conditions:

i. $H(t) = 0$

ii. $\lim_{t\to\infty} H(t) = \infty, H(t)$

iii. $H(t)$ is increasing (=non-decreasing).

The cumulative hazard rate function also known as the integrated hazard rate function and may be interpreted as the expected number of events that occur up to a

given time (Collett, 2015). From Equations (2.10) and (2.14) we easily find

$$S(t) = \exp\{-H(t)\}, \tag{2.15}$$

and furthermore

$$H(t) = -\log\{S(t)\}. \tag{2.16}$$

The relationship in Equation (2.16) plays an important role to check adequacy of a parametric hazard-based regression models, and to formulate likelihood functions for censored time-to-event data (Khosa, 2019).

2.2.6 Reversed Hazard Function

The reversed hazard rate (also known as the retro hazard) is defined as the ratio of the corresponding pdf to the corresponding cdf. The following is the retro hazard:

$$\lambda_r(x;\boldsymbol{\theta}) = \frac{f(x;\boldsymbol{\theta})}{F(x;\boldsymbol{\theta})}. \tag{2.17}$$

Reversed hazard rate function plays an important role in the analysis of left-censored data and in the estimation of the survival function. The following equation gives us the basic relationship between hazard rate function and the reversed hazard rate function.

$$\lambda_r(x;\boldsymbol{\theta}) = \frac{h(x;\boldsymbol{\theta})S(x;\boldsymbol{\theta})}{1 - S(x;\boldsymbol{\theta})}. \tag{2.18}$$

2.2.7 Odds Function

Another function in survival analysis that has gained more attention recently is the odds function (and its associated derivative). Although the odds definition is traditional in epidemiological case-control studies, its use in survival models appeared only with the work of (Bennett, 1983a), who proposed the Proportional Odds (PO)

model class.

The odds function express, for a fixed time t, how much an event of interest is more likely to occur than not to occur (Panaro, 2020). Thus, denoting the odds function by $R(x;\theta)$ its mathematical expression is defined as follows:

$$R(x;\boldsymbol{\theta}) = \frac{F(x;\boldsymbol{\theta})}{S(x;\boldsymbol{\theta})} = \frac{1 - exp[-H(x;\theta)]}{exp[-H(t;\theta)]} = exp[H(t;\theta)] - 1. \tag{2.19}$$

2.2.8 Derivative of the Odds Fnction

The derivative of the odds function, denoted by $r(x;\theta)$ is defined as follows:

$$r(x;\boldsymbol{\theta}) = \frac{d}{dx}[R(x;\theta)] = \frac{d}{dx}[exp[H(x;\theta)] - 1] = \frac{h(x;\theta)}{S(t;\theta)} = \frac{f(x;\theta)}{[S(x;\theta)^2]}. \tag{2.20}$$

2.2.9 Mean Residual Life Function

The mean residual life (mrl) function is defined as:

$$\mu(x;\boldsymbol{\theta}) = E(T - t|T > t). \tag{2.21}$$

The mrl can characterize the x's distribution in the same way as any of the aforementioned functions. To see this, for instance, the $\mu(x)$ can be expressed in $S(x)$, and vice versa, as in

$$\mu(x;\boldsymbol{\theta}) = \frac{\int_t^\infty S(u)du}{S(x)}. \tag{2.22}$$

$$S(x;\boldsymbol{\theta}) = \frac{\mu(0)}{\mu(x)} exp \int_0^x \frac{du}{\mu(u)}. \tag{2.23}$$

The other representatives of a lifetime distribution can also be expressed interms of $\mu(x)$:

$$F(x;\boldsymbol{\theta}) = 1 - \frac{\mu(0)}{\mu(x)} exp \int_0^x \frac{du}{\mu(u)}. \tag{2.24}$$

$$H(x;\boldsymbol{\theta}) = 1 - ln[\frac{\mu(0)}{\mu(x)}] + \int_0^x \frac{1}{\mu(u)}du = \int_0^x \frac{1+\mu'(u)}{\mu(u)}du. \tag{2.25}$$

$$h(x;\boldsymbol{\theta}) = \frac{1+\mu'(u)}{\mu(u)}. \tag{2.26}$$

$$f(x;\boldsymbol{\theta}) = \frac{1+\mu'(u)}{\mu^2(u)}\mu(0)exp[-\int_0^x \frac{1}{\mu(u)}du]. \tag{2.27}$$

2.2.10 Vitality Function

Let X be a non-negative random variable (r.v.) having absolutely continuous cdf and pdf. The vitality function of X is defined as follows :

$$v(x;\boldsymbol{\theta}) = E(X|X > t) = \frac{\int_t^\infty xf(x)}{1-F(x)}. \tag{2.28}$$

The vitality function has a relationship with the hrf and the mrl functions as follows:

$$h(x;\boldsymbol{\theta}) = \frac{v'(x;\boldsymbol{\theta})}{v(x;\boldsymbol{\theta}) - t'}. \tag{2.29}$$

$$m(x;\boldsymbol{\theta}) = v(x;\boldsymbol{\theta}) - t. \tag{2.30}$$

2.2.11 Fundamental Relationships between Functions

The ten fundamental probabilistic functions in survival analysis described above. Mathematically, they can all be written in terms of one another.

Hazard function:

$$h(t) = \frac{f(t)}{S(t)} \tag{2.31}$$

Cumulative hazard function:

$$\begin{gathered} H(t) = \int_0^t h(u)du = \int_0^t \frac{f(u)}{S(u)}du = \int_0^t \frac{-dS(u)}{S(u)}du = -\log\{S(t)\} \\ f(t) = h(t)S(t) = h(t)\exp\{-H(t)\} \end{gathered} \tag{2.32}$$

From Equation (2.15) we have that

$$S(t) = \exp\{-H(t)\}$$

And furthermore

$$\begin{aligned} F(t) &= 1 - \exp\{-H(t)\} \\ f(t) &= -\frac{d\exp\{-H(t)\}}{dx} \end{aligned} \tag{2.33}$$

Hence,

$$R(t) = \exp\{H(t)\} - 1. \tag{2.34}$$

The relationship in Equation (2.34) plays an important role to formulate the likelihood function for the parametric odds-based regression models in the context of censored survival data (Demarqui et al., 2022).

Using the above equations, we have the following relations between hrf, sf, cdf, pdf, odds function, and mrl on the one hand and chf on the other hand:

$$\begin{aligned} H(t) &= \int_0^t h(u)du \\ &= -\log[S(t)] \\ &= -\log[1 - F(t)] \\ &= -\log[\int_t^\infty f(u)du] \\ &= \log[R(t) + 1] \\ &= \int_0^x \frac{1 + \mu'(u)}{\mu(u)}du \end{aligned} \tag{2.35}$$

$H(t)$ plays a key role in survival and reliability analysis, because of the exponentiated formula in Equation (2.15) which says that with $H(t)$ specified, we have

$$\Pr(X > x) = e^{-H(t)}, x \geq 0. \tag{2.36}$$

2.2.12 Why hazard rate function?

When compared to the other representations of a lifetime distribution, the hrf is usually more illuminating about the underlying cause of failure. As a result, taking the hrf into account may be the dominant method for summarizing survival statistics. According to Cox and Oakes (2018); there are several reasons why considering the hrf is a good idea:

i. It may be physically enlightening to consider the immediate "risk" attaching to an individual known to be alive at age t,

ii. Through the hrf, comparisons of groups of individuals are sometimes made intensively,

iii. Hazard-based regression models are often convenient when there is censoring or there are several types of failure

iv. The comparison with an exponential distribution is particularly simple in terms of the hazard,

v. The hazard is a special form for the "single failure" system of the complete intensity function for more elaborate point processes, i.e., systems in which multiple point events can occur for each individual."

The hrf is more informative than the survival function because different survival functions can have similar shapes. while the hrf's can differ dramatically (Colosimo and Giolo, 2006). The hrf is perhaps the most popular of the ten representatives modelling and analyzing lifetime data. This is owing to its intuitive reading as the degree of failure risk associated with a unit at age x. Another reason for its popularity is because it is a specific instance of the intensity function for a non-homogeneous POISSON process. An hrf models the occurrence of only one event, namely the first

event (=failure), whereas an intensity function depicts the occurrence of a succession of events over time.

2.3 Features of Time-to-event Data

To begin, we must consider why survival data are not accessible to typical statistical approaches employed in data analysis. In general, time-to-event data has three main features:

1. Skewness
2. Truncation
3. Censoring Mechanism

2.3.1 Skewness

The first feature of survival data is that time-to-event data are not symmetrically distributed in general. A histogram created from the survival times of a collection of comparable individuals will typically be favorably skewed, with a longer "tail" to the right of the interval containing the greatest number of observations. As a result, assuming that data of this type has a normal distribution is illogical. This problem could be solved by first modifying the data to get a more symmetric distribution, such as by using logarithms. However, a more satisfying solution would be to use a different distributional model for the original data (Collett, 2015).
The skewness of time-to-event data is an important character. As a result, normal linear model theory does not work, and models such as Weibull, log-logistic, log-normal, and their modifications are frequently used.

2.3.2 Truncation

Another common character of survival data is truncation. Truncation occurs when there is a late entry of a participant in to the study (Lee and Wang, 2003). Hence, due to the truncation and censoring, standard statistical methods cannot handle an analysis of survival data.

Truncation occurs when only those individuals whose event time lies within a certain observational window (T_L, T_R) are observed. An individual whose event time is not in this interval is not observed and no information on this subject is available to the investigator. This situation contrasts to censoring where there is at least partial information on the censored individuals. Because individual event times belong to the observational window, the inference for truncated data is restricted to conditional estimation (Klein et al., 2014). This is a an issue for doing classical inference but has a simple and natural approach within the Bayesian inference (Lázaro Hervás, 2018).

2.3.2.1 Left Truncation

Left truncation occurs when T_R is infinite so the observation frame is $[T_L, \infty)$ and only those individuals are observed whose event time T is greater than the truncation time T_L, that is $T > T_L$, it is also known as delayed entry.

2.3.2.2 Right Truncation

Right truncation occurs when $T_L = 0$ and the observational frame is $(0, T_R]$, hence survival times T are only observed when $T \leq T_R$. In other words, individuals who have experienced the event are only observable.

2.3.3 Censoring

The third and main feature of time-to-event data that makes traditional statistical inference methods inapplicable is that survival times are frequently censored. The survival time of an individual is said to be censored when the information about his lifetime is incomplete.

Censoring schemes that are well known include left censoring, right censoring, interval censoring, double censoring, and middle censoring. We describe all of the basic censoring schemes in this sub-section.

2.3.3.1 Left Censoring

Left censoring occurs when the exact time the patient entered the study is unknown but the exact time of death is known. i.e., If we follow people until they become HIV positive, we may see a failure when they first test positive for the virus. However, we may not know when we were first exposed to the virus, and thus we do not know when the failure occurred. As a result, the lifetime is censored because the true lifetime, which ends at exposure, is shorter than the follow-up time, which ends when the individual tests positive.

2.3.3.2 Right Censoring

Individuals who are followed from the beginning of the study until a point where they are lost during the follow up are referred to as right censored; that is, we know the exact time of entry of a patient but do not have the availability of the exact time of death.

In practice, there are various types of right censorship. Type I censoring, Type II censoring, and independent random censoring schemes are the most common. Below is a brief description of these schemes.

i. **Type I Censoring**: occurs when the study is designed to end after a certain

period of follow-up. The number of failures in this case is random.

ii. **Type II Censoring**: occurs when the study is terminated after a predetermined number of failures. The follow-up time in this case is arbitrary.

iii. **Independent Random Censoring**: or Type III censoring occurs when an individual is removed from the study without failing, or when the individual dies for a reason other than the one studied (Colosimo and Giolo, 2006).

2.3.3.3 Interval Censoring

Interval censoring occurs when the exact times of a patient's death and entry into the study are unknown. This type of censored data tells us that the individual was alive at specific points in time, so we know that the survival time was greater than a certain value t. i.e., Time to cosmetic deterioration of breast cancer patients.

2.3.3.4 Double Censoring

Consider a survival study in which two related events, E_1 and E_2, occur at times T_1 and T_2, respectively, with $P(T_1 \leq T_2) = 1$. Assume that the inter-occurence time, $T = T_2 - T_1$, is the survival time of interest. When both T_1 and T_2 observations are interval censored, the survival time T is said to be doubly censored or doubly interval censored (Prasad and Sankaran, 2017).

2.3.3.5 Middle Censoring

Middle-censoring, as proposed by Jammalamadaka and Mangalam (2003), occurs when a data point is rendered unobservable if it falls within a random censoring interval. In such cases, some people have access to their precise lifetimes, while others are subjected to random censoring intervals to which their true lifetimes belong. To be more specific, consider T to be a random variable representing the lifetime of interest, and (U, V) to be a bivariate random variable representing the censoring

interval such that $P(U < V) = 1$. The exact lifetime T becomes unobservable under the middle-censoring setup if $T \in (U, V)$, and we only witness the censoring interval in such cases (U, V). Exact observations on T, on the other hand, will be accessible if $T \notin (U, V)$.

2.3.3.6 Informative and Non-informative Censoring

There are two other types of censoring mechanism:

i. **Informative Censoring**: occurs when the probability of censoring is related to the expected failure time.

ii. **Non-informative Censoring**: occurs when the probability of censoring depends on covariates (i.e., sex, age, etcetra.) unrelated to the event process.

2.3.4 Why Right-Censored Observations?

The most common type of censored data in practice is right censoring scheme, which is especially common in cancer statistics and oncology studies. As a result, the parametric hazard-based regression models developed in this thesis are for data that has been right-censored and under the assumption of a non-informative censoring mechanism (Gu, 2020).

In this thesis, right censoring is of type I, in which the end of the study period C_r is known and prefixed before it begins, and/or random dropout is permitted.

2.4 Common Survival Distributions

This section reviews at some of the continuous probability distributions that can be used to analyze survival data. Exponential, Gompertz, Weibull, log-normal, log-logistic, Burr-XII, and gamma continuous probability distributions are commonly used to represent survival data. These are more likely in survival analysis due to

(a) model simplicity, (b) approach flexibility, (c) capacity to satisfactorily represent data that are typically encountered in survival analysis, and (d) widely accessible statistical software programs (Khosa, 2019). Another modified distribution named exponentiated Weibull distribution is also discussed because it was used as a fundamental survival data generator for some simulation studies in the thesis as well as applied for some of the study's chapters for model comparison.

2.4.1 Exponential Distribution

Exponential distribution is a continuous probability distribution with only one unknown parameter k. It is the simplest distribution for lifetime distribution models. The distribution is not flexible enough to describe commonly encountered hazard rate shapes for survival data. The pdf, cdf, sf, hrf and chf of the exponential random variable are, respectively

Let $X \sim \text{Exponential}(k)$

$$f(t) = k \exp\{-kt\} \tag{2.37}$$

$$F(t) = 1 - \exp\{-kt\} \tag{2.38}$$

$$S(t) = \exp\{-kt\} \tag{2.39}$$

$$h(t) = k \tag{2.40}$$

$$H(t) = -\log S(t) = -\log(\exp\{-kt\} = kt \tag{2.41}$$

where $k > 0$ is the rate parameter (the scale parameter is $\alpha = k^{-1}$ is often used in place of k) and $t \geq 0$. The distribution with $k = 1$ is known as the standard exponential distribution. A short value of k shows low risk and long survival, where as a large value shows high risk and short survival.

Figure 2.2 shows several examples of the density, hazard, and survival function shapes for various rate parameter values.

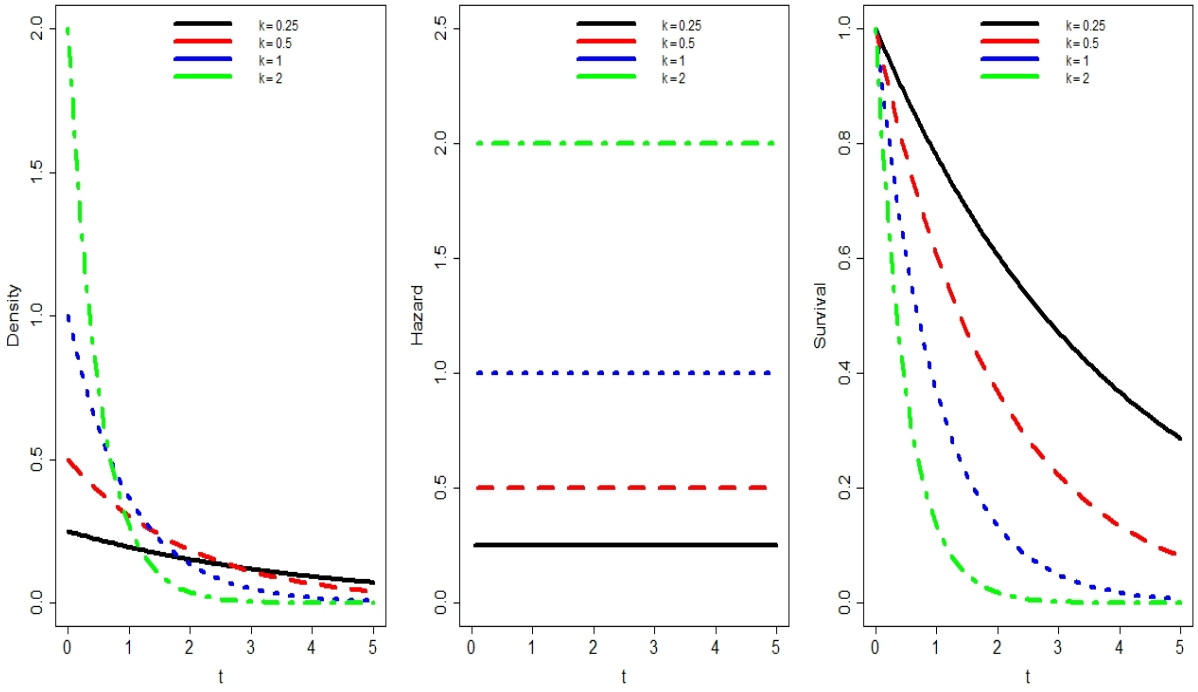

Figure 2.2: Density, hazard and survival shapes for the exponential distribution for different values of the rate parameter

2.4.2 Weibull Distribution

Weibull distribution, named after Waloddi Weibull, a Swedish mathematician and physicist, who developed it in 1951 (Weibull, 1951). It is a continuous probability distribution closed under the hazard-based regression models. This distribution is one of the most widely used lifetime distributions in survival and reliability analysis, it is also applied in many different fields of applications like the physical, social, economics, business, hydrology and weather. Several extensions have been made on the Weibull distribution and many different estimation techniques and mathematical properties have been discussed by a number of authors.

Weibull distribution is a generalization of the exponential distribution. It is a versatile distribution that can take on the characteristics of other types of continuous distributions. It has an additional parameter compared to the exponential. The additional parameter describes the shape of the hazard functions, based on the value

of the shape parameter. The pdf, cdf, sf, hrf and chf of the Weibull random variable are, respectively.

Let $X \sim$ Weibull (k, α)

$$f(t) = \alpha k(kt)^{\alpha-1} \exp\{-(kt)^{\alpha}\} \tag{2.42}$$

$$F(t) = 1 - \exp\{-(kt)^{a}\} \tag{2.43}$$

$$S(t) = \exp\{-(kt)^{\alpha}\} \tag{2.44}$$

$$h(t) = \alpha k(kt)^{\alpha-1}, \tag{2.45}$$

$$H(t) = -\log S(t) = -\log\left(\exp\{-(kt)^{a}\} = (kt)^{\alpha}\right. \tag{2.46}$$

where $\beta > 0$ is the shape parameter and $k > 0$ is the rate parameter (the scale parameter is scale $= k^{-1}$ is often used in place of k).

The hazard rate function increases when $\alpha > 1$, decreases for $\alpha < 1$, and constant for $\alpha = 1$. When $\alpha = 1$ the Weibull distribution reduces to exponential. It is worth mentioning that the Weibull distribution does not accommodate non-monotone (i.e., unimodal) hazard functions.

Figure 2.3 shows several examples of the density, hazard, and survival function shapes for various rate and shape parameter values.

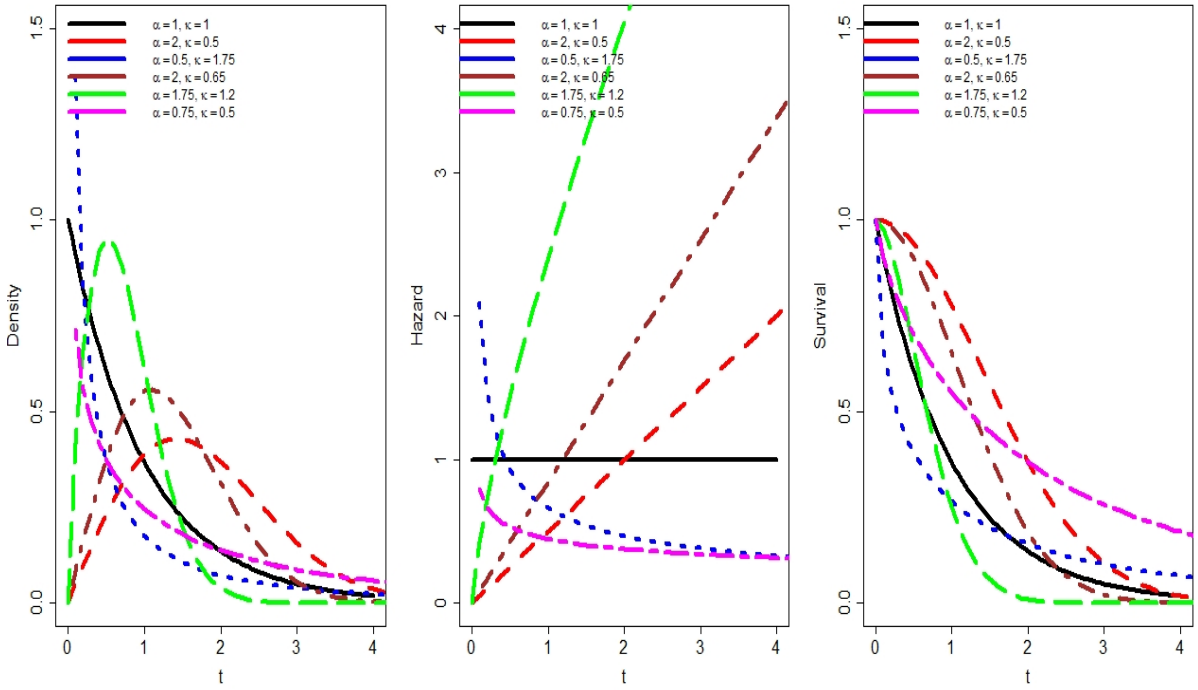

Figure 2.3: Density, hazard and survival shapes for the Weibull distribution for different values of the rate and shape parameters

Regarding the odds function and its derivative, they are expressed as:

$$R(t) = exp[H(t)] - 1 = exp(kt)^{\alpha} - 1. \tag{2.47}$$

$$r(t) = h(t)exp[H(t)] = \alpha\kappa(\kappa t)^{\alpha-1}exp(\kappa t)^{\alpha}. \tag{2.48}$$

2.4.3 Gompertz Distribution

Gompertz distribution, named after Benjamin Gompertz, a British mathematician and actuary., who developed it in 1825 (Gompertz, 1825). It is a continuous probability distribution used for modelling adult life spans and other application under different disciplines like; actuarial science, demography, survival and reliability analysis. This distribution is flexible and can be skewed both in right and in left. The pdf, the cdf, sf, hrf and chf of the exponential random variable are, respectively

Let $T \sim$ Gompertz (k, α)

$$f(t) = \alpha c \cdot e^{tk} \exp\left\{-\alpha\left(e^{tk} - 1\right\}\right., \quad t \in [0, \infty) \tag{2.49}$$

$$F(t) = 1 - \exp\left\{-\alpha\left(e^{tk} - 1\right\}\right., \tag{2.50}$$

$$S(t) = \exp\left\{-\alpha\left(e^{tk} - 1\right\}\right., \tag{2.51}$$

$$h(t) == \alpha k \cdot e^{tk}, \tag{2.52}$$

$$H(t) = -\log S(t) = -\log\left[\exp\left\{-\alpha\left(e^{tk} - 1\right\}\right] = \alpha\left(e^{tk} - 1\right) \tag{2.53}$$

Where $k > 0$ is the rate parameter (the scale parameter is scale $= k^{-1}$ is often used in place of k), α is the shape parameter, and $t \geq 0$. The distribution with $k = 1$. Is known as the standard Gompertz distribution. When $k = 0$, the survival time then have an exponential distribution, therefore Gompertz distribution is an extension of exponential distribution. For the Gompertz distribution, $\log(h(t))$ is linear with t.

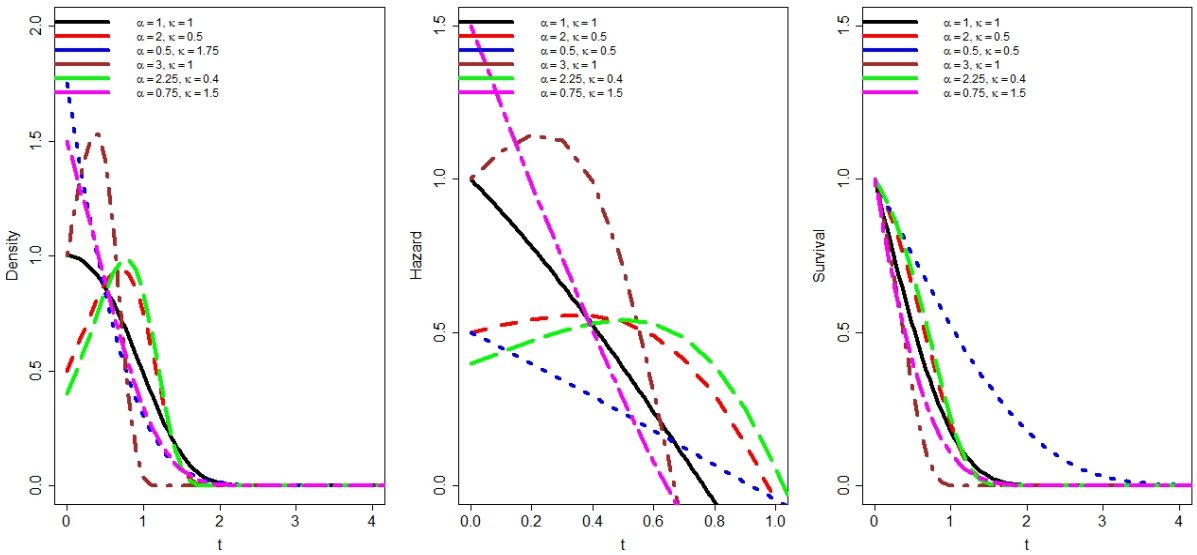

Figure 2.4: Density, hazard and survival shapes for the Gompertz distribution for different values of the rate and shape parameters

2.4.4 Log-logistic Distribution

The log-logistic (LL) distribution (also known as the Fisk distribution) is very effective for modeling unimodal (non-monotone) hazard rate curves. The loglogistic distribution's pdf, cdf, sf,hrf and chf are computed as follows:

$$f(t) = \frac{\alpha\kappa(\kappa t)^{\alpha-1}}{[1+(\kappa t)^{\alpha}]^2} \tag{2.54}$$

$$F(t) = \frac{(\kappa t)^{\alpha}}{1+(\kappa t)^{\alpha}} \tag{2.55}$$

$$S(t) = \frac{1}{1+(\kappa t)^{\alpha}} \tag{2.56}$$

$$h(t) = \frac{\alpha\kappa(\kappa t)^{\alpha-1}}{1+(\kappa t)^{\alpha}} \tag{2.57}$$

$$H(t) = log(1+(\kappa t)^{\alpha}) \tag{2.58}$$

where $t > 0$ is the support of the distribution, and $\kappa > 0$ and $\alpha > 0$ are the parameters, where κ is the rate parameter and α is the shape parameter. The numerator of the hazard function for the log-logistic distribution is the same as the Weibull hazard function, but the total hazard rate has the following features: monotone decreasing for $\alpha \leq 1$, while for $\alpha > 1$ the hazard rate increases initially to a maximum at time $[(\alpha-1)/\kappa]^{1/\alpha}$ and then decreases to zero as time approaches infinity. It is easy to verify that the hazard function of the log-logistic distribution is monotone decreasing for $k \leq 1$, and unimodal for $k > 1$. The log-logistic distribution is commonly used to represent the course of a disease in which mortality peaks after a finite period and then gradually diminishes (Bennett, 1983b). Regarding the odds function and its derivative, they are expressed as:

$$R(t) = exp[H(t)] - 1 = exp\left[log(1+(\kappa t)^{\alpha})\right] - 1 = \kappa t^{\alpha}. \tag{2.59}$$

$$r(t) = h(t)exp[H(t)] = \left[\frac{\alpha\kappa(\kappa t)^{\alpha-1}}{1+(\kappa t)^{\alpha}}\right] exp\left[log(1+(\kappa t)^{\alpha})\right] = \alpha\kappa(\kappa t)^{\alpha-1}. \quad (2.60)$$

It's important to note that the odds function and its derivative for the LL distribution resemble the cumulative hazard function and the hazard rate function for the Weibull distribution, making the LL distribution closed under odds-based regression models.

Figure 2.5 shows several examples of the density, hazard, and survival function shapes for various rate and shape parameter values of the 2-parameter log-logistic distribution.

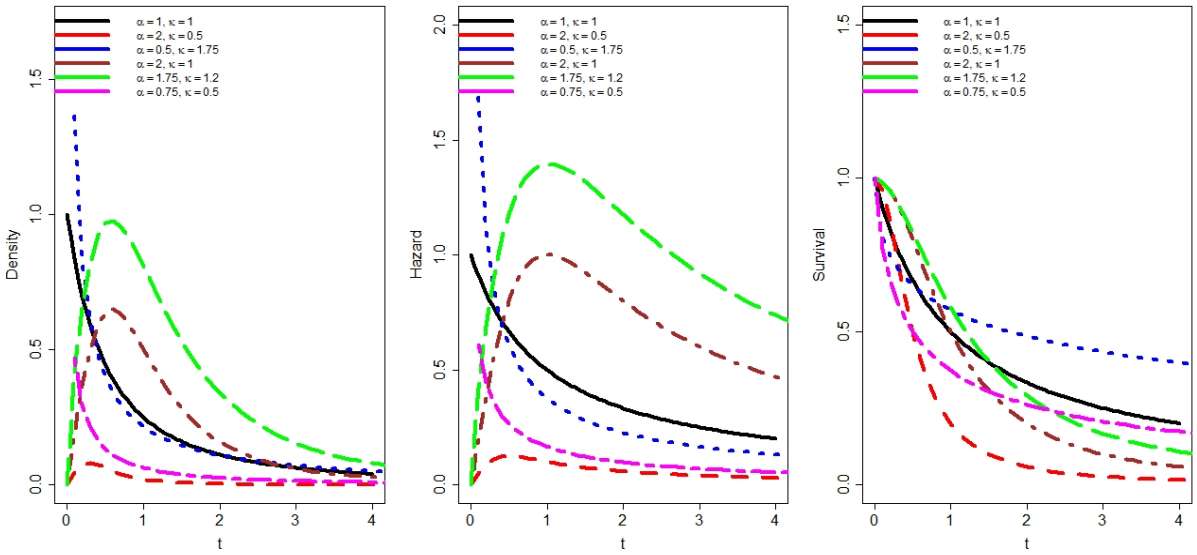

Figure 2.5: Density, hazard and survival shapes for the log-logistic distribution for different values of the rate and shape parameters.

2.4.5 Log-Normal Distribution

The log-normal distribution is a well-known model for representing non-monotone hazard rate data. The pdf, cdf, sf, hrf, and chf equations for the log-normal distribution are as expressed follows:

$$f(t) = \frac{1}{t\alpha\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\log \kappa t}{\alpha}\right)^2\right\} \quad (2.61)$$

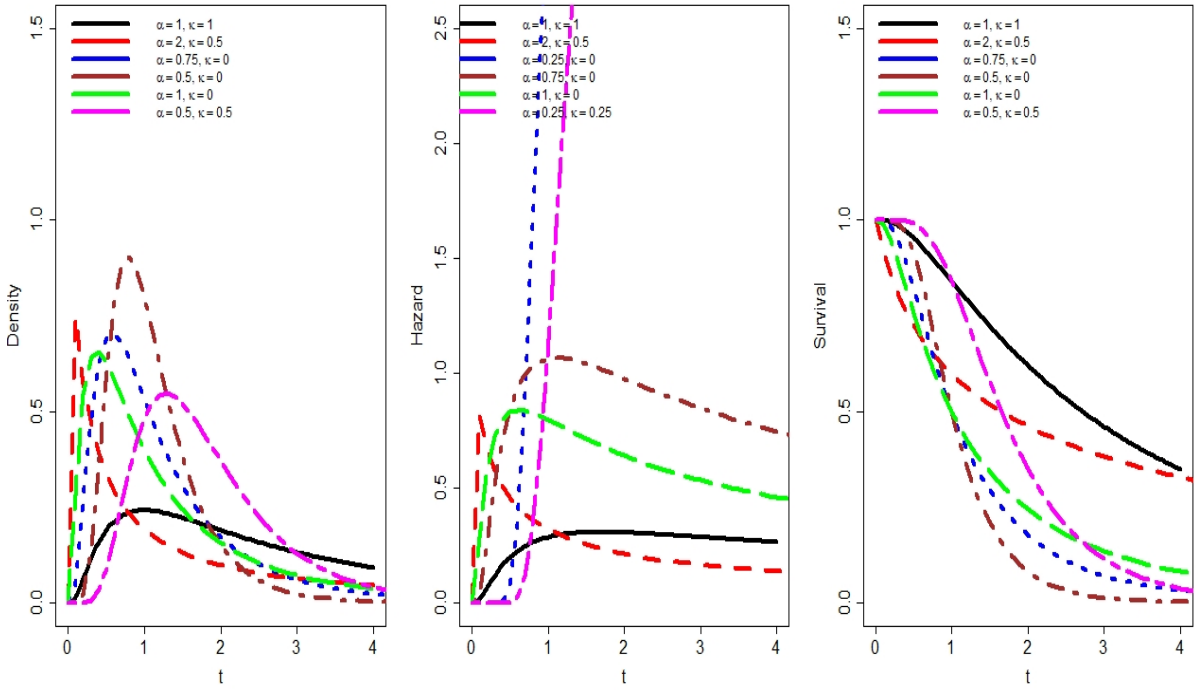

Figure 2.6: Density, hazard and survival shapes for the log-normal distribution for different values of the rate and shape parameters.

$$F(t) = \Phi\left(\frac{\log \kappa t}{\alpha}\right) \tag{2.62}$$

$$S(t) = 1 - \Phi\left(\frac{\log \kappa t}{\alpha}\right) \tag{2.63}$$

$$h(t) = \frac{f(t)}{S(t)} = \frac{\frac{1}{t\alpha\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\log \kappa t}{\alpha}\right)^2\right\}}{1 - \Phi\left(\frac{\log \kappa t}{\alpha}\right)} \tag{2.64}$$

$$H(t) = -log[S(t)] = -log\left[1 - \Phi\left(\frac{\log \kappa t}{\alpha}\right)\right] \tag{2.65}$$

where $t > 0$ is the support of the distribution, $\Phi(\cdot)$ is the standard normal cdf, and $\alpha > 0$ and $\kappa > 0$ are the parameters. Note that if T is log-normal with parameters κ and α, then $Y = \log T$ is n with mean $\alpha = -\log \kappa$ and variance α^2 (Lawless, 2011). Figure 2.6 shows several examples of the density, hazard, and survival function shapes for various rate and shape parameter values of the 2-parameter log-normal model.

2.4.6 Gamma Distribution

Another popular distribution is the gamma distribution. Its relationship to exponential and normal distributions accounts for a considerable portion of its significance. The random variable T has a gamma distribution if

$$f(t) = \frac{\kappa^{\alpha} t^{\alpha-1} \exp\{-\kappa t\}}{\Gamma(\alpha)} \quad t > 0. \tag{2.66}$$

where $\alpha > 0$ is a shape parameter, and $\kappa^{-1} > 0$ is a scale parameter. The exponential distribution is a sub-model of the gamma distribution when $(\alpha = 1)$.

The sf of the gamma distribution is expressed as follows:

$$S(t) = 1 - I_K(\kappa t), . \tag{2.67}$$

where $I_{\alpha}(t) = \int_0^t \frac{u^{\alpha-1} \exp\{-u\}}{\Gamma(\alpha)} du$ is the incomplete gamma function.

The hrf is computed as

$$h(t) = \frac{\kappa^{\alpha} t^{\alpha-1} \exp\{-\kappa t\}}{\Gamma(\alpha) \{1 - I_{\alpha}(\kappa t)\}}. \tag{2.68}$$

which is monotone increasing for $\alpha > 1 (h(t) = 0$ at $t = 0$, and $h(t) \to \kappa$ as $t \to \infty)$, and monotone decreasing for $0 < \alpha < 1 (h(t) \to \infty$ as $t \to 0$, and $h(t) \to \kappa$ as $t \to \infty)$.

Figure 2.7 shows several examples of the density, hazard, and survival function shapes for various rate and shape parameter values of the 2-parameter gamma distribution.

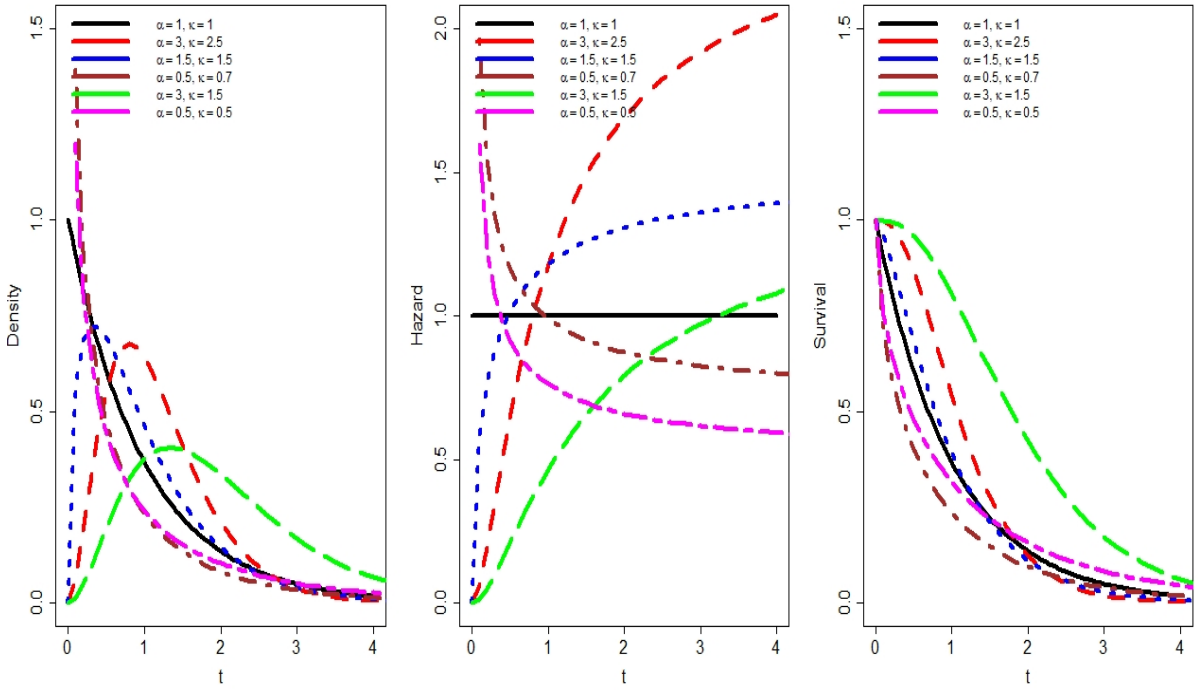

Figure 2.7: Density, hazard and survival shapes for the gamma distribution for different values of the rate and shape parameters.

Weibull, log-logistic, and log-normal distributions are more frequently used in survival analysis than the gamma distribution, in part because its hazard and survivor functions are computationally challenging. Additionally, this can make it computationally difficult to estimate the parameters using maximum likelihood. But the gamma distribution has been used to predict the lifetime of technological systems that have been fixed numerous times after failure, rainfall data in meteorology, and loan and insurance claim data in business (Khosa, 2019).

2.4.7 Burr-XII Distribution

Burr (1942) developed the two-parameter Burr type XII distribution as a result of his investigation into techniques for fitting distribution functions—rather than probability density functions—to frequency data. The Burr type XII distribution also covers a significant portion of the region covered by the gamma and lognormal dis-

tributions in the skewness-kurtosis plane; for more information, see Danish (2016). For the distribution function or the hazard function, it has a good closed-form formulation that is not available for the gamma and log-normal distributions. One of the general parametric families, it offers a wide range of skewness and kurtosis values and covers a number of curve forms. The pdf of a random variable's Burr type XII distribution T is expressed as follows:

$$f_T(t;\kappa,\alpha) = \kappa\alpha t^{\alpha-1}\left(1+t^{\alpha}\right)^{-\kappa-1}; \quad t>0, \tag{2.69}$$

where $\kappa > 0$ and $\alpha > 0$ are the shape parameters of the distribution. The distribution is unimodal with mode at $t = \left(\frac{\alpha-1}{\alpha\kappa+1}\right)^{\frac{1}{\alpha}}$ for $\alpha > 1$ and L-shaped for $\kappa \leq 0$.
The cdf, hrf, and sf of the Burr type XII distribution are computed as follows:

$$F(t;\kappa,\alpha) = 1-\left(1+t^{\alpha}\right)^{-\kappa} \tag{2.70}$$

$$h(t;\kappa,\alpha) = \frac{\kappa\alpha t^{\alpha-1}}{1+t^{\alpha}} \tag{2.71}$$

$$S(t;\kappa,\alpha) = \left(1+t^{\alpha}\right)^{-\kappa} \tag{2.72}$$

It can be easily seen from (2.71) that the hrf is monotone decreasing for $\alpha \leq 1$. It is upside down bathtub shaped curve for $\alpha > 1$, that is, it initially increases, attains a maximum at $t = (\alpha-1)^{\frac{1}{\alpha}}$ and then decreases to zero as t approaches ∞.
Figure 2.8 shows several examples of the density, hazard, and survival function shapes for various shape parameter values of the 2-parameter Burr-XII distribution.

2.4.8 Exponentiated Weibull Distribution

Mudholkar and Srivastava (1993) introduced an extension of the Weibull distribution named exponentiated Weibull (EW) distribution. The cdf, pdf, hrf, and sf of the

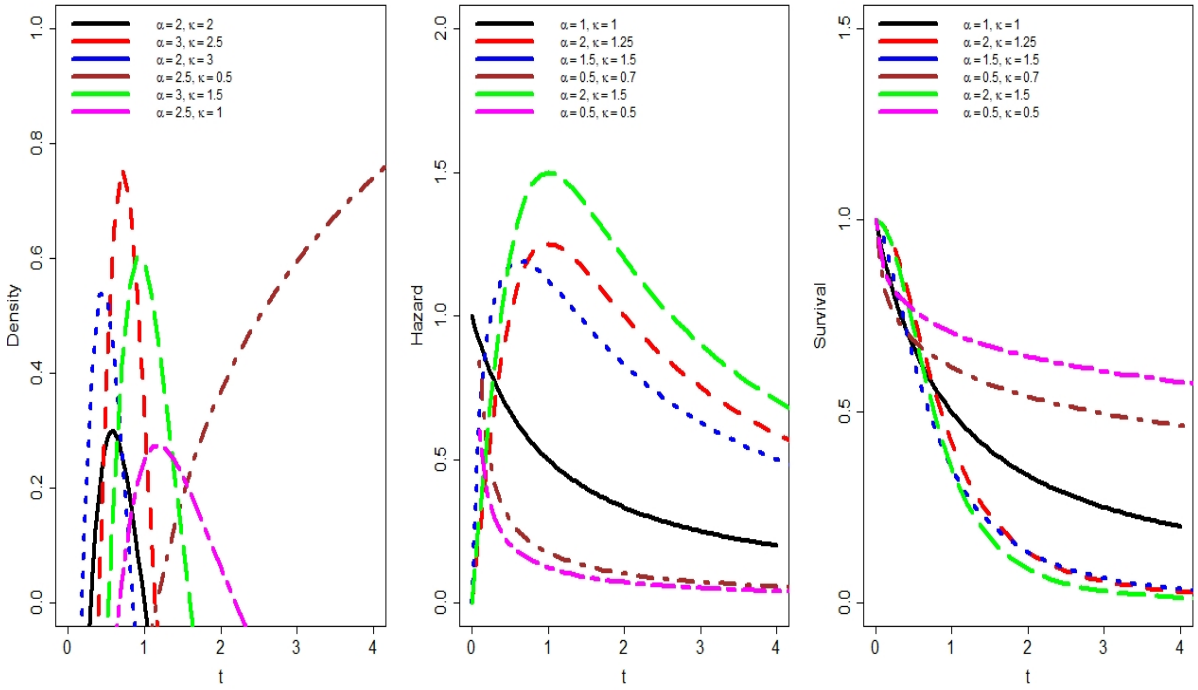

Figure 2.8: Density, hazard and survival shapes for the Burr-XII distribution for different values of the rate and shape parameters.

EW distribution are, respectively,

$$F(t) = \left(1 - \exp\left\{-(\kappa t)^{\alpha}\right\}\right)^{\eta}, \tag{2.73}$$

$$f(t) = \alpha\kappa\eta(\kappa t)^{\alpha-1}\left(1 - \exp\left\{-(\kappa t)^{\alpha}\right\}\right)^{\eta-1}\exp\left\{-(\kappa t)^{\alpha}\right\}, \tag{2.74}$$

$$h(t) = \frac{\alpha\kappa\eta(\kappa t)^{\alpha-1}\left(1 - \exp\left\{-(\kappa t)^{\alpha}\right\}\right)^{\eta-1}\exp\left\{-(\kappa t)^{\alpha}\right\}}{1 - \left(1 - \exp\left\{-(\kappa t)^{\alpha}\right\}\right)^{\eta}}, \tag{2.75}$$

$$S(t) = 1 - \left(1 - \exp\left\{-(\kappa t)^{\alpha}\right\}\right)^{\eta}, \tag{2.76}$$

where $t > 0$ is the support of the distribution, $\kappa > 0$ is a rate parameter, and $\alpha > 0$ and $\eta > 0$ are shape parameters. Note that $\eta = 1$ reduces the exponentiated Weibull to the two-parameter Weibull distribution. Mudholkar and Srivastava (1993) illustrated that the hazard function is (a) monotone increasing for $\alpha \geq 1$ and $\alpha\eta \geq 1$, (b) monotone decreasing for $\alpha \leq 1$, and $\alpha\eta \leq 1$, (c) unimodal for $\alpha < 1$ and $\alpha\eta > 1$,

and (d) bathtub-shaped for $\alpha > 1$ and $\alpha\eta < 1$.

Figure 2.9 demonstrated the hrf shapes for the EW distribution with different values of the scale and shape parameters

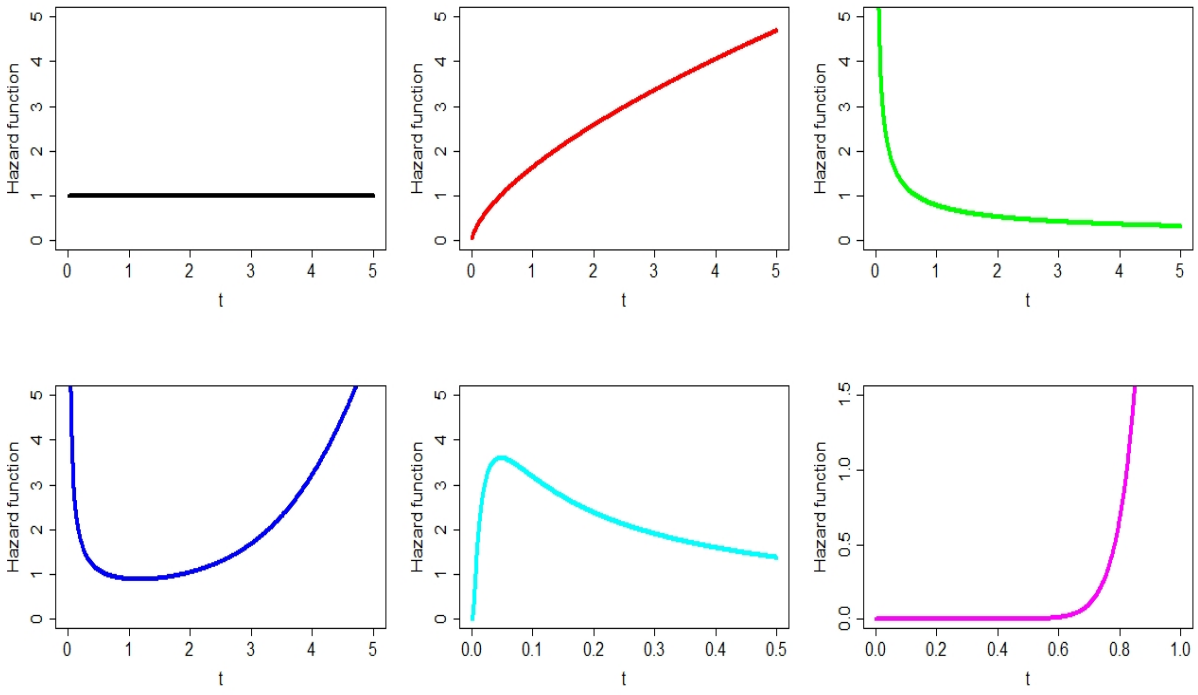

Figure 2.9: Hazard rate shapes for the Exponentiated Weibull distribution with different values of the scale and shape parameters

The exponentiated weibull distribution has shown significant promise in characterizing several types of survival data. The model is adaptable and economical since it can accept both monotone and non-monotone hazard rate functions while estimating only three parameters. It has been effectively used to model bathtub failure rates of electrical devices (Mudholkar and Srivastava, 1993), bus-motor failure data (Mudholkar and Hutson, 1996), and cancer survival data (Khan and Basharat, 2022); and (Rubio et al., 2019).

2.4.9 Why Log-logistic Distribution? and Why its Generalizations?

We chose the generalization of the classical log-logistic distribution as our baseline distribution in this thesis for numerous reasons:

i. It is cdf has an explicit closed-from expression, which is very useful for analyzing time-to-event data with incomplete information (e.g. censoring and truncation) (Bennett, 1983a);

ii. It has a similar shape of pdf and hazard function as the log-normal distribution but has heavier-tails and the tail properties are what the inference is based on (Singh et al., 1988);

iii. It has a non-monotonic hazard function: the hazard function is unimodal when shape parameter is greater than 1 and is decreasing monotonically when shape parameter is less than or equal to 1; this is what makes to be different from the Weibull distribution (Singh, 1998);

iv. It has the potential for analysis of time-to-event data whose rate increases initially and decreases later (Khosa, 2019);

v. It is also used to analyze skewed data, which is prevalent in time-to-event analysis (Lawless, 2011);

vi. The LL distribution can be adopted as the basis of an accelerated failure time (AFT) model by allowing the scale parameter α to differ between groups (Reath et al., 2018).

vii. It has also closed under the proportional odds model (Bennett, 1983a); and the last but not the least

viii. The generalization of the LL distribution has an attractive feature of being a member to all hazard-based regression models (Singh, 1998).

2.5 Survival Regression Models

Many new emerging approaches for analyzing right-censored data have been proposed in recent decades, with the ultimate goal of estimating the covariate effects on the hazard rate function or the odds function. Censored time-to-event data are analyzed using parametric regression models, which make assumptions about how covariates affect the hazard rate or odds function and survival time. There are eight common approaches to how covariates affect hazard or odds function and time scale, namely:

i. Proportional Hazard (PH) Model (Cox, 1972).

ii. Accelerated Failure Time (AFT) Model (Kalbfleisch and Prentice, 1973).

iii. Proportional Odds (PO) Model (Bennett, 1983a).

iv. Accelerated Hazard (AH) Model (Chen and Wang, 2000).

v. General hazard (GH) Model Class (Ciampi and Etezadi-Amoli, 1985; Etezadi-Amoli and Ciampi, 1987; Chen and Jewell, 2001).

vi. Yang and Prentice (YP) Model Class (Yang and Prentice, 2005).

vii. General Odds-Rate (GOR) Hazards Model (Banerjee et al., 2007).

viii. Super Model (SM) (Zhang et al., 2019).

Two other survival regression models based on the mean residual life and vitality functions are proposed respectively;

ix. Proportional mean residual life (PMRL) Model (Oakes and Dasu, 1990)

x. Proportional vitalities (PVIT) Model (Shrahili et al., 2020)

In this section, we reviewed 5 Survival Regression models including: YP, GOR, SM, PMRL and PVIT models that are not discussed in the next chapters.

2.5.1 Yang and Prentice (YP) Model

Yang and Prentice (2005) presented a novel odds-based regression model class to accommodate crossing survival curves. These can occur when, for example, a treatment may be effective in the long run but may have adverse effects on subjects in the early stages of follow-up, or when, in clinical trials of a curable disease, proportions of long-term survival (i.e., near the end of the study) for treatment and control groups may be the same, but deaths (the event of interest) may appear earlier in the control patients. Separately, the abovementioned proportional hazard and proportional odds models are inadequate for survival data with crossing curves.
The odds function of this model is expressed as follows:

$$R_{YP}\left(t;\beta_1,\beta_2,x\right)=e^{x'(\beta_1-\beta_2)}R_0(t)^{e^{x'\beta_2}}. \tag{2.77}$$

The associated derivative of the odds function for the YP model is computed as follows:

$$r_{YP}\left(t;\beta_1,\beta_2,x\right)=r_0(t)e^{x'\beta_1}R_0(t)^{e^{x'\beta_2-1}}. \tag{2.78}$$

The hrf, sf, and chf of the YP model is defined as follows:

$$h_{YP}\left(t;\beta_1,\beta_2,x\right)=\frac{e^{x'(\beta_1+\beta_2)}h_0(t)}{F_0(t)e^{x'\beta_1}+S_0(t)e^{x'\beta_2}}. \tag{2.79}$$

$$S_{YP}\left(t;\beta_1,\beta_2,x\right)=\left[1+\left\{\frac{F_0(t)}{S_0(t)}\right\}e^{x'(\beta_1-\beta_2)}\right]^{-e^{x'\beta_2}}. \tag{2.80}$$

Note that, $\beta_2=0$, and $\beta_1=\beta_2$ give PO and PH as formally nested special cases,

respectively.

2.5.2 General Odds-rate Hazards Model

Banerjee et al. (2007) proposed a generalized odds-rate hazards (GORH) model to analyze censored survival data sets when the PH assumption is not valid. The GOR model contains the PH, AFT, and PO models as special cases. The survival function of this model is expressed as follows:

$$S_{GORH}(t;\theta,\beta,x) = \{1+\theta\exp(\varphi_0(t)+x'\beta)\}^{-\theta^{-1}}. \tag{2.81}$$

Where $t>0, \theta>0$ is the GORH parameter, and $\beta \in R^p$. The hrf of the GORH model is expressed as follows:

$$h_{GORH}(t;\theta,\beta,x) = \varphi_0'(t)\exp(\varphi_0(t)+x'\beta)\{1+\theta\exp(\varphi_0(t)+x'\beta)\}^{-1}. \tag{2.82}$$

Note that, $\theta \to 0, \theta = 1$, and $\varphi_0(t) = \log(t)$, give PH, PO, and AFT as formally nested special cases, respectively.

2.5.3 Super Model

Zhang et al. (2019) introduced a semi-parametric super model, which contains six common survival regression models including PH, PO, AFT, AH, YP and GH models. The super model has closed form hrf that they expressed as follows:

$$h_{SM}(t;\beta_1,\beta_2,\beta_3,x) = \frac{e^{x'(\beta_2-\beta_3+\beta_1)}h_0\left\{te^{\beta_1 x'}\right\}}{e^{x'(\beta_2+\beta_1)}F_0\left\{te^{\beta_1 x'}\right\}+e^{\beta_3 x'}S_0\left\{te^{\beta_1 x'}\right\}} \tag{2.83}$$

The odds function for the super model is expressed as follows:

$$R_{SM}(t;\beta_1,\beta_2,\beta_3,x) = e^{x'(\beta_2-\beta_3+\beta_1)}R_o\left\{te^{\beta_1 x'}\right\}^{e^{(\beta_3-\beta_1)'x}}. \tag{2.84}$$

The survivor function for the SM model is expressed as follows:

$$S_{SM}\left(t;\beta_1,\beta_2,\beta_3,x\right)=\left[1+e^{x'(\beta_2-\beta_3+\beta_1)}\frac{F_o\left\{te^{\beta_1 x'}\right\}}{S_0\left\{te^{\beta_1 x'}\right\}}\right]^{-e^{(\beta_3-\beta_1)'x}}. \tag{2.85}$$

Note that, $\beta_1=0$ YP obtains, $\beta_3=\beta_1+\beta_2$ GH obtains, $\beta_3=0,\beta_1+\beta_2=0$ AH obtains, $\beta_1=0,\beta_2=\beta_3$ PH obtains, $\beta_3=\beta_1=0$ PO obtains, and $\beta_3=\beta_1,\beta_2=0$ AFT obtains as formally nested special cases, respectively.

2.5.4 Proportional Mean Residual Life Model

Oakes and Dasu (1990) introduced the proportional mean residual life (PMRL) model, and they expressed as follows:

$$\mu(t\mid\theta)=\theta\mu_0(t),t\geq 0, \tag{2.86}$$

where θ is a positive parameter, $\mu(\cdot\mid\theta)$ is the mrl function of a response variable, and μ_0 is the baseline mrl function. Then, Zahedi [16] highlighted the role of this model to be played by a regression model. This way, the effect of data in changing the behavior of a baseline mrl function appears in terms of some regression coefficients. Estimation procedures for coefficients of the regression PMRL model have been conducted by Maguluri and Zhang [17] and Chen et al. [18].

2.5.5 Proportional Vitalities Model

Shrahili et al. (2020) introduced the proportional vitalities (PVIT) model as an alternative to the PH, PO, and PMRL models. Unlike the prevalent PH and PMRL models, in the PVIT model, the sf and the pdf of the dependent random variable do not have an explicit closed form. According to Shrahili et al. (2020) the PVIT

model is expresed as follows:

$$v(t \mid \theta) = \theta v_0(t), t \geq 0 \tag{2.87}$$

where $\theta > 0$ is the vitality growth parameter with a time-dependent domain so that $\theta \geq t/v_0(t)$, for all $t \geq 0$, the function $v(\cdot \mid \theta)$ is the vitality function of the population or the dependent variable and v_0 is the baseline vitality function. The model 2.87 is a partial model but it stands valid and qualified under more general circumstances (Kayid, 2021).

2.5.6 Why a Parametric Hazard-based Regression Models?

Despite the fact that the Cox PH model is still the most commonly used survival model (Cox, 1972), there is growing interest in parametric hazard-based regression models (Crowther, 2014). There are several advantages to using a parametric method in the study of time-to-event data. One critical difference is that the Cox PH model does not explicitly estimate the baseline hazard function, but a parametric model does. Measures of absolute risk, such as hazard rates, can be determined directly by simply modeling the baseline hazard function, including estimating the related uncertainty.

This is especially useful to demonstrate how hazard rates alter and evolve over time, both epidemiologically and clinically. If the baseline hazard is correctly stated, a completely parametric approach can be more efficient, with smaller standard errors, than the corresponding Cox model generated using partial likelihood (Collett, 2015). When compared to a Cox PH model with time-dependent effects, modeling time-dependent effects (non-proportional hazards or accelerated models) is significantly easier within a parametric framework and is typically computationally more efficient. Modeling both the baseline hazard and time-dependent effects in continuous time provides the added benefit of permitting predictions both in and out of

sample, which is very important in prognostic modeling. Hazard ratios can also be converted to an absolute scale, which can be used to calculate the number of people who need to be treated. Finally, parametric survival models must be employed inside the economic decision modeling framework to model time-to-event data, as extrapolation is frequently required, frequently across a lifetime horizon, in order to assess anticipated treatment effects and costs (Latimer, 2013).

In particular, modified continuous probability distributions are increasingly being used as a baseline hazard for parametric hazard-based regression models (Rubio et al., 2019). A number of practical studies have recently lauded the merits of the modified distributions baseline method (Khan and Khosa, 2016; Khosa, 2019; Crowther, 2014; Pang, 2020; Burke, 2014). In direct comparison to the Cox PH model, they discovered that "on balance, we prefer the parametric hazard-based regression model over the Cox PH model because it provides the benefits of parametric models while closely matching the Cox estimates," and concluded that "the application of modified baseline distributions that can incorporate different hazard rate shapes proved a powerful method" (Alvares and Rubio, 2021), and "Researchers working in cancer epidemiology and patient survival are encouraged to consider Parametric hazard-based regression models with the application of modified flexible baseline continuous probability distributions as an alternative to the Cox PH model" (Rubio et al., 2019).

2.6 Bayesian and Frequentist Approaches

All of the parametric hazard-based regression models determine a probabilistic mechanism that leads to survival data. The mechanism depends further on a vector of unknown parameters, denoted by θ, which represents the relevant information we wish to pick up from the observed data. The assumed probabilistic mechanism together with the observed data determines the likelihood function, $L(\theta)$, which is the

corner stone to draw the inference about the unknown parameter vector θ.

There are two common philosophical approaches exist in statistics of how to use the likelihood in order to draw the inference about θ, namely the frequentist or classical approach (deductive inferential approach) and Bayesian approach (Inductive inferential approach). The frequentist approach analysis which was founded by Prof. R. A. Fisher in a series of papers round about 1930 considers the variation of data and derived quantities of fixed (deterministic), unknown model parameters. While the Bayesian approach analysis which was first discovered by Thomas Bayes, probability distributions are associated with the parameters of the likelihood, as if the parameters were random variables. They considered the parameters of the model as random variables (has an unknown distribution) and requires that prior distributions specified for them and data are considered as fixed. The key differences between the two schools is fundamentally a question of philosophy (the definition of probability).

In frequentist approach, several techniques exist to estimate the true value of the parameter θ, maximum likelihood estimation (MLE) being one of the most popular ones.

2.6.1 Likelihood for Censored and Truncated Observations

The likelihood function is an important part of the inferential process. Its formulation in the context of survival data analysis requires special consideration because it is dependent on the type of truncation and censoring observations. Assuming that lifetimes and censoring are independent, the probability function of the model's parameters can be formulated by including the corresponding elements such as:

i. **Complete or uncensored case:** when the exact survival time is known, the density of the survival time at the observed time t, $f(t)$.

ii. **Right-censored case:** In the case of a right-censored observations, the sur-

vival function at the censoring time, $S(C_R)$.

iii. **Left-censored case:** In the case of left-censored observations, the cumulative distribution function at the censoring time, $F(C_R) = 1 - S(C_R)$.

iv. **Interval-censored case:** In the case of interval-censored observation, the difference between the survival function at times $C_L - S(c_R)$ and $C_R - S(C_R)$.

v. **Right-truncated case:** In the case of right-truncated observation in which it is assumed that $T \leq T_R$, the density of the survival time at observed time t conditional on the survival time is less than T_L, $f(t)/(1 - S(T_R))$

vi. **Left-truncated case:** In the case of right-truncated observation in which it is assumed that $T > T_L$, the density of the survival time at observed time t conditional on the survival time is greater than T_L, $f(t)/(1 - S(T_L))$

For truncated data we are restricted to use conditional distribution in constructing the likelihood. The likelihood function for truncated data, say i with truncation interval (T_L, T_R) is given by

$$L_i = \begin{cases} f(t_i) / [S(T_L) - S(T_R)] & \text{if } i \text{ is interval-truncated and uncensored} \\ S(C_i) / [S(T_L) - S(T_R)] & \text{if } i \text{ is interval-truncated and censored} \\ f(t_i) / S(T_L) & \text{if } i \text{ is left truncated and uncensored} \\ S(C_i) / S(T_L) & \text{if } i \text{ is left truncated and censored} \\ f(t_i) / [1 - S(T_R)] & \text{if } i \text{ is right truncated} \end{cases}$$

Figure 2.10 summarized the formulation of the likelihood function for the parameteric survival models under censored and truncated observations.

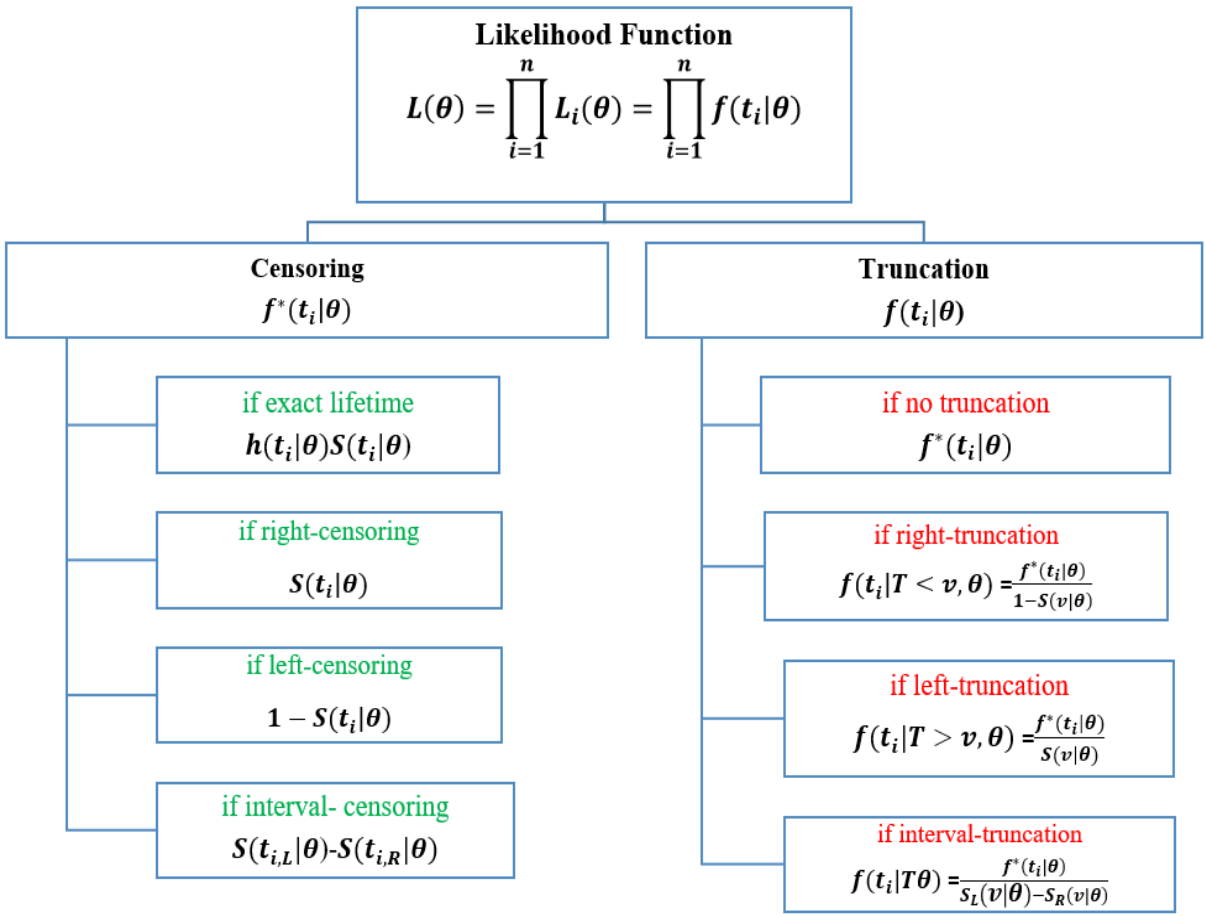

Figure 2.10: The formulation of the likelihood function under the presence of censored and truncated observations

2.6.2 Bayesian Inference

In this section, we present an overview of fundamental concepts and procedures of Bayesian inference. In contrast to the frequentist approach, the Bayesian approach treats parameters as random variables and uses data to update previous knowledge about the parameters (Mwalili, 2006). Bayesian inference is a method for estimating parameters and analyzing data that is based on Baye's theorem. All observable and unobserved parameters in a statistical model are given a joint probability distribution, known as the prior and data distributions, which is unique to Bayesian statistics.

2.6.3 Bayes' Theorem

All forms of uncertainty in Bayesian inference are always defined in terms of probability distributions (van de Schoot et al., 2021). A typical Bayesian workflow includes three major steps Fig. 2.11:

i. **Prior distribution $\pi(\theta)$:** capturing available knowledge about a given parameter in a statistical model via the prior distribution, which is typically determined before data collection. The prior distribution of the parameters reflects prior knowledge of these parameters. A non-informative prior expresses ambiguous knowledge about a variable, whereas an informative prior expresses explicit or definite information about a variable. When more than one parameter is involved, it is common to assume that the prior distribution of each parameter is a priori independent.

ii. **Likelihood Function $L(\theta)$:** determining the likelihood function using the information about the parameters available in the observed data.

iii. **Posterior distribution $\pi(\theta|D)$:** the prior distribution can be combined with the likelihood function using Bayes' Theorem to yield the posterior distribution.

All of these steps are part of the Bayes' theorem, which claims that our updated understanding of the parameters of interest given our current data is dependent on our past knowledge about the parameters of interest weighted by the data's current evidence.

2.6.4 Bayesian Research Cycle

The steps required for a Bayesian research cycle comprise those of a regular research cycle as well as a Bayesian-specific workflow:

a) The standard research cycle consists of reading literature, identifying a problem, and formulating the study question and hypothesis. To increase transparency, the analytic strategy might be pre-registered.

b) A Bayesian-specific workflow includes formalizing prior distributions based on prior elicitation and background knowledge, defining the likelihood function by specifying a data-generating model and including observed data, and gaining the posterior distribution as a function of both the defined prior and the likelihood function. Following the acquisition of the posterior data, conclusions can be drawn that can then be used to initiate a new research cycle.

Although the Bayesian methodology's foundation is basic and intuitive, its application to complicated real-life situations in non-standard probability scenarios and high-dimensional problems was first problematic (**?**). In particular, in a large number of models and applications, m(D) lacks an analytic closed expression, and hence the posterior distribution p lacks a closed form (Ibrahim et al., 2001).
The following modelling uncertainty scenarios are handled by the Bayesian technique (van de Schoot et al., 2021):

i. Overfitting of models

ii. Limited data

iii. If we want to model the data but the information is missing from the data and we have to assume that some facts are more likely than others.

.

2.6.5 Numerical Methods for Bayesian Inference

It is necessary to evaluate integration problems in order to determine the normalized joint posterior distribution, the marginal distribution for a single parameter, the

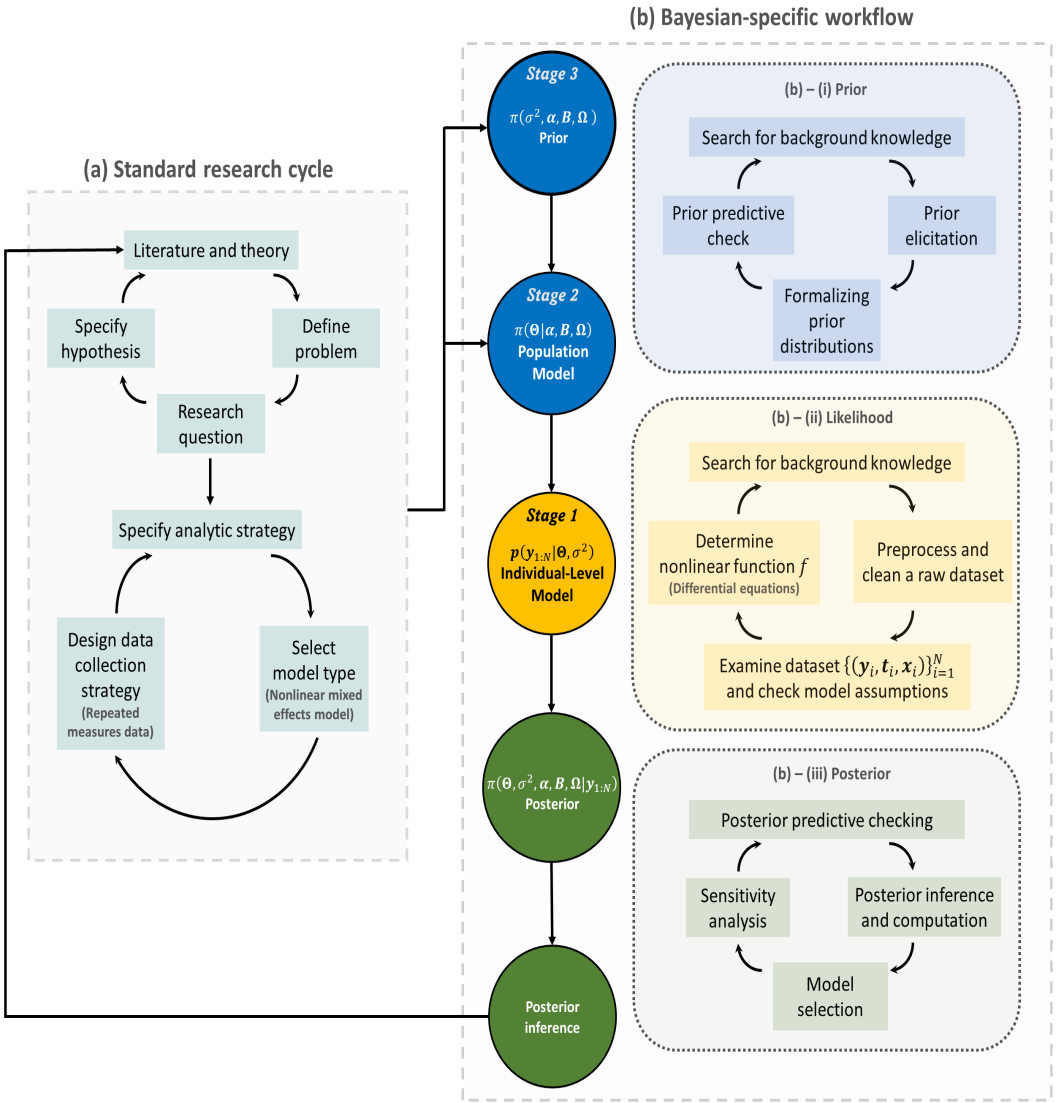

Figure 2.11: Bayesian Research Cycle (van de Schoot et al., 2021)

summary measures of the posterior distribution, and for many other inferential processes. However, analytical solutions are rarely available, so numerical techniques are unavoidable for obtaining the desired results for Bayesian statistics. According to McElreath (2020), the most popular and widely applied numerical approaches are:

i. Grid approximation

ii. Quadratic approximation

iii. Markov chain Monte Carlo (McMC) techniques

2.6.5.1 Grid approximation

The simplest method to approximate the posterior distribution is called grid approximation, and it is as follows:

i. To produce the grids, a finite number of equally spaced points are constructed in the range of values for each parameter.

ii. The posterior is calculated by multiplying the prior and likelihood at the grid points for each set of parameters.

iii. The final step in normalizing posterior involves dividing each value by the total sum of all values.

The number of grids will be equal to 100^k if there are k (let's say) model parameters and 100 (let's say) points for each parameter. The number of grids grows exponentially as the number of points inside the range of values and the number of parameters rises. Even for somewhat big situations, grid approximation is not a practical solution (Christensen et al., 2010). Despite the fact that the method does not require formulas to be derived analytically to calculate the posterior distribution.

2.6.5.2 Quadratic Approximation

In reality, Gaussian or normal approximation is what is meant by the term "quadratic approximation." Because the logarithm of a Gaussian distribution creates a quadratic function parabola, a Gaussian approximation is known as a "quadratic approximation." Because of the parabola, this approximation essentially reflects any log-posterior. The primary tenet of quadratic approximation is that a Gaussian distribution can accurately represent the posterior distribution. When the posterior is unimodal and the sample size is large, this approximation can be fairly accurate. The posterior expectation of any function $g(\theta)$ of the random parameter θ is approximated analytically using Laplace's approach as well. To assess the expectation, Turkman et al. (2019) provided an analytical method:

$$E[g(\theta)|y] = \int g(\theta)p(\theta|y)d\theta \tag{2.88}$$

using the Laplace method to approximate integral.

2.6.5.3 Markov chain Monte Carlo Methods

The exact posterior cannot be discovered analytically except in a few rare instances, and determining the numerical posterior requires challenging numerical integration, especially when there are many parameters. These are the two fundamental obstacles that limit the use of Bayesian statistics (Christensen et al., 2010). However, samples are taken from the posterior distribution and conclusions are based on these samples rather than approximating integrals. There are two methods for simulating samples taken from the posterior distribution: (1) the Monte Carlo (MC) simulation method, which draws independent samples from the posterior distribution; and (2) the Markov chain Monte Carlo (McMC), which draws dependent samples from the posterior distribution.

2.7 McMC Performance and Convergence Diagnostics

In Bayesian inference, McMC algorithms are frequently employed to simulate samples from intricate, high-dimensional posterior distributions. In Bayesian computational statistics, evaluating the chains' performance and the McMC sampling process' convergence is a crucial topic. because the posterior distribution's simulated samples are used for all inference. When the McMC method converges, it does so to the intended posterior density rather than to a single point (Basharat, 2019).

2.7.1 McMC Performance

Every time Markov chains are used, it is crucial to evaluate their performance. The McMC output should be examined using at least two chains, it is recommended. In this sub-section, crucial characteristics for evaluating the performance of Markov chains are described.

2.7.1.1 McMC Mixing

A McMC algorithm's mixing capacity determines how effective it is. A chain's mixing property describes two features (Cowles and Carlin, 1996).

i. How soon a chain forgets its original values comes first.

ii. How rapidly a chain can investigate the target distribution's shape and full support.

2.7.1.2 Number of Chains

In the literature, there are contrasting viewpoints on how many chains should be employed in the McMC algorithm. While some researchers (Gelman and Rubin,

1992) advise for numerous long chains, others only argue for one long chain. The benefits and drawbacks of each strategy were examined by (Consul, 2016; Gilks et al., 1994). The fundamental contention is that, even for very long chains, the chain can wrap itself around the mode of the target distribution and remain there indefinitely. In this situation, even though the chain does not thoroughly explore the support and shape of the target distribution, the convergence diagnostic may nevertheless show that the chain has converged.However, by executing numerous chains, it is possible to guarantee that at least one of them will examine the characteristics of the target distribution and will eliminate the impact of beginning values. In actual practice, two or more chains are typically performed with the expectation that at least one of them will converge to the chosen distribution and investigate all of its features. We build three Markov chains that unavoidably converge to an underlying stationary distribution to implement McMC for our hazard-based regression models.

2.7.1.3 Thinning

The samples produced by McMC are not independent, but rather correlated. Only the k^{th} sampled value is stored, discarding the remaining values in order to reduce autocorrelation, where $k > 1$ is the lag time beyond which autocorrelation is minimal. In most cases, a chain's poor chain mixing can be seen in the autocorrelation's delayed decay. As a result, it is recommended that the inference be based on every i^{th} iteration of chains, with i set to a value that is sufficiently high to ensure that subsequent draws are roughly independent (Gelman and Rubin, 1992). In the literature, this tactic is referred to as "thinning."

Thinned samples have a lower autocorrelation and a larger effective sample size (n_eff). Additionally, it reduces storage requirements and speeds up calculation for high-dimensional problems (Cowles and Carlin, 1996). However, thinning isn't always effective or necessary (Basharat, 2019).

2.7.1.4 Burn-in Period

Burn-in period is the term used to describe the initial phase, or the number of iterations prior to convergence that are abandoned in order to minimize the impact of the initial values. If the sample size is large enough, the impact of this time period is reduced (McElreath, 2020). Even after running for a long time, a Markov chain's dependence on its starting value may continue to be stable. As a result, this reliance may cause the chain to converge slowly if the initial values chosen are significantly different from the posterior mode (Cowles and Carlin, 1996). The chains may need to go through a few rounds before they reach the high probability zone, where they are more indicative of the desired distribution. In essence, the chain is made independent of its beginning values and converges to the target distribution quickly by discarding the first S iterations as part of the burn in period (Gelman, 1996).

2.7.1.5 Stopping Time

A chain should typically be stopped at a specific moment after running for a long enough amount of time to have good mixing. Making a choice about the stopping time in practice might be challenging. Gilks et al. (1994) recommended comparing the estimations (posterior means/medians) from each chain to determine the stopping time after running numerous lengthy chains. The run length should be extended if the estimations from other chains do not closely concur.

2.7.2 McMC Convergence Diagnostics

When all of the chains' outputs are indistinguishable and the chains have "lost" their initial values, McMC convergence is achieved (Mwalili, 2006). However, it might be difficult to decide whether it would be appropriate to assume that the samples closely represent the Markov chain's underlying stationary distribution. A chain's convergence is used to determine how well it has resembled its stationary

distribution (Cowles and Carlin, 1996). Formal test statistics and graphical tools are introduced in the literature to determine the convergence of the chains including the following.

2.7.2.1 Geweke's Convergence Diagnostic

Geweke's diagnostic, also called Geweke's z-score diagnostic, focuses on comparing the first and last parts of a chain. It is, in fact, a frequentist comparison, of means, with 95 percent of the values falling between -2 and 2 , as proposed by (Geweke et al., 1991).

2.7.2.2 Heidelberger and Welch's Convergence Diagnostic

Schruben (1982) and Schruben et al. (1983) proposed detecting nonstationarity in simulation output using a spectral analysis approach to estimate the sample mean variance. They applied the Cramer-von Mises statistic and Brownian bridge theory to test the null hypothesis of stationarity of the Markov chain.
Heidelberger and Welch (1983) applied the aforementioned test to introduce a comprehensive method for generating a confidence interval of a predetermined width for the mean of a parameter when the chain has an initial transient (a state when the algorithm has not reached stationarity yet). They computed a test statistic (based on the Cramer-von Mises test statistic) to reject or accept the null hypothesis that the Markov chain belongs to a stationary distribution. A single chain was subjected to diagnostic.

2.7.2.3 Raftery and Lewis's Diagnostic

Raftery and Lewis (1992) proposed "a method for a single chain that tests for chain convergence to the target distribution and estimates the run-lengths required to properly estimate quantiles of functions of the parameters.

In this thesis, we applied a quantile of interest (0.025), the desired level of accuracy of ±0.0005, and a probability of 0.95 to attain the indicated degree of accuracy (Raftery and Lewis, 1995).

2.7.2.4 Brooks-Gelman-Rubin (BGR) Convergence Diagnostic

Gelman and Rubin (1992) proposed a convergence diagnostic technique to check the McMC algorithms simulation and is based on within chain variance and between chain variance. Gelman et al. Gelman et al. (2013) suggested that the limit of acceptance of potential scale reduction factor (PSRF) to be less than 1.1.

2.7.2.5 Graphical Tools

Four frequently used graphical tools can also be used to assess a chain's convergence and mixing properties: (1) Ergodic mean plot, (2) Autocorrelation plot, (3) Trace plot, (4) Density plot (Cowles and Carlin, 1996).

1. Ergodic mean plot: The Ergodic mean (also referred to running mean) is a well-known convergence diagnostic for McMC algorithms. The Ergodic mean is defined as the mean of all simulated sample values of up to a specific iteration (Smith and Roberts, 1993). Ergodic mean is used to observe the convergence pattern of the McMC chains.

2. Autocorrelation plot: Although the autocorrelation plot is not strictly a convergence diagnostic tool, it does aid in indirectly assessing the convergence of the McMC simulation process (Basharat, 2019).

(3) Density plots: Density can be compared to the fundamental shapes associated with typical analytic distributions, and density plots can reveal behaviour in the tails, skewness, existence of multimodal behaviour, and data outliers (Consul, 2016).

(4) Trace plot: A trace plot, often known as a time series plot, is one of the most popular diagnostics of a McMC simulation [(Cowles and Carlin, 1996). The chain's mixing speed is displayed on the trace plot. The realization of the chains versus the

iteration count is represented by a trace plot. No matter where it began, a well-behaved chain will swiftly depart from its initial values, and the samples will jitter around erratically in the area where the posterior density is strongest, showing that there is no discernible pattern between any two chains. If a distinct trend can be detected in the trace plot, it may be assumed that good chain mixing and stationary distribution have not been attained.

2.8 Markov chain Monte Carlo Sampling

The extensive development of Markov chain Monte Carlo (McMC) sampling methods during the previous decades has made Bayesian inference a viable tool for solving many statistical problems correctly (Mwalili, 2006). McMC simulation methods are a type of stochastic algorithm that is used to sample from posterior distributions. These methods allow for the selection of samples from a probability distribution without knowing the density of the distribution.

Markov chain Monte Carlo (MCMC) methods are a class of algorithms for sampling the posterior distribution based on constructing a Markov chain that has the desired distribution as its limiting (stationary) distribution. The idea of MCMC sampling is to simulate a random walk in the space of parameters of interest, $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_d)'$, which converges to the joint posterior distribution $p(\boldsymbol{\theta} \mid \boldsymbol{y})$. The samples are drawn sequentially, with the distribution of the sampled draws depending on the last value drawn; hence, the draws form a Markov chain. The states of the chain after a large number of iterations is then used as a sample from the desired posterior distribution. As a result, we get a sample of values from the posterior rather than a closed version of it when we use McMC. These samples can then be used to draw conclusions about important derived quantities of interest (Jackson, 2015). In the next subsections, we will briefly outline the most common McMC algorithms:

1. The Gibbs sampler and

2. The Metropolis-Hastings sampling

2.8.1 Gibbs Sampler Algorithm

The Gibbs Sampler (Gilks et al., 1994; Smith and Roberts, 1993) is a multidimensional McMC algorithm that has been shown to be particularly useful. It is defined in terms of θ subvectors. The Gibbs sampler cycles over the subvectors of θ at each iteration t, pulling θ_j from the conditional distribution given all of the remaining components of θ.

Gibbs Sampler Algorithm is written as follows:

1. Choose an arbitrary initial value of $\theta^{(0)} = \left\{\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_k^{(0)}\right\}$

2. For I $= 0, 1, 2, \ldots,$ N $- 1$, generate each component of θ as follows:

   a. Draw $\theta_1^{(I+1)} =$ from $\pi\left(\theta_1 \mid \theta_2^i, \theta_3^i, \ldots, \theta_k^i, Y, X\right)$

   b. Draw $\theta_2^{(I+1)} =$ from $\pi\left(\theta_2 \mid \theta_1^{i+1}, \theta_3^i, \ldots, \theta_k^i, Y, X\right)$

   c. $\theta_3^{(I+1)} =$ from $\pi\left(\theta_3 \mid \theta_1^{i+1}, \theta_2^{i+1}, \theta_4^i, \ldots, \theta_k^i, Y, X\right)$

   d. ...

   e. Draw $\theta_k^{(I+1)} =$ from $\pi\left(\theta_3 \mid \theta_1^{i+1}, \theta_2^{i+1}, \ldots, \theta_{k-1}^{i+1}, Y, X\right)$

3. Repeat step 2 until convergence

4. Return

$$\theta^{(b+1)} = \left\{\theta_1^{(b+1)}, \theta_2^{(0b+2)}, \ldots, \theta_k^{(b+1)}\right\}, \theta_1^{(b+2)}, \theta_2^{(b+2)}, \ldots, \theta_k^{(N)}$$

As a rule of thumb, the iteration should be run until the Monte Carlo error for each parameter of interest should be less than about 5% of the sample standard deviation.

2.8.2 Metropolis-Hastings

Given a target posterior distribution $p(\boldsymbol{\theta} \mid \boldsymbol{y})$, known up to a normalizing constant, the Metropolis algorithm generates a sequence of random vectors $\left(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \cdots\right)$ whose distribution converges to the target distribution. Each sequence can be considered a random walk whose stationary distribution is $p(\boldsymbol{\theta} \mid \boldsymbol{y})$.

Following are the steps in the algorithm (Gelman et al., 2013; Mwalili, 2006; Gelman et al., 1995). Start with some initial value $\boldsymbol{\theta}^0$. For $t = 1, 2, \cdots$, obtain $\boldsymbol{\theta}^{(t)}$ from $\boldsymbol{\theta}^{(t-1)}$ using the following steps:

1. Sample a candidate point $\boldsymbol{\theta}^*$ from a proposal distribution at time t, $q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(t-1)}\right)$. The proposal distribution distribution must be symmetric; that is, $q\left(\boldsymbol{\theta}_a \mid \boldsymbol{\theta}_b\right) = q\left(\boldsymbol{\theta}_b \mid \boldsymbol{\theta}_a\right)$ for all $\boldsymbol{\theta}_a$ and $\boldsymbol{\theta}_b$.

2. Calculate the ratio of the densities,

$$r = \frac{p\left(\boldsymbol{\theta}^* \mid y\right)}{p\left(\boldsymbol{\theta}^{(t-1)} \mid y\right)}.$$

3. Set

$$\boldsymbol{\theta}^{(t)} = \begin{cases} \boldsymbol{\theta}^* & \text{with probablity } \min(r, 1), \\ \boldsymbol{\theta}^{(t-1)} & \text{otherwise.} \end{cases}$$

The algorithm requires the ability to draw θ^* from the proposal (jumping) distribution $q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}\right)$ for all $\boldsymbol{\theta}$.

2.8.3 The Metropolis-Hastings Algorithm

The fundamental Metropolis algorithm mentioned above is generalized in two ways by the Metropolis-Hastings (M-H) algorithm. First off, symmetry in the proposal distribution q is no longer required. That is to say, it is not necessary that $q\left(\boldsymbol{\theta}_a \mid \boldsymbol{\theta}_b\right) =$

$q(\boldsymbol{\theta}_b \mid \boldsymbol{\theta}_a)$. Second, the acceptance ratio has been adjusted to account for the asymmetry in the proposal density (Gelman et al., 1995)

$$r = \frac{p(\boldsymbol{\theta}^* \mid \boldsymbol{y})\, q\left(\boldsymbol{\theta}^{(t-1)} \mid \boldsymbol{\theta}^*\right)}{p\left(\boldsymbol{\theta}^{(t-1)} \mid \boldsymbol{y}\right) q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(t-1)}\right)}.$$

Allowing an asymmetric proposal distribution can be useful in increasing the speed of the random walk.

2.8.4 Why the Bayesian Technique to Inference?

Real-world data sets frequently contain several confounders and necessitate complicated models in order to achieve reasonable results. When we come across generally applicable models, we frequently require complex computational approaches to suit them. As a result, methods like Markov chain Monte Carlo (McMC), that enable sampling from the posterior distribution when no analytical solution exists, are frequently used (Mwalili, 2006). Furthermore, recent technological improvements have supported the use of Bayesian methods for inference in the context of complex models, giving a flexible and powerful alternative to traditional frequentist approaches (Owino, 2014).

The Bayesian technique to inference will be useful for creating flexibility in the form of hazard dependence on covariates. By describing correlations in the prior distribution between log-hazards for nearby covariate profiles, a compromise is achieved that relaxes the assumption of a parametric form of relationship while imposing enough structure to use the information in finite data sets. As a result, selecting a prior distribution might be critical for deriving good posterior inferences.

2.9 Model Comparison

2.9.1 Classical Model Comparison

The comparison between GH, AFT, AH, and PH models based on GLL baseline hazard was evaluated using different information criterions, and the nested structure of the GH model and its special cases was evaluated using the likelihood ratio test (LRT) as discussed below:

2.9.1.1 Nested Models

When the models are nested, we can compare them using a likelihood ratio test (LRT). Assuming we have two models: $f = f(t \mid \theta \in \Theta)$ and $f_0 = f\left(t \mid \theta_0 \in \Theta_0 \subset \Theta\right)$, where $\dim(\Theta) = m$ and $\dim\left(\Theta_0\right) = m - r$, respectively. In other words, f is reduced to f_0 by adjusting r of its parameters to constants. The likelihood ratio test (LRT) is expressed as follows:

$$LRT = -2\log\frac{L(\hat{\theta})}{L(\hat{\theta_0})} = 2\left[\ell(\hat{\theta}) - 2\ell_0\left(\widehat{\theta_0}\right)\right] \sim {x^2}_r, \tag{2.89}$$

where $\hat{\theta}$ is the restricted Maximum likelihood (ML) estimates under the null hypothesis (H_0) and $\hat{\theta}_0$ is the unrestricted ML estimates under the alternative hypothesis (H_1), L is the likelihood function, and ℓ is the log-likelihood function.

Under the null hypothesis, the LRT follows Chi-square distribution with degrees of freedom (df) ($df_{\text{alt}} - df_{\text{null}}$). If the p-value is less than 0.05 the null hypothesis is rejected. In other words, if $LRT > X^2_{r,1-\tau}$, we conclude that the fit provided by f is significantly better than f_0 (at the τ level of significance).

2.9.1.2 Non-nested Model

More generally, models can be non-nested, which means that there is no parameter configuration that makes the two models' equivalent. As a result, we are unable to use the likelihood ratio test. The Akaike information criterion (AIC) is one of the most extensively used methods for comparing non-nested models. The AIC rewards goodness of fit but penalizes the model for increasing the number of estimated parameters and is expressed as follows:

$$AIC = 2(j + p) - 2l, \tag{2.90}$$

where l represents the log-likelihood function evaluated as the MLEs, p the number of covariates and j the number of distributional parameters of the assumed baseline probability distribution (i.e., $j = 3$ for generalized log-logistic baseline). Burnham and Anderson (2004) provided some basic rules of thumb for the use of AIC.

Table 2.1: Rule of Thumb for AIC Differences

Δ_M	Level of support of Model $\boldsymbol{M}$
$0 - 2$	Substantial
$4 - 7$	Considerably less
> 10	Essentially none

Other approaches for model comparison tools for both nested and non-nested models to decide which model best fits the provided data are available. Specifically, Bozdogan's Consistent Akaike Information Criterion (BCAIC), the Bayesian information criteria (BIC), the CAIC (Consistent Akaike Information Criterion), and the Hannan Quin Information Criterion (HQIC).

In scenarios where the sample size is fairly small when compared to the number of parameters in the model, the CAIC fixes the AIC for overfitting of the data and is

calculated as follows:

$$CAIC = AIC + \frac{2(j+p)(j+p+1)}{n-(j+p)-1}, \tag{2.91}$$

Contrary to the AIC, which is asymptotically efficient, the HQIC is frequently quoted in the literature. It is calculated as follows:

$$HQIC = 2(j+p)\log(\log(n)) - 2l, \tag{2.92}$$

The BCAIC is another adjusted form of AIC which is consistent and is computed as follows:

$$BCAIC = 2(j+p)\log(\log(n)) - 2l, \tag{2.93}$$

The BIC also known as Schwarz information criterion (Schwarz, 1978), is used in the same way as AIC (we aim to minimize its value) but has a larger penalty for complexity when $n \geq 8$ (which is typically is). The BIC is computed as follows:

$$BIC = (j+p)\log(n) - 2l, \tag{2.94}$$

where l represents the log-likelihood function evaluated as the MLEs, p the number of covariates and j the number of distributional parameters of the assumed baseline probability distribution (i.e., $j = 3$ for generalized log-logistic baseline).

2.9.2 Bayesian Model Comparison

The Watanabe Akaike information criterion (WAIC), the leave-one-out information criterion (LOOIC), and the deviation information criterion (DIC) were the three Bayesian model selection criteria that we employed in this thesis.

In chapters 3, 4, and 5, DIC was taken into account for the Bayesian model selection, but chapters 6, 7, and 8 take WAIC and LOOIC into account. As a general rule, the

model with the smaller DIC is chosen as the best fitting if there is a DIC difference between two models of greater than 3. The DIC is computed as follows:

$$DIC = \bar{D} + pD = \widehat{D} + 2pD \tag{2.95}$$

where; $\bar{D}$ is a goodness of fit test for statistical model and is represented the deviane's posterior mean, and pD is computes as the difference between $pD = \bar{D} - \widehat{D}$, and it is represented the effective number of model parameters.

WAIC and LOOIC were employed as full Bayesian model selection criteria in this study. They are both techniques for calculating pointwise out-of-sample prediction accuracy using a fitted Bayesian model. Asymptotically, they are equivalent since WAIC is based on the series expansion of leave-one-out cross-validation (LOO). It is helpful to be able to compute both WAIC and cross-validation because they address different prediction questions with finite data. The log-likelihood assessed at the posterior simulations of the parameter values can be used to directly estimate the WAIC and an approximated LOO based on importance sampling. Compared to more basic estimates of prediction error like AIC and DIC, LOOIC and WAIC have a number of advantages, but they are less frequently employed in practice since they require additional computing steps (Vehtari et al., 2017; Magnusson et al., 2020).

2.10 Total Time on Test

The theory of total time on test (TTT) transform is familiar for its use in different fields of study such as stochastic modeling, econometrics, survival and reliability analysis, and ordering of distributions (Nasiru, 2018). Researchers in survival and reliability analysis are often interested in how the shape of the hrf of a given data set looks like. The TTT-transform provides the researchers with a graphical way of viewing the shape of the hrf.

In the literature, the major share of the TTT-transform is concerned with survival and reliability problems that include the characterization of aging properties, tests of hypothesis, ordering life distributions, model identification, and defining new classes of lifetime distributions. The method was developed by (Barlow and Doksum, 1972), for statistical inference problems under order restrictions.

For a random variable representing a lifetime, with a distribution function F(x) and survival function S(x), the function defined on $[0, 1]$ by

$$H_F^{-1}(p) = \int_0^{F^{-1}(P)} S(u)du, p \in [0, 1] \tag{2.96}$$

It is called the TTT-transform of F. Where $S(u) = 1 - F(u)$ is the survival function. The scaled TTTtransform is computed using

$$\varphi G(p) = \frac{H^{-1}(p)}{H^{-1}(1)} \tag{2.97}$$

The curve $\varphi G(p)$ versus $0 \leq p \leq 1$ is the scaled TTT-transform curve.

According to Barlow and Doksum (1972), the shape of the hrf can be classified as one of the following using the scaled TTT-transform curve:

1. The hrf is said to be **constant** if the scaled TTT-transform curve is on the 45^0 line.

2. The hrf is **increasing** if the scaled TTT-transform curve is concave above the 45^0 line.

3. The hrf is **decreasing** if the scaled TTT-transform curve is convex below the 45^0 line.

4. The hrf exhibits a **bathtub shape** if the scaled TTT-transform curve is first convex below the 45^0 line and then concave above the line.

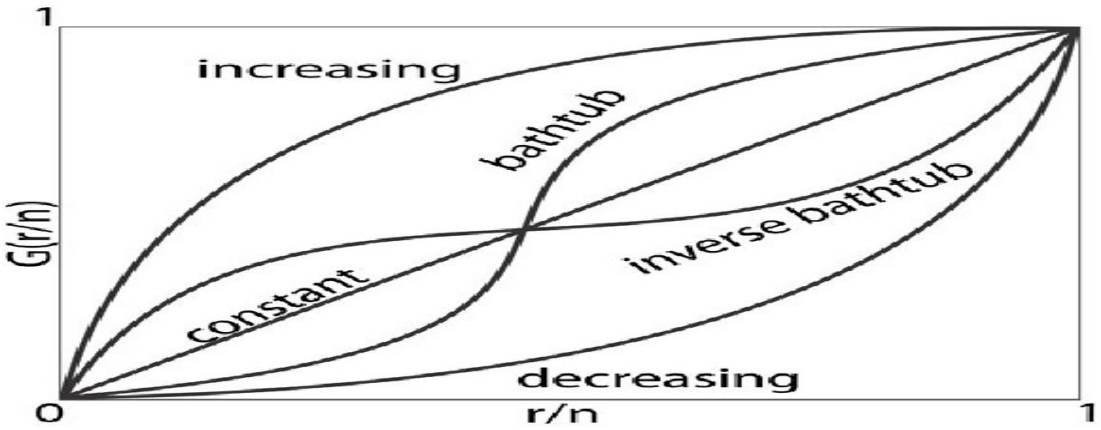

Figure 2.12: Shapes for the TTT-Plot (Ramos et al., 2014)

5. The hrf is **uni-modal** or **inverse bathtub** shape if the scaled TTTtransform is first concave above the 45^0 line and then convex below the 45^0 line.

2.11 Concluding Remarks

The chapter provided some of the key concepts of modelling censored survival data and a comprehensive explanation of the several methods used in order to attain the objectives of the study.

CHAPTER 3

Bayesian and Classical Inference for the Generalized Log-Logistic Distribution with Applications to Survival Data

In this chapter, we present our second published manuscript [1] about our proposed baseline hazard named generalized log-logistic distribution with application to survival data set. Note that the materials of this chapter have been reproduced from our article.

3.1 Introduction

Applied Statisticians use many probability distributions for reliability and survival studies. The distributions could be applied in different fields such as medicine, engineering, economy, industrial and physical fields, and so many other fields. Exponential distributions, Generalized Exponential distributions, Gamma distributions, Generalized Gamma distributions, extreme value distributions, Weibull distributions, log-logistic distributions, log-Normal distributions, Burr XII, and Generalized Weibull distributions are among the most commonly used distributions in survival and reliability analysis.

Typically, researchers in reliability and survival analysis are concerned with the development of new probability models. Log-Logistic distribution (LL) is one of the parametric distributions that can be used as a life testing distribution because of the

[1] Muse, A. H., Mwalili, S., Ngesa, O., Almalki, S. J., and Abd-Elmougod, G. A. (2021). Bayesian and classical inference for the generalized log-logistic distribution with applications to survival data. Computational intelligence and neuroscience, 2021.

simplicity of its cumulative distribution and survival function can both be stated in closed form and because it belongs to the Scale-Shape family (Lawless, 2011). LL is one of the right-skewed, heavy-tailed functions that can be used as an alternative to a log-normal distribution. It resembles the log-normal distribution in shape but has heavier tails. Log-logistic distribution is particularly applicable to model non-monotone (i.e., unimodal) hazard functions.

It is well understood that the log-logistic model is not appropriate for modelling where the failure rate is monotonic when analysing time-to-event data with parametric models. It is suitable to use an extension of the model which has a monotone hazard function. Departures from the monotonicity of distribution are typically studied in terms of its shape, or more specifically in terms of its skewness (also referred to as asymmetry) and kurtosis.

In this chapter, we focus on a modification of the log-logistic model because it resembles the log-normal distribution in shape but is better suited for the application in the analysis of survival data when dealing with incomplete data, such as censored observations which are common in survival data analysis (Bennett, 1983b). The presence of incomplete observations causes difficulties when using log-normal or inverse Gaussian models since the survival functions in these cases are complicated. On the other hand, since the logarithms of small positive numbers are large negative numbers, the log-normal distribution may give undue weight to very short survival times (Lawless, 2011). For the reasons stated above, we will focus on the log-logistic model whose hazard rate exhibits the aforementioned behaviour.

However, due to the log-logistic model's symmetric property, it may be inadequate for cases where the hazard rate is heavily tailed or skewed, as well as for modelling censored survival data (Singh et al., 1988; Bennett, 1983b; Kalbfleisch and Prentice, 1973). In this study, we studied a modification (or generalization) of the log-logistic parametric survival model and referring to this as the generalized log-logistic distribution given in (Khan and Khosa, 2016). The generalized log-logistic distribution

reflects the structure of the heavy tails and the skewness and it significantly outperformed the log-logistic distribution in general.

In the statistical literature, with the aim of increasing the versatility of the log-logistic distribution in modelling survival time data, different generalized forms of the distribution have recently been proposed and includes: a new extension of the log-logistic distribution with applications to actuarial data sets (Alfaer et al., 2021); Alpha power transformed log-logistic distribution (Aldahlan, 2020), (Malik and Ahmad, 2020); transmuted four-parameters generalized log-logistic distribution (Adeyinka and Olapade, 2019; Granzotto and Louzada, 2015); a new three-parameter of the log-logistic distribution (Shakhatreh, 2018); extended log-logistic distribution (Lima and Cordeiro, 2017); exponentiated log-logistic geometric distribution (Mendoza et al., 2016); the log-logistic Weibull distribution (Oluyede et al., 2016); beta log-logistic distribution (Lemonte, 2014); McDonal log-logistic distribution (Tahir et al., 2014); transmuted log-logistic distribution (Aryal, 2013); Marshall-Okin extended log-logistic distribution (Gui, 2013); the Zografos-Balakrishnan log-logistic distribution (Hamedani, 2013); and exponential- log-logistic distribution (Rosaiah et al., 2014). More details about the modifications and recent generalizations of the log-logistic distribution can be found in (Muse et al., 2021b).

In addition, other authors have studied the Bayesian inference of the log-logistic distribution and some of its generalizations. Dos Santos et al. (2018) developed a Bayesian analysis of the transmuted log-logistic distribution. Yahaya and Dewu (2017) studied the Bayesian estimation of the scale parameter for the log-logistic distribution using chi-square and Maxwell priors. Abbas and Tang (2016) studied the objective Bayesian analysis of the log-logistic distribution using the reference and Jeffreys prior. Al-Shomrani et al. (2016) focused on the application of the Markov chain Monte Carlo (McMC) techniques for estimating the unknown parameters of the log-logistic distribution. Guure et al. (2015) explored the Bayesian inference of the log-logistic distribution for the interval-censored data. Kang et al.

(2014) proposed the non-informative priors for the log-logistic distribution. Chaudhary (2019) studied the Bayesian estimation of the three-parameter exponentiated log-logistic distribution. Akhtar et al. (2014) discussed the Bayesian analysis of the log-logistic distribution using the Laplace approximation. Chaudhary (2019) proposed the Bayesian analysis of the two-parameter exponentiated log-logistic distribution.

The log-logistic distribution has large-scale applications in analyzing time-to-event data. The model is closed under both proportionality (multiplication) of failure time and proportionality of odds. though, it is not a proportional hazard (PH) model. However, regarding this issue Khan and Khosa (2016) presented generalized log-logistic distribution that belongs to the proportional hazard models. The proposed distribution has similar properties to the 2-parameter log-logistic distribution and approaches the Weibull distribution in limit. However, its statistical and mathematical properties, as well as inferential procedures, have not received attention so far. On the other hand, they discussed the classical inference of the proposed distribution under the PH regression framework. However, much work still has to be done. In this paper, we focused on the Bayesian and classical inference of the generalized log-logistic distribution as a generalized distribution, not as a regression model.

Additionally, for the applied cases, especially in the survival modelling, the GLL model could be applicable in the following cases; (1) modelling the "asymmetric monotonically right-skewed" heavy tail data sets; (2) modelling the "bathtub-shaped hazard rate" data sets like data set I; (2) in "survival analysis", the GLL distribution could be chosen for modelling Proportional hazard frameworks; (4) in the medical field, the GLL distribution could be considered in modelling the "Bladder cancer data sets" which have "reversed bathtub-shaped" as illustrated in data set I; (5) in the reliability and survival analysis, the proposed distribution can be an alternative to the Weibull distribution since it can be closed under both accelerated failure time

(AFT) and PH models since the Weibull distribution fails to model unimodal data. For these based on ground reasons, we are motivated to study and introduce the GLL distribution.

Thus, the main goal of this chapter is to propose and study a generalized log-logistic distribution, which extends the exponential, Weibull, log-logistic, and BurrXII distributions, with the hope that the proposed distribution may have a better fit compared to these distributions and other 3-parametric distributions in certain practical situations. In addition, we would provide a comprehensive account of the mathematical and statistical properties of the proposed model. The proposed model's formulae are simple and tractable, and with the use of modern computer software and its numerical capabilities, the proposed model could be a great addition to the arsenal of applied mathematicians and statisticians in the areas like medicine, engineering, economics, social sciences, biology, among others. Finally, we discussed the Bayesian model formulation of the proposed distribution

The rest of the chapter is organized as follows. Section 3.2 describes the distribution functions for the GLL distribution, its sub-model distributions, and some of its basic properties. Some mathematical properties of the GLL distribution are derived in Section 3.3. Section 3.4 describes the Maximum likelihood for the estimation parameters of GLL distribution. Section 3.5 discusses the results of a simulation study aimed at investigating and comparing the performance of the proposed estimators. Section 3.6 presents an analysis of a real-life data set. The Bayesian Model formulation for the proposed distribution is discussed in section 3.7. Section 3.8 presents the Bayesian analysis of a real-life data set using Markov chain Monte Carlo techniques. Finally, Section 3.9 summarizes the study with some concluding remarks.

3.2 The Generalized Log-logistic Distribution

The generalized log-logistic distribution is a continuous probability distribution with positive support R on a subset of $(0,\infty)$ with three parameters. It is a generalization of the two-parameter log-logistic distribution. The generalization of log-logistic distribution for censored survival data can be traced back to (Singh et al., 1988) who discussed a generalized log-logistic distribution and applied it to censored survival data, and proposed a generalized log-logistic model and introduced the shape parameter and then they used to fit a lung cancer data. Prentice (1976) proposed a generalization for quantile response data and discussed several of its uses.

Since many continuous probability distributions are commonly applied for parametric models in survival analysis like the exponential, Gompertz, Weibull, log-normal, log-logistic, and the gamma distribution. GLL is also applicable for survival data analysis. There are a number of probability functions that are related to continuous probability distributions, we will concentrate on functions that are related to the lifetime distributions as a random variable in this study. The most common ones are; hazard rate function (or failure rate function) hrf, survival (or reliability function), probability density function (pdf), cumulative distribution function (cdf), cumulative hazard rate function (CHF), and the reversed hazard rate function (rhrf) or retro hazard. The advantage of these functions is that they completely describe the lifetime distribution, and if you have any of them, determining the others is simple.

3.2.1 Hazard Rate Function

The hazard (failure) rate function plays an important role in survival analysis. It is the most popular function for analyzing and modelling lifetime data because of its intuitive interpretation of the amount of risk to fail associated with a unit time t. applicable for describing the lifetime distribution of engineered and other

components. The hazard rate is more informative than all of the other functions in lifetime distributions. Because of this (Khan and Khosa, 2016) started their work by defining the hazard rate of the GLL distribution. Cox and Oakes (2018) described the reason that why the hazard rate is considering when we are dealing with the survival data. They give a number of reasons include; hazard rate-based models are often convenient when there is incomplete information (censoring) or there are several types of failure rates, also hazard rate is a special form of the intensity function, and last but not least is that hazard rate function can be derived from all other functions that we use to describe lifetime distributions.

The hazard rate function describes how the instantaneous failure rate changes over time. For the GLL distribution, the hazard rate function is computed as:

$$h(x;\boldsymbol{\theta}) = \frac{\alpha k(kx)^{\alpha-1}}{[1+(\eta x)^{\alpha}]}, \quad x \geq 0, \quad k, \alpha, \eta > 0 \tag{3.1}$$

where $k > 0, \beta > 0, \eta > 0$ are the distributional parameters and $\boldsymbol{\theta} = (k, \alpha, \eta)'$. It can be easily seen from Equation (3.1) that the hazard rate function is monotone decreasing for $\alpha \leq 1$, and unimodal when $\alpha \leq 1$. That is, it initially increases to a maximum at $t = \left[\frac{\alpha-1}{\lambda^{\alpha}}\right]^{\frac{1}{\alpha}}$, and then decreases to zero monotonically as $t \to \infty$. The plots of the hrf are shown in Figure 3.1

3.2.2 Sub-Models

The proposed distribution consists of a number of important sub-models that are widely used in parametric survival modelling. These include the log-logistic distribution, the standard log-logistic distribution, the Burr XII distribution, the Weibull distribution, and the exponential distribution. The propositions below relate the GLL to the log-logistic, standard log-logistic, Burr XII, Weibull, and exponential distributions.

1. Log-logistic Distribution:

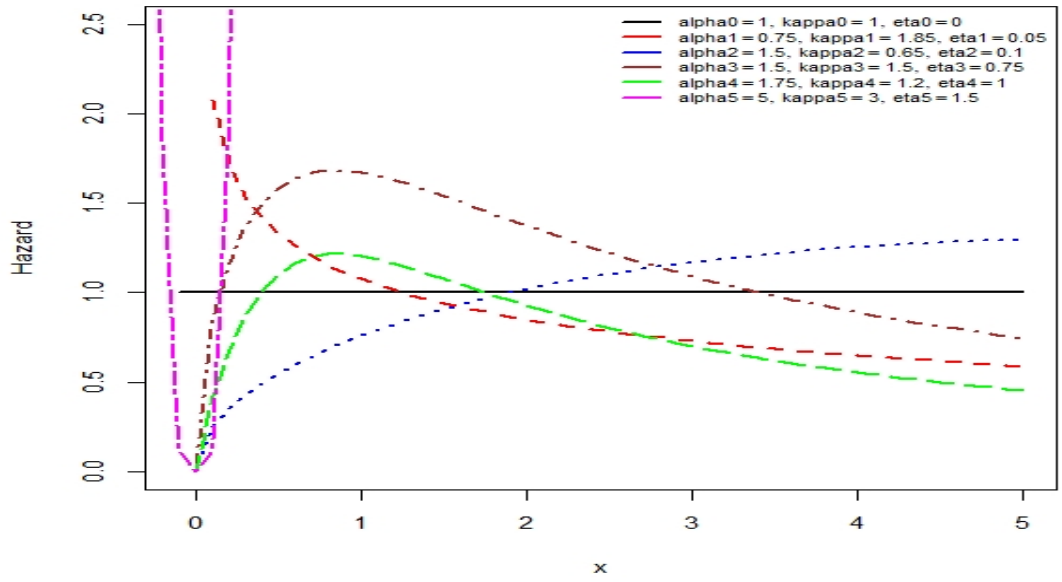

Figure 3.1: The hazard rate shapes for the GLL distribution

Proposition 3.2.1 *Let $X \sim \text{GLL}(\alpha, k, \eta)$. If η depends on k via $k = \eta$, then the hazard rate function of Equation (3.1) reduces to the hazard rate function of the log-logistic distribution.*

Proof 3.2.1 *From the hazard rate function of the generalized log-logistic distribution given by*

$$h(x;\boldsymbol{\theta}) = \frac{\alpha k(kx)^{\alpha-1}}{[1+(\eta x)^{\alpha}]}$$

If we replace $\eta = k$, it gives us

$$h(x;\boldsymbol{\theta}) = \frac{\alpha k(kx)^{\alpha-1}}{[1+(kx)^{\alpha}]} = \frac{\alpha k(kx)^{\alpha-1}}{[1+(kx)^{\alpha}]} \tag{3.2}$$

Which is the hazard rate function form of a log-logistic distribution with the two unknown parameters (k, α). Where $\theta = (k, \alpha)', k = \frac{1}{\beta}$ is the rate parameter.Hence, the proof

It is easy to verify that the hazard rate function of the log-logistic distribution is

monotone decreasing for $0 < \alpha \leq 1$, and uni-modal for $\alpha > 1$ (decreases and then increases with the maximum at $x = \frac{1}{k}(\alpha - 1)^{\frac{1}{\alpha}}$.

2. Standard Log-logistic Distribution

Proposition 3.2.2 *Let $X \sim GLL(\alpha, k, \eta)$. If η depends on k via $k = \eta = 1$, then the hazard rate function of Equation (3.1) reduces to the hazard rate function of the standard log-logistic distribution.*

Proof 3.2.2 *From the hazard rate function of the generalized log-logistic distribution given by*

$$h(x; \boldsymbol{\theta}) = \frac{\alpha k (kx)^{\alpha - 1}}{[1 + (\eta x)^{\alpha}]}$$

If we replace $\eta = k = 1$, it gives us

$$\begin{aligned} h(x; \boldsymbol{\theta}) =& \frac{\alpha \cdot 1 (1 \cdot x)^{\alpha - 1}}{[1 + (1.x)^{\alpha}]} \\ =& \frac{\alpha (x)^{\alpha - 1}}{[1 + x^{\alpha}]} \end{aligned} \tag{3.3}$$

Which is the hazard rate function form of a standard log-logistic distribution with one unknown parameter (α). Hence, the proof

Note that, where $x > 0$, is the support of the distribution, and $\alpha > 0$ is the shape parameter. It is easy to verify that the hazard rate function of the log-logistic distribution is monotone decreasing for $0 < \alpha \leq 1$, and uni-modal for $\alpha > 1$ (decreases and then increases with the maximum at $x = (\alpha - 1)^{\frac{1}{\alpha}}$.

3. Burr-XII Distribution

Proposition 3.2.3 *Let $X \sim GLL(\alpha, k, \eta)$. If $\eta = 1$, then the hazard rate function of Equation (3.1) reduces to the hazard rate function of the Burr-XII distribution.*

Proof 3.2.3 *From the hazard rate function of the generalized log-logistic distribution given by*

$$h(x; \boldsymbol{\theta}) = \frac{\alpha k (kx)^{\alpha - 1}}{[1 + (\eta x)^{\alpha}]}$$

If we replace $\eta = 1$, it gives us

$$h(x;\boldsymbol{\theta}) = \frac{\alpha k(kx)^{\alpha-1}}{[1+(1.x)^{\alpha}]} = \frac{\alpha k x^{\alpha-1}}{[1+x^{\alpha}]} \tag{3.4}$$

Which is the hazard rate function form of a Burr XII distribution with two unknown parameters (α, k). Hence, the proof

The Burr XII hazard function is monotone decreasing for $\alpha \leq 1$, and upside-down bathtub shapes curve for $\alpha > 1$ (which means that it initially increases, attains a maximum at $x = (\alpha - 1)^{\frac{1}{\alpha}}$, and then decreases to zero at $x \to \infty$).

4. Weibull Distribution

Proposition 3.2.4 *Let $X \sim GLL(\alpha, k, \eta)$. If $\eta^{\alpha} \to 0$. then the hazard rate function of the GLL Equation (3.1) approaches to the hazard rate function of the Weibull distribution.*

Proof 3.2.4 *If we now let If $\eta^{\alpha} \to 0$ then, from the hazard rate function of the GLL given by:*

$$h(x;\boldsymbol{\theta}) = \frac{\alpha k(kx)^{\alpha-1}}{[1+(\eta x)^{\alpha}]},$$

we have that

$$h(x;\boldsymbol{\theta}) = \frac{\alpha k(kx)^{\alpha-1}}{[1+(0]},$$

simplifying gives,

$$h(t;\boldsymbol{\theta}) = \alpha k(kx)^{\alpha-1} \tag{3.5}$$

which is a hazard function of a Weibull distribution with the unknown parameters (α, k). This property of the GLL enables it to handle monotone increasing hazard satisfactorily with $\alpha > 1$, and λ close to zero (very small).

It is clear from Equationn (3.5) that $0 < \alpha < 1$, the hazard rate function decreases, for $\alpha > 1$, the hazard rate function increases, for $\alpha = 1$, the hazard rate function

decreases. The distribution reduces to exponential for $\alpha = 1$.

5. Exponential Distribution

Proposition 3.2.5 *Similarly, if we now let $\alpha = 1$, then the hazard rate function of Equation (3.5) reduces to the hazard rate function of the exponential distribution.*

Proof 3.2.5 *From Equation (3.5) we have that the hazard rate function is*

$$h(t;\boldsymbol{\theta}) = \alpha k(kx)^{\alpha-1}$$

if we replace $\alpha = 1$

$$h(t;\boldsymbol{\theta}) = \alpha k(kx)^{\alpha-1}$$

simplifying gives,

$$h(t;\boldsymbol{\theta}) = k \cdot 1(1.t)^{1-1}$$

$$h(t;\boldsymbol{\theta}) = k \tag{3.6}$$

which is the hazard rate function of an exponential distribution.

This property makes the exponential distribution to be inadequate to describe survival data. Hence, the proof.

The summary of the submodels for the proposed distribution are summarized in Table 3.1.

Table 3.1: Summary of submodels from the GLL distribution.

Distributions	α	η	k
Log-logistic distribution	α	$\eta = k$	$k = \eta$
Weibull distribution	$\eta^{\alpha} \longrightarrow 0$	$\eta^{\alpha} \longrightarrow 0$	k
Exponential distribution	$\alpha = 1$	$\eta \longrightarrow 0$	k
Standard log-logistic distribution	α	$\eta = k = 1$	$k = \eta = 1$
Burr XII distribution	α	$\eta = 1$	k

3.2.3 The Probability Density Function

The pdf of the GLL distribution with three unknown parameters can be obtained by applying the following equation and the plots of the pdf are shown in Figure 3.2

$$f(x;\boldsymbol{\theta}) = h(x,\theta)\exp\left\{-\int_0^x h(x)dx\right\} \tag{3.7}$$

Simplifying gives;

$$f(x;\boldsymbol{\theta}) = \frac{\alpha k(kx)^{\alpha-1}}{[1+(\eta x)^{\alpha}]^{\frac{k^{\alpha}}{\eta^{\alpha}+1}}}, \quad x \geq 0, \quad k,\alpha,\eta > 0 \tag{3.8}$$

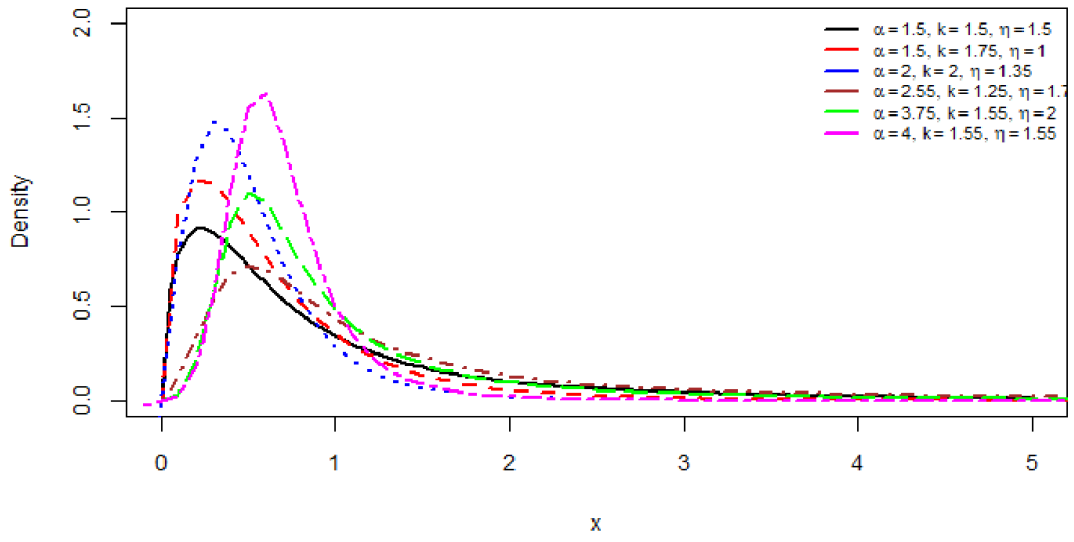

Figure 3.2: The pdf shapes for the GLL distribution

3.2.4 The Survival (or Reliability) Function

The survival (reliability) function of the GLL distribution that represents the probability that observation does not fail until t is given below and its plots are shown

in Figure 3.3 .

$$S(x;\boldsymbol{\theta}) = \frac{f(x;\boldsymbol{\theta})}{h(x;\boldsymbol{\theta})} \tag{3.9}$$

Simplifying gives;

$$S(x;\boldsymbol{\theta}) = [1+(\eta x)^{\alpha}]^{-\frac{k^{\alpha}}{\eta^{\alpha}}}, \quad x \geq 0, \quad k,\alpha,\eta > 0 \tag{3.10}$$

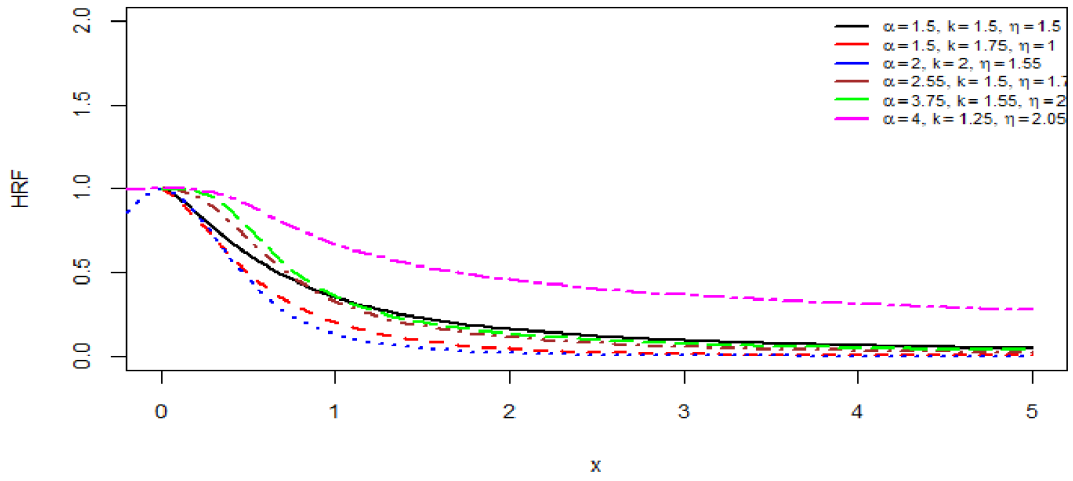

Figure 3.3: The survival curves for the GLL distribution

3.2.5 The Cummulative Distribution Function

The cumulative distribution function (CDF) also known as the lifetime distribution function of the GLL distribution is of the form below and the plots of the cdf are shown in Figure 3.4

$$F(x;\boldsymbol{\theta}) = \frac{[1+(\eta x)^{\alpha}]^{\frac{k^{\alpha}}{\eta^{\alpha}}} - 1}{[1+(\eta x)^{\alpha}]^{\frac{k^{\alpha}}{\eta^{\alpha}}}}, x \geq 0, \quad k,\alpha,\eta > 0 \tag{3.11}$$

where $k > 0, \beta > 0, \eta > 0$ are parameters and $\boldsymbol{\theta} = (k,\alpha,\eta)'$.

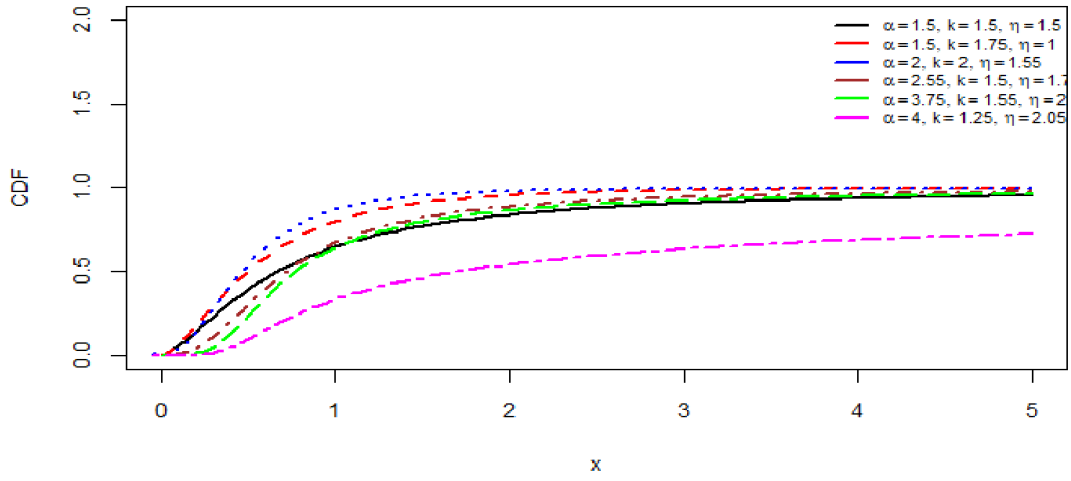

Figure 3.4: The cdf plots for the GLL distribution

3.2.6 The Retro Hazard Function

The reversed hazard rate (also known as the retro hazard) is defined as the ratio of pdf to the corresponding cdf. The retro hazard is written as follows:

$$\lambda_r(x;\boldsymbol{\theta}) = \frac{f(x;\boldsymbol{\theta})}{F(x;\boldsymbol{\theta})}, \tag{3.12}$$

Reversed hazard rate function plays an important role in the analysis of censored data and in the estimation of the survival function. The following equation gives us the basic relatio between hazard rate function and the reversed hazard rate function.

$$\lambda_r(x;\boldsymbol{\theta}) = \frac{h(x;\boldsymbol{\theta})S(x;\boldsymbol{\theta})}{1 - S(x;\boldsymbol{\theta})} \tag{3.13}$$

The applications of hazard rate function in survival analysis are well known. Recently the reversed hazard rate function has gained popularity among applied statisticians, for information see (Gupta and Wu, 2001; Cox and Oakes, 2018; Legrand, 2021). Block et al. (1998) showed that the hazard rate function play essential role

in the analysis of right-censored data while the retro hazard plays an essential role in the analysis of left-censored data. The reversed hazard rate function of the GLL distribution takes the form

$$r(x;\boldsymbol{\theta}) = \frac{f(x;\boldsymbol{\theta})}{F(x;\boldsymbol{\theta})} = \frac{[1+(\eta x)^{\alpha}]^{\frac{k^{\alpha}}{\eta^{\alpha}}+1}}{[1+(\eta x)^{\alpha}]^{\frac{k^{\alpha}}{\eta^{\alpha}}}} \tag{3.14}$$

Simplifying gives;

$$r(x;\boldsymbol{\theta}) = \frac{\alpha k(kx)^{\alpha-1}}{[1+(\eta x)^{\alpha}]^{\frac{k^{\alpha}}{\eta^{\alpha}+1}} - [1+(\eta x)^{\alpha}]}, \quad x \geq 0, \quad k,\alpha,\eta > 0 \tag{3.15}$$

The reversed hazard rate plots are shown in Figure 3.5

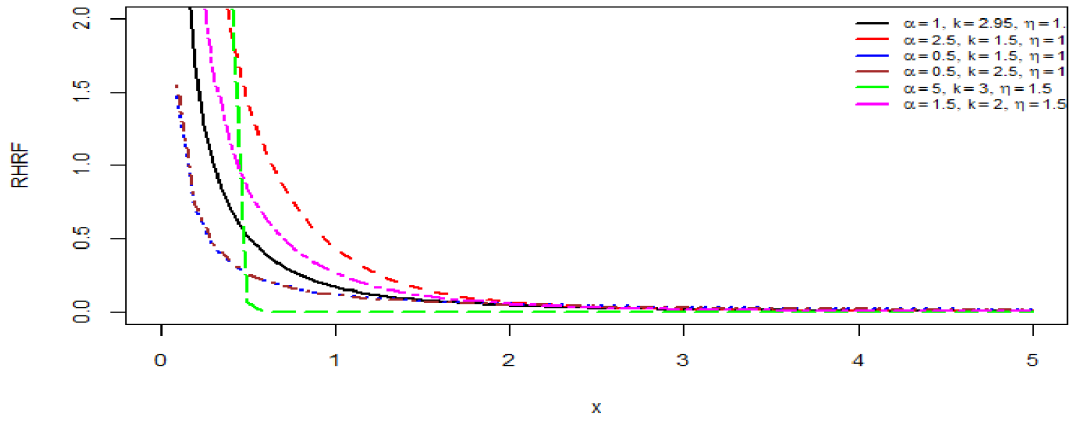

Figure 3.5: The rerto hazard plots for the GLL distribution

3.2.7 The Cummulative Hazard Function

The cumulative hazard function of the GLL distribution takes the form:

$$H(x;\boldsymbol{\theta}) = -\log S(x;\boldsymbol{\theta}) = \int_0^x h(x;\boldsymbol{\theta})dx \tag{3.16}$$

Simplifying gives;

$$H(x;\boldsymbol{\theta}) = \frac{k^{\alpha}}{\eta^{\alpha}} \log\left[1 + (\eta x)^{\alpha}\right], \quad x \geq 0, \quad k, \alpha, \eta > 0 \tag{3.17}$$

where $k > 0, \alpha > 0, \eta > 0$ are parameters and $\boldsymbol{\theta} = (k, \alpha, \eta)'$.

3.2.8 The Hazard Rate Average (HRA) function

The HRA function of X is expressed as;

$$HRA(x;\boldsymbol{\theta}) = \frac{H(x;\boldsymbol{\theta})}{x} = \frac{\int_0^x h(x;\boldsymbol{\theta})dx}{x}, x > 0 \tag{3.18}$$

Where $H(x;\boldsymbol{\theta})$ is the cumulative hazard function. An analysis of $HRA(x;\boldsymbol{\theta})$ on t enables us to find increasing hazard rate average and decreasing hazard rate average.

3.3 Some Mathematical Properties of the Proposed Distribution

In this section, we present some mathematical properties of the GLL distribution. The functions that we discussed in section (2) are not the only ways that we can define the GLL distribution, but there are other mathematical functions that we can use to describe the lifetime distributions of a random variable X. These include; quantile function and its related results, moments and its related properties, rth central moments, residual life and reversed residual life functions, and other mathematical properties.

3.3.1 The Quantile Function and Related Results

The quantile function (which is the inverse of the CDF) is crucial in statistical and quantitative data analysis. A probability distribution can be defined in terms of

either the quantile function or the cumulative distribution function (Midhu et al., 2013). The quantiles of the proposed distribution with various parameter values are given in Table 3.2.

Theorem 3.3.1 *If $T \sim GLL(k, \alpha, \eta)$, then the quantile function, lower quartile, median, and the upper quartile of the GLL distribution, respectively are given by*

$$X_{q=}F^{-1}(q; k, \alpha, \eta) = \frac{\left\{\left[\frac{1}{1-p}\right]^{\frac{\eta^\alpha}{k^\alpha}} - 1\right\}^{\frac{1}{\alpha}}}{\eta}, \tag{3.19}$$

$$X_{q_1} = \frac{\left\{\left[\frac{4}{3}\right]^{\frac{\eta^\alpha}{k^\alpha}} - 1\right\}^{\frac{1}{\alpha}}}{\eta}, \tag{3.20}$$

$$X_{q_2} = \text{ Median } = \frac{\left\{2^{\frac{\eta^\alpha}{k^\alpha}} - 1\right\}^{\frac{1}{\alpha}}}{\eta}. \tag{3.21}$$

$$X_{q_3} = \frac{\left\{4^{\frac{\eta^\alpha}{k^\alpha}} - 1\right\}^{\frac{1}{\alpha}}}{\eta}. \tag{3.22}$$

Proof 3.3.1 *The quantile function of GLL distribution is derived by finding the value of Q for which;*

$$1 - [1 + (\eta x)^\alpha]^{-\frac{k^\alpha}{\eta^\alpha}} = p$$

$$X_{q=}F^{-1}(q; k, \alpha, \eta) = [1 + (\eta q)^\alpha]^{-\frac{k^\alpha}{\eta^\alpha}} = 1 - p$$

$$= \frac{1}{[1 + (\eta q)^\alpha]^{\frac{k^\alpha}{\eta^\alpha}}} = 1 - p$$

$$= [1 + (\eta q)^\alpha]^{\frac{k^\alpha}{\eta^\alpha}} = \frac{1}{1-p}$$

$$= 1 + (\eta q)^\alpha = \left(\frac{1}{1-p}\right)^{\frac{\eta^\alpha}{k^\alpha}}$$

$$= (\eta q)^\alpha = \left(\frac{1}{1-p}\right)^{\frac{\eta^\alpha}{k^\alpha}} - 1$$

Table 3.2: Quantiles of the proposed distribution for different parameter values.

Quantiles	(κ, α, η)				
	$(0.5, 0.5, 0.5)$	$(5.0, 1.5, 1.5)$	$(4.0, 4.0, 2.5)$	$(3.0, 2.0, 3.0)$	$(5.0, 3.0, 2.0)$
0.1	0.0247	0.0449	0.1427	0.1111	0.0945
0.2	0.1250	0.0745	0.1725	0.1667	0.1216
0.3	0.3673	0.1026	0.1945	0.2182	0.1424
0.4	0.8889	0.1314	0.2134	0.2722	0.1608
0.5	1.9999	0.1627	0.2312	0.3333	0.1783
0.6	4.4999	0.1985	0.2489	0.4082	0.1961
0.7	10.8889	0.2421	0.2681	0.4082	0.2155
0.8	32.0000	0.3006	0.2906	0.6667	0.2385
0.9	162.0000	0.3972	0.3222	1.0000	0.2707

$$= \eta q = \left\{ \left[\frac{1}{1-p} \right]^{\frac{\eta^\alpha}{k^\alpha}} - 1 \right\}^{\frac{1}{\alpha}}$$

$$\therefore q = \frac{\left\{ \left[\frac{1}{1-p} \right]^{\frac{\eta^\alpha}{k^\alpha}} - 1 \right\}^{\frac{1}{\alpha}}}{\eta}.$$

where; $p \in [0, 1).k > 0, \alpha > 0, \eta > 0$. *Hence the proof.*

Similarly, we can prove (3.20), (3.21) and (3.22) by applying the following values;

Lower quartile $= 1/4$, median $= 2/4 = 1/2$, and the upper quartile $= 3/4$.

Lower quartile is

$$X_{q_1} = \frac{\left\{ \left[\frac{4}{3} \right]^{\frac{\eta^\alpha}{k^\alpha}} - 1 \right\}^{\frac{1}{\alpha}}}{\eta}$$

Medium is

$$X_{q_2} = \text{ Median } = \frac{\left\{ 2^{\frac{\eta^\alpha}{k^\alpha}} - 1 \right\}^{\frac{1}{\alpha}}}{\eta}$$

Upper quartile is

$$X_{q_3} = \frac{\left\{ 44^{\frac{\eta^\alpha}{k^\alpha}} - 1 \right\}^{\frac{1}{\alpha}}}{\eta}$$

3.3.2 Skewness and Kurtosis

The following relationship defines the mathematical form of the Galton Skewness and Moors Kurtosis of the GLL model with three parameters:

$$S_K = \frac{Q\left(\frac{3}{4}\right) + Q\left(\frac{1}{4}\right) - 2Q\left(\frac{2}{4}\right)}{Q\left(\frac{3}{4}\right) - Q\left(\frac{1}{4}\right)}, \tag{3.23}$$

$$K_M = \frac{Q\left(\frac{7}{8}\right) + Q\left(\frac{3}{8}\right) - Q\left(\frac{5}{8}\right) - Q\left(\frac{1}{8}\right)}{Q\left(\frac{6}{8}\right) - Q\left(\frac{2}{8}\right)} \tag{3.24}$$

where Q describes different quartile values.

The above equations can be determined as functions of the GLL quantile function. The advantages of these measures are that they are less sensitive in the presence of outliers and that they exist even when the distribution is lacking moments.

3.3.3 The Random Deviate Generation Function

Let U be a random variable with a uniform distribution $(0, 1)$ and an inverse cdf, $F(.)$, Then any sample drawn from $F^{-1}(u)$ is assumed to have been drawn form $F(.)$. As a result, using GLL(k, α, η), the random deviate can be generated as follows:

$$x = \frac{\left\{\left[\frac{1}{1-u} - 1\right]^{\frac{\lambda^\alpha}{k^\alpha}}\right\}^{\frac{1}{\alpha}}}{\lambda}, 0 < u < 1 \tag{3.25}$$

where u follows $U(0, 1)$ distribution.

3.3.4 The r^{th} Moments and Related Results.

Numerous important characteristics and properties of a probability distribution such as mean, variance, kurtosis, and skewness can be obtained from its moments. Moments are extremely important and play a central role in statistical analysis, especially in applications. The important moment functions, such as the moments,

r^{th} moment, r^{th} central moment, mean, variance, skewness, and kurtosis of the proposed distribution, are presented.

Theorem 3.3.2 *If $T \sim GLL(k, \alpha, \eta)$, then the r^{th} power, negative moments, and logarithmic moments are given, respectively, by*

$$E\left(T^r\right) = \frac{k^\alpha}{\eta^{\alpha+r}} \frac{\Gamma\left((k^\alpha/\eta^\alpha) - (r/\alpha)\right)\Gamma((r/\alpha)+1)}{\Gamma\left((k^\alpha/\eta^\alpha)+1\right)}, \quad \text{for } \frac{\alpha k^\alpha}{\eta^\alpha} > r, \tag{3.26}$$

$$E\left(T^{-r}\right) = \frac{\lambda^{\alpha+r}}{k^\alpha} \frac{\Gamma\left((k^\alpha/\eta^\alpha)+1\right)}{\Gamma\left((k^\alpha/\eta^\alpha) - (r/\alpha)\right)\Gamma((r/\alpha)+1)}. \tag{3.27}$$

Proof 3.3.2 *We have*

$$\begin{aligned} E\left(T^r\right) &= \int_0^\infty t^r f(t; k, \alpha, \eta)\mathrm{d}t \\ &= \int_0^\infty t^r \frac{\alpha k (kt)^{\alpha-1}}{\left[1+(\eta t)^\alpha\right]^{\left(k^\alpha/\eta^\beta\right)+1}}\ \mathrm{d}t \\ &= \frac{\alpha k}{\Gamma\left((k^\alpha/\eta^\alpha)+1\right)} \int_0^\infty t^r \frac{(kt)^{\alpha-1}}{1+(\eta t)^\alpha}\mathrm{d}t \\ &= \frac{k^\alpha}{\eta^{\alpha+r}} \frac{\Gamma\left((k^\alpha/\eta^\alpha) - (r/\alpha)\right)\Gamma((r/\alpha)+1)}{\Gamma\left((k^\alpha/\eta^\alpha)+1\right)}, \quad \text{for } \frac{\alpha k^\alpha}{\eta^\alpha} > r. \end{aligned}$$

Similarly, we can prove (3.27)

3.3.4.1 Mean and Variance

Corollary 3.3.1 *If $T \sim GLL(k, \alpha, \eta)$, then the mean and variance are given, respectively, as follows:*

The mean of the GLL distribution is

$$\mu = E(T) = \frac{k^\alpha}{\eta^\alpha} \frac{\Gamma\left((k^\alpha/\eta^\alpha) - (1/\alpha)\right)\Gamma((1/\alpha)+1)}{\Gamma\left((k^\alpha/\eta^\alpha)+1\right)} \tag{3.28}$$

This is provided that $(\alpha k^\alpha/y^\alpha) > 1$.

The variance of the GLL distribution is

$$\begin{aligned}\sigma^2 = V(T) &= E\left(T^2\right) - (E(T))^2 \\ &= \frac{k^\alpha}{\eta^{\alpha+2}} \frac{\Gamma\left((k^\alpha/\eta^\alpha) - (2/\alpha)\right)\Gamma((2/\alpha)+1)}{\Gamma\left((k^\alpha/\eta^\alpha)+1\right)} - \left(\frac{k^\alpha}{\eta^\alpha} \frac{\Gamma\left((k^\alpha/\eta^\alpha) - (1/\alpha)\right)\Gamma((1/\alpha)+1)}{\Gamma\left((k^\alpha/\eta^\alpha)+1\right)}\right)^2.\end{aligned} \tag{3.29}$$

This is provided that $(\alpha k^\alpha/y^\alpha) > 2$.

3.3.4.2 The r^{th} Central Moments

Corollary 3.3.2 *If* $T \sim GLL(k, \alpha, \eta)$*, then the cumulants of the first, second, and* th *central moments, are given, respectively, by:*

$$\begin{aligned}c_1 =& \mu_1' = E(T) = \frac{k^\alpha}{\eta^\alpha} \frac{\Gamma\left((k^\alpha/\eta^\alpha) - (1/\alpha)\right)\Gamma((1/\alpha)+1)}{\Gamma\left((k^\alpha/\eta^\alpha)+1\right)} \\ c_2 =& \mu_2' - \mu_1^{\hat{k}} = E\left(T^2\right) - (E(T))^2 \\ &= \frac{k^\alpha}{\eta^{\alpha+2}} \frac{\Gamma\left((k^\alpha/\eta^\alpha) - (2/\alpha)\right)\Gamma((2/\alpha)+1)}{\Gamma\left((k^\alpha/\eta^\alpha)+1\right)} \\ &- \left(\frac{k^\beta}{\eta^\alpha} \frac{\Gamma\left((k^\alpha/\eta^\alpha) - (1/\alpha)\right)\Gamma((1/\alpha)+1)}{\Gamma\left((k\alpha/\eta^\alpha)+1\right)}\right)^2, \\ c_r =& \mu_r' - \sum_{n=1}^{r-1} \begin{pmatrix} r-1 \\ n-1 \end{pmatrix} c_n \mu_{r-m}' = \frac{k^\alpha}{\eta^{\alpha+r}} \frac{\Gamma\left((k^\alpha/\eta^\alpha) - (r/\alpha)\right)\Gamma((r/\alpha)+1)}{\Gamma\left((k^\alpha/\eta^\alpha)+1\right)} \\ &- \sum_{n=1}^{r-1} \begin{pmatrix} r-1 \\ n-1 \end{pmatrix} c_n \frac{k^\alpha}{\eta^{\alpha+(}(r-n)} \frac{\Gamma\left((k^\alpha/\eta^\alpha) - ((r-n)/\alpha)\right)\Gamma(((r-n)/\alpha)+1)}{\Gamma\left((k^\alpha/\eta^\alpha)+1\right)}.\end{aligned} \tag{3.30}$$

Hence, from Corollary 2, we can derive the skewness and kurtosis of the GLL distribution by computing, respectively:

$$\text{Skewness } = \frac{c_3}{\left(\sigma^2\right)^{3/2}}, \tag{3.31}$$

$$\text{Kurtosis } = \frac{c_4}{\left(\sigma^2\right)^2}. \tag{3.32}$$

3.3.5 Residual and Reverse Residual Life

The residual life has broader applications in survival analysis and risk management. The residual lifetime of the GLL random variable is calculated as follows:

$$R_{(t)}(x) = \frac{S(x+t)}{S(t)} \tag{3.33}$$

$$R_{(t)}(x) = \frac{[1+(\eta(x+t))^{\alpha}]^{-(k+n\eta^{\alpha})}}{[1+(\eta t)^{\alpha}]^{-(kk^{\alpha}(\eta\eta^{\alpha})}}. \tag{3.34}$$

In addition, the reverse residual life of the generalized log-logistic random variable can be calculated as follows:

$$\hat{R}_{\langle t\rangle}(x) = \frac{S(x-t)}{S(t)}, \tag{3.35}$$

$$\hat{R}_{\langle t\rangle}(x) = \frac{[1+(\eta(x-t))^{\alpha}]^{-(k+n\eta^{a})}}{[1+(\eta t)^{\alpha}]^{-(kk^{\alpha}\eta^{\alpha})}} \tag{3.36}$$

From Table 3.3, the GLL distribution is clearly numerically versatile in terms of means and variance. Furthermore, the values of skewness (CS) show that it can be right-skewed, nearly symmetrical, or slightly left-skewed. The kurtosis (CK) values show that the GLL distribution can be mesokurtic, leptokurtic, or platykurtic. All of these characteristics demonstrate the GLL distribution flexibility, which remains appealing for modelling purposes.

The mean and variance plots for different values of alpha and kappa parameters are shown in Figure 3.6, while the skewness and kurtosis plots are shown in Figure 3.7.

Table 3.3: First five moments, standard deviation, skewness, and kurtosis of the GLL distribution for some parameter values.

Moments	(κ, α, η)						
	(0.5,0.5,0.5)	(1.0,1.5,1.5)	(1.5,2.0,2.5)	(2.0,5.0,3.0)	(1.0,1.0,2.0)	(4.0,4.5,0.2)	(5.0,4.0,0.5)
μ_1'	0.1034	0.2065	0.2432	0.2795	0.1547	0.2281	0.1813
μ_2'	0.0567	0.1292	0.1482	0.1741	0.0893	0.0554	0.0354
μ_3'	0.0388	0.0925	0.1036	0.1204	0.0619	0.0141	0.0073
μ_4'	0.0294	0.0715	0.0787	0.0900	0.0471	0.0037	0.0016
μ_5'	0.0237	0.0581	0.0631	0.0711	0.0380	0.0010	0.0004
SD	0.2146	0.2943	0.2984	0.3098	0.2557	0.0575	0.0509
CV	2.0743	1.4250	1.2270	1.1081	1.6529	0.2521	0.2805
CS	2.3743	1.1784	0.9109	0.6100	1.6648	-0.1784	-0.0871
CK	7.8842	3.0318	2.5240	2.0238	4.6628	2.8081	2.7479

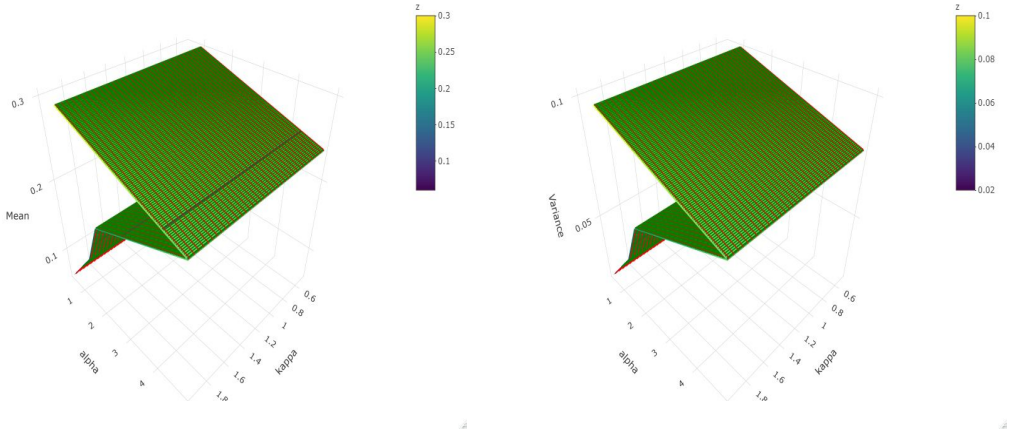

Figure 3.6: The mean and variance plots for several combinations of alpha and kappa parameters

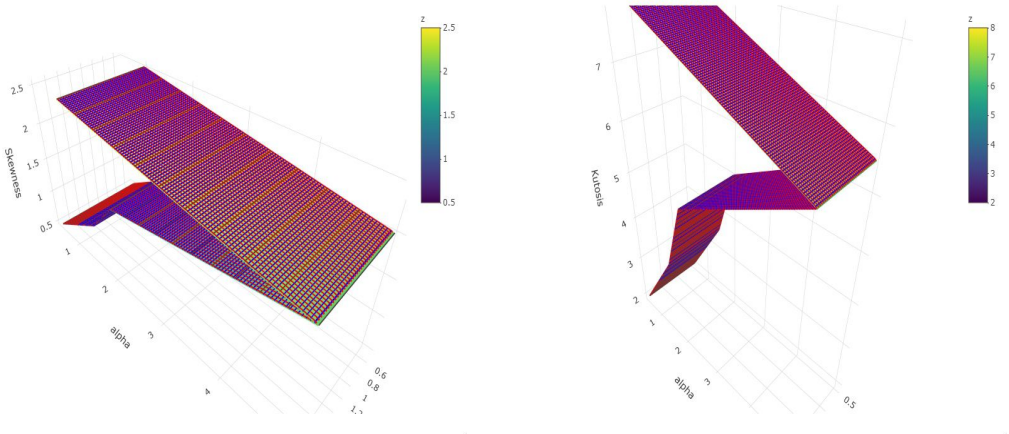

Figure 3.7: The skewness and kurtosis plots for several combinations of alpha and kappa parameters

3.4 Maximum Likelihood Estimation

In this section, the unknown parameters of the generalized log-logistic distribution based on a complete sample is estimated using maximum likelihood method. Let $X_1, X_2, \ldots, X_n$ indicate a random sample of the complete GLL data, and then the sample's likelihood function is given as:

$$\begin{aligned} L &= \prod_{i=1}^{n} f\left(x_i, \alpha, k, \eta\right) \\ L(x; \alpha, k, \eta) &= \prod_{i=1}^{n} \left[\frac{\alpha k \left(k x_i\right)^{\alpha-1}}{\left[1 + \left(\eta x_i\right)^{\alpha}\right]^{\frac{k^{\alpha}}{\eta^{\alpha+1}}}} \right. \end{aligned} \tag{3.37}$$

The log-likelihood function may be expressed as;

$$\begin{aligned} \ell = n \log(\alpha k) + (\alpha - 1) \sum_{i=1}^{n} \log\left(k x_i\right) - \sum_{i=1}^{n} \log\left[1 + \left(\eta x_i\right)^{\alpha}\right] \\ - \left(\frac{k}{\eta}\right)^{\alpha} \sum_{i=1}^{n} \log\left[1 + \left(\eta x_i\right)^{\alpha}\right]. \end{aligned} \tag{3.38}$$

By taking the first derivatives of the log-likelihood function in Equation 3.38 with respect to α, k, and η, and fixing the outcome to zero, we have:

$$\begin{aligned} \frac{\partial \ell}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n} \log\left(k x_i\right) - \sum_{i=1}^{n} \left\{ \frac{\left(\eta x_i\right)^{\alpha} \log\left(\eta x_i\right))}{\left[1 + \left(\eta x_i\right)^{\alpha}\right]} \right\} - \left(\frac{k}{\eta}\right)^{\alpha} \left(\frac{k}{\eta}\right) \sum_{i=1}^{n} \log\left[1 + \left(\eta x_i\right)^{\alpha}\right] \\ - \left(\frac{k}{\eta}\right)^{\alpha} \sum_{i=1}^{n} \left\{ \frac{\left(\eta x_i\right)^{\alpha} \log\left(\eta x_i\right))}{\left[1 + \left(\eta x_i\right)^{\alpha}\right]} \right\}. \end{aligned} \tag{3.39}$$

$$\frac{\partial \ell}{\partial k} = \frac{n}{k} + \frac{(\alpha - 1)}{k} - \left(\frac{\alpha}{k}\right) \left(\frac{k}{\eta}\right)^{\alpha} \sum_{i=1}^{n} \log\left[1 + \left(\eta x_i\right)^{\alpha}\right] \tag{3.40}$$

$$\begin{aligned} \frac{\partial \ell}{\partial \eta} = - \left(\frac{\alpha}{\eta}\right) \sum_{i=1}^{n} \left\{ \frac{\left(\eta x_i\right)^{\alpha}}{\left[1 + \left(\eta x_i\right)^{\alpha}\right]} \right\} + \left(\frac{\alpha}{k}\right) \left(\frac{k}{\eta}\right)^{\alpha} \sum_{i=1}^{n} \log\left[1 + \left(\eta x_i\right)^{\alpha}\right] - \\ \left(\frac{\alpha}{\eta}\right) \left(\frac{k}{\eta}\right)^{\alpha} \sum_{i=1}^{n} \left\{ \frac{\left(\eta x_i\right)^{\alpha}}{\left[1 + \left(\eta x_i\right)^{\alpha}\right]} \right\} \end{aligned} \tag{3.41}$$

Note that the MLE's $\hat{\alpha}, \hat{k}$, and $\hat{\eta}$ of α, k, and η, respectively can be obtained by equating the outcomes to zero and solving the system of non-linear equations numerically. In order to construct confidence intervals for the parameters, the observed information matrix $J(\theta)$ is used since the expected information matrix is complicated. The observed information matrix is given by:

$$J(\theta) = - \begin{bmatrix} \frac{\partial^2 \ell}{\partial^2 \alpha} & \frac{\partial^2 \ell}{\partial \alpha \partial k} & \frac{\partial^2 \ell}{\partial \alpha \partial \eta} \\ & \frac{\partial^2 \ell}{\partial^2 k} & \frac{\partial^2 \ell}{\partial k \partial \eta} \\ & & \frac{\partial^2 \ell}{\partial^2 \eta} \end{bmatrix}$$

Where; $\theta = (\alpha, k, \eta)'$. When the usual regularity conditions are fulfilled and that the parameters are within the interior of the parameter space, but not on the boundary. $\sqrt{n}(\cong \theta - \theta)$ converges in distribution to $N_3\left(\mathbf{0}, I^{-1}(\theta)\right)$, were $I(\theta)$ is the expected information matrix. The asymptotic behavior is still valid when $I(\theta)$ is replaced by the observed information matrix evaluated at $J(\theta)$. The asymptotic multivariate normal distribution $N_3\left(\mathbf{0}, J^{-1}(\theta)\right)$ can be used to construct an appropriate $100(1-\tau)\%$ two-sided confidence intervals for the model parameters, where τ is the significant level.

3.5 Monte Carlo Simulation Study

In this section, we assess the performance of the MLEs estimators for a finite sample of size n using a Monte Carlo simulation study. The simulation study based on the generalized log-logistic distribution is carried out to examine the average biases (ABs), the mean square errors (MSEs), the root mean square errors (RMSEs), and maximum likelihood estimates (MLEs) for the model parameters α, k, and η. The simulation experiment was carried out using a variety of simulations with varying sample sizes and parameter values. To generate random samples for the GLL, the quantile function is given in Equation (3.19). The simulation study was repeated

1500 times each with sample sizes $n = 50, 100, \ldots, 1500$ and the following parameter scenarios in a set I: $\alpha = 0.9, k = 0.5$, and $\eta = 2.5$, and set II: $\alpha = 0.8, k = 0.4$, and $\eta = 2.0$.

The MLEs of the GLL model are determined via the nlminb () R-function with the argument method = "BFGS"; see supplementary materials. For each piece of simulated data, say, $(\hat{\alpha}, \hat{k}, \hat{\eta})$ for $i = 1, 2, \ldots, 1000$, and the AB, RMSE, and CP of the parameters were computed by:

$$\text{AB} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i - \theta), \tag{3.42}$$

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i - \theta)^2, \tag{3.43}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i - \theta)^2}, \tag{3.44}$$

where $\theta = \alpha, k$ and η.

The MLE, AB, and RMSE values of the parameters α, k, and η are displayed from various sample sizes. Based on these findings, we conclude that the MLEs perform quite well in estimating the model parameters and that the estimates are fairly stable and are nearer to the true values for these sample sizes. The Table 3.4 and Figures 3.8 to 3.11 show that as the sample size increases, the MSE and RMSE decrease as expected. Furthermore, as the sample size increases, the AB decreases. In addition, the MLEs of the parameters of the model are very close to the true value. As a result, the maximum likelihood estimates and their asymptotic results can be applied to construct confidence intervals for the model parameters even for a small sample size.

Table 3.4: Monte Carlo simulation results for the GLL distribution: MLE, AB, MSEs, and RMSEs

Parameters	n	I			II		
		MLE	AB	RMSE	MLE	AB	RMSE
α	50	2.320	1.420	5.273	2.330	1.530	7.149
	100	1.281	0.381	2.386	1.097	0.297	2.263
	300	0.995	0.095	1.369	0.937	0.137	1.512
	600	0.921	0.021	0.5207	0.840	0.040	0.619
	900	0.908	0.008	0.067	0.804	0.004	0.058
	1200	0.905	0.005	0.060	0.803	0.003	0.049
	1500	0.904	0.004	0.054	0.804	0.004	0.045
k	50	1.246	0.746	2.306	1.217	0.817	2.802
	100	0.792	0.292	1.235	0.613	0.213	1.093
	300	0.571	0.071	0.665	0.463	0.063	0.467
	600	0.511	0.011	0.135	0.422	0.022	0.199
	900	0.508	0.008	0.095	0.405	0.005	0.081
	1200	0.507	0.007	0.082	0.404	0.004	0.066
	1500	0.505	0.005	0.073	0.405	0.005	0.063
η	50	3.280	0.780	3.404	3.033	1.033	3.944
	100	2.780	0.280	1.800	2.281	0.281	1.614
	300	2.612	0.112	0.904	2.056	0.056	0.806
	600	2.554	0.054	0.588	2.046	0.046	0.543
	900	2.542	0.042	0.500	2.019	0.019	0.442
	1200	2.526	0.026	0.409	2.014	0.014	0.360
	1500	2.520	0.020	0.370	2.030	0.030	0.340

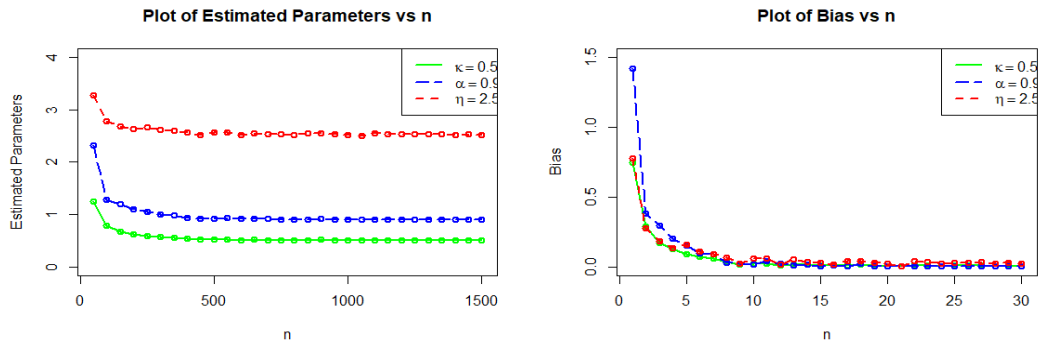

Figure 3.8: Plots for MLEs and biases of the GLL model for set I of the table.

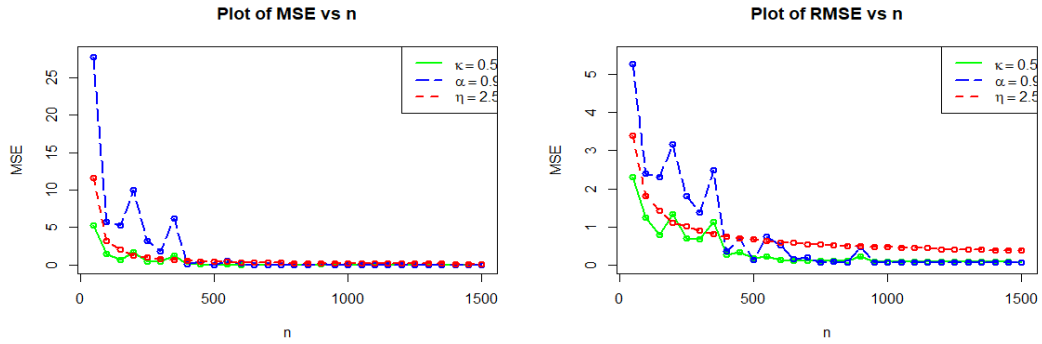

Figure 3.9: Plots for MSEs and RMSEs of the GLL distribution for the values of set I in the table

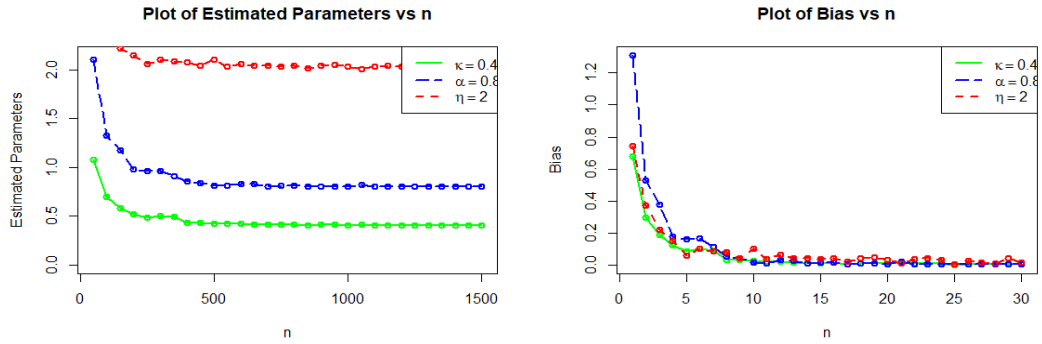

Figure 3.10: Plots for MLEs and biases of the GLL distribution for the values of set II in the table.

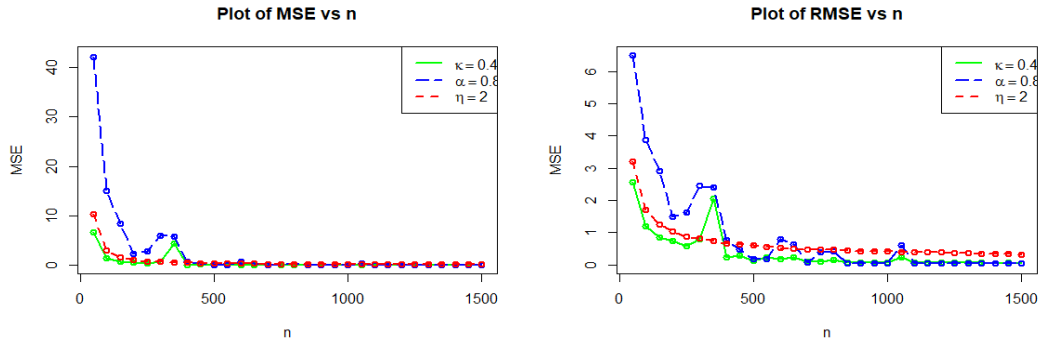

Figure 3.11: Plots for MSEs and RMSEs of the GLL distribution for the values of set II in the table.

3.6 Applications to Real-life Data

In this section, the proposed distribution is fully applied to real-world data set which is taken from literature to demonstrate the ability of the new model. We compare the proposed distribution with the other three parametric survival distributions including (gamma, log-normal, log-logistic, exponentiated Weibull, and the Weibull distribution). Also, we have compared the GLL distribution with some of its sub-models with two-parameter distribution namely; Weibull, log-logistic, and the BurrXII distributions.

The density functions of the fitted models are:

1. Weibull

$$f(t) = \alpha k (kt)^{\alpha-1} \exp\left\{-(kt)^{\alpha}\right\}, \tag{3.45}$$

2. Log-logistic

$$f(t) = \frac{\alpha k (kt)^{\alpha-1}}{\left[1 + (kt)^{\alpha}\right]^2} \tag{3.46}$$

3. Burr-XII distribution

$$f(t) = \frac{\alpha k t^{\alpha-1}}{\left[1 + t^{\alpha}\right]^{-k-1}}, \tag{3.47}$$

4. Exponentiated Weibull

$$f(t) = \alpha k \lambda (kt)^{\alpha-1} \left(1 - \exp\left\{-(kt)^{\alpha}\right\}\right)^{\lambda-1} \exp\left\{-(kt)^{\alpha}\right\}, \tag{3.48}$$

5. Three parameter log-logistic (or shifted log-logistic distribution)

$$f(t) = \frac{\frac{\alpha}{\beta}\left(\frac{t-\mu}{\beta}\right)^{\alpha-1}}{\left[1 + \left(\frac{t-\mu}{k\beta}\right)^{\alpha}\right]^2} \tag{3.49}$$

6. Three parameter log-normal

$$f(t) = \frac{\alpha}{\beta}\left(\frac{t-\mu}{\beta}\right)^{\alpha-1} \exp\left\{-\left(\frac{t-\mu}{k\beta}\right)^{\alpha}\right\}, \tag{3.50}$$

7. Three parameter Weibull

$$f(t) = \frac{\exp\left\{-\frac{1}{2}\left(\frac{\log(t-\mu)-\alpha}{\beta}\right)^2\right\}}{\sqrt{2\pi}\beta(x-\mu)} \tag{3.51}$$

8. Three parameter Gamma

$$f(t) = \frac{(t-\mu)^{\alpha-1}\exp -\left(\frac{t-\mu}{\beta}\right)}{\beta^{\alpha}\Gamma(\alpha)} \tag{3.52}$$

Where; $t > \mu$. Certain analytical measures are taken into account in order to determine which distribution best fits the applied data. These analytical measures include four discrimination measures: AIC (Akaike Information Criterion), CAIC (Consistent Akaike Information Criterion), BIC (Bayesian Information Criterion), and HQIC (Hannan Quin Information Criterion). In addition, there are two goodness-of-fit tests: Anderson-Darlin (A^*), and Cramervon Mises (W^*).
The AIC is

$$AIC = 2k - 2l \tag{3.53}$$

The BIC is

$$BIC = k\ln(n) - 2l \tag{3.54}$$

The CAIC is

$$CAIC = \frac{2nk}{n-k-1} - 2l \tag{3.55}$$

The HQIC is

$$HQIC = 2k\ln(\ln(n)) - 2l, \tag{3.56}$$

where; l represents the log-likelihood function evaluated as the MLEs, n the sample size and k the number of model parameters The goodness-of-fit measures under consideration are as follows:

The Anderson-Darlin (A^*) test statistic is given by

$$A^* = -n - \frac{1}{n}\sum_{i=1}^{n}(2l-1) \times [\ln G(X_i) + \ln\{1 - G(X_{n-i+1})\}] \tag{3.57}$$

The Cramer-von Mises (W*) test statistics is given by

$$W^* = \frac{1}{12n} + \sum_{i=1}^{n}\left[\frac{2i-1}{2n} + G(X_i)\right]^2 \tag{3.58}$$

Where; x_i is the ith observation in the sample and n is the sample size, x_i calculated when the data is sorted in ascending order.

The best model is the one with the lowest AIC, BIC, CAIC, and HQIC, also the A*, W*, and K $-$ S tests. Moreover, the best model is also chosen as the one having the highest value of the log-likelihood function, and p-values for the K $-$ S statistics. are also used to compare the competitive models.

3.6.1 Likelihood Ratio Test for Sub-Models

The GLL distribution has five sub-models, namely; log-logistic distribution, Weibull distribution, Burr-XII distribution, exponential distribution, and the standard log-logistic distribution. Hence, we have employed the likelihood ratio criterion to test the following hypotheses:

1. $H_0 : \eta^\alpha \to 0$, that is the sample is from Weibull distribution. $H_1 : \eta^\alpha \underset{not}{\longrightarrow} 0$, that is the sample is GLL

2. $H_0 : \eta = k$, that is the sample is from log-logistic distribution. $H_1 : \eta \neq k$, that is the sample is GLL

3. $H_0 : \eta = 1$, that is the sample is from Burr XII distribution. $H_1 : \eta \neq 1$, that is the sample is GLL

4. $H_0 : \eta = k = 1$, that is the sample is from the Standard log-logistic distribution. $H_1 : \eta \neq 1, k \neq 1$, that is the sample is GLL

5. $H_0 : \eta = 0 \& \alpha = 1$, that is the sample is from an exponential distribution. $H_1 : \eta \neq 0 \& \alpha \neq 1$, that is the sample is GLL

The likelihood ratio test (LRT) is given by:

$$LR = -2 \ln \frac{L(\hat{\theta}*; x)}{L(\hat{\theta}; x)}, \tag{3.59}$$

Where; $\hat{\theta}^*$ is the restricted Maximum likelihood estimates under the null hypothesis H_0 and $\hat{\theta}$ is the unrestricted Maximum likelihood estimates under the alternative hypothesis H_1. Under the null hypothesis, the LRT follows Chi-square distribution with degrees of freedom (df) ($df_{alt} - df_{\text{null}}$). If the p-value is less than 0.05 the null hypothesis is rejected.

3.6.2 An Application to Bladder Cancer Data Set

The following real-world data set is used to demonstrate the proposed methodology. The data in Table 3.5 below show the remission times (in months) of a sample of 128 bladder cancer patients. The data set is available in (Lee and Wang, 2003). The descriptive statistics for the data set are shown in Table 3.6 and the likelihood ratio test statistics for the data set are given in Table 3.7.

Table 3.6: Descriptive statistics of the Bladder cancer data set

Mean	Median	Mode	Variance	Skewness	Kurtosis	Minimum	Maximum
9.365	6.395	5.00	110.435	3.286	15.481	0.08	79.05

Table 3.5: The remission times (in months) of a sample of 128 bladder cancer patients

3.88, 5.32, 7.39, 10.34, 14.83, 34.26, 0.90, 2.69, 4.18, 5.34, 7.59, 10.66, 15.96, 36.66, 1.05,2.69, 4.23, 5.41, 7.62, 10.75, 16.62, 43.01, 1.19, 2.75, 4.26, 5.41, 7.63, 17.12, 46.12, 1.26, 2.83, 4.33, 5.49, 7.66, 11.25, 17.14, 79.05, 1.35, 2.87, 5.62, 7.87, 11.64, 17.36, 1.40, 3.02, 4.34, 5.71, 7.93, 0.08, 2.09, 3.48, 4.87, 6.94, 8.66, 13.11, 23.63, 0.20, 2.23, 3.5, 4.98, 6.97, 9.02, 13.29, 0.40, 2.26, 3.57, 5.06, 7.09, 9.22, 13.80, 25.74, 0.50, 2.46, 3.64, 5.09, 7.26, 9.47, 14.24, 25.82, 0.51, 2.54, 3.70, 5.17, 7.28, 9.74, 14.76, 26.31, 0.81, 2.62, 3.82, 5.32, 7.32, 10.06, 14.77, 32.15, 2.64, 11.79, 18.10, 1.46, 4.40, 5.85, 8.26, 11.98, 19.13, 1.76, 3.25, 4.50, 6.25, 8.37,12.02, 2.02, 3.31, 4.51, 6.54, 8.53, 12.03, 20.28, 2.02, 3.36, 6.76, 12.07, 21.73, 2.00, 3.36, 6.93, 8.65, 12.63, 22.69 .

Table 3.7: Likelihood ratio test statistic for bladder cancer data set

Distribution	Hypothesis	LRT	p values
W2	$H_0 : \eta^* \longrightarrow 0$ vs $H_1 : H_0$ is false	8.676	0.003
L.2	$H_0 : \eta^\alpha = k$ vs $H_1 : H_0$ is false	10.819	0.001
Burr XII	$H_0 : k\lambda^{-(1/\alpha)}, \lambda > 0$ vs $H_1 : H_0$ is false	87.472	< 0.001
Ex	$H_0 : \eta = 08\alpha = 1$ vs $H_1 : H_0$ is false	9.182	0.010
Standard LL	$H_0 : \eta = k = 1$ vs $H_1 : H_0$ is false	190.150	< 0.001

For data set, the asymptotic variance-covariance matrix for the estimated GLL parameters is given by

$$J^{-1} = \begin{bmatrix} 3.0929 \times 10^{-4} & 1.7255 \times 10^{-3} & 5.8513 \times 10^{-4} \\ 1.7255 \times 10^{-3} & 3.1612 \times 10^{-2} & 5.9347 \times 10^{-3} \\ 5.8513 \times 10^{-4} & 5.9347 \times 10^{-3} & 1.5958 \times 10^{-3} \end{bmatrix}$$

The information criterion values in Table 3.8 and the goodness-of-fit tests in Table 3.9 both demonstrate the superiority of the proposed model over the other competing models. The estimated pdf and CDF of the proposed distribution corresponding to the real-world data set are shown in Figure 3.12 and the Kaplan-Meier and PP plots for the proposed distribution are shown in Figure 3.13.

Table 3.8: Information criterion for the bladder cancer data set.

Dist	AIC	BIC	CAIC	HQIC
GLL	**825.564**	**834.120**	**825.756**	**829.040**
LN3	826.723	835.279	826.916	830.199
LL2	826.937	835.641	827.033	829.254
ExpW	827.393	835.949	827.586	830.869
LL3	827.458	836.014	827.651	830.934
G3	831.955	840.511	832.148	835.431
W2	832.163	837.868	832.259	834.481
W3	832.665	841.221	832.858	836.141
Burr XII	910.959	916.663	911.055	913.276

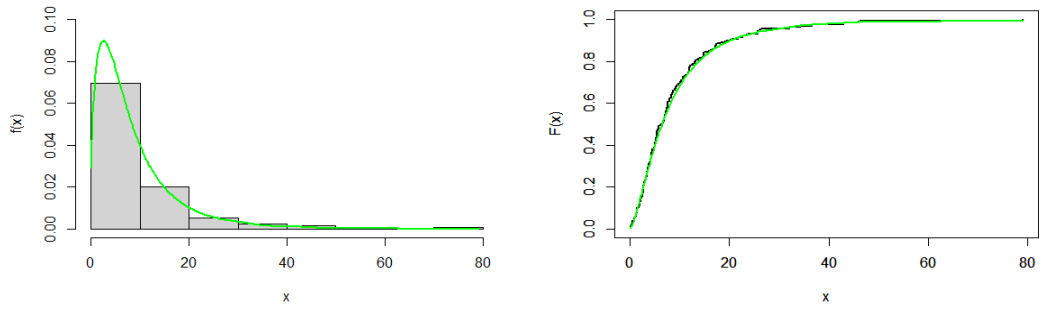

Figure 3.12: Estimated pdf and CDF of the GLL distribution corresponding to bladder data set.

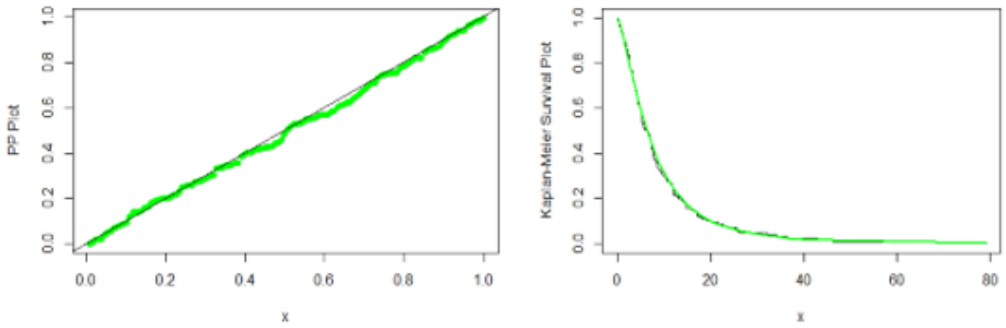

Figure 3.13: PP and Kaplan–Meier plots of the GLL distribution corresponding to bladder data set.

Table 3.9: MLE estimators of the model parameters, the log-likelihood, and goodness-of-fit statistics for data set I.

Distributions	Estimates (SEs)	ℓ	W^*	A^*	K – S (p value)
GLL (α, $k\eta$)	$\alpha = 1.410$ (0.174) $k = 0.134$ (0.017) $\eta = 0.077$ (0.038)	**–409.78**	**0.019**	**0.128**	**0.034** **(0.999)**
ExpW (α, $k\lambda$)	$\alpha = 0.275$ (0.146) $k = 0.676$ (0.136) $\lambda = 2.636$ (1.161)	–410.70	0.045	0.291	0.044 (0.967)
LL3 (α, β, γ)	$\alpha = 0.535$ (0.061) $\beta = 1.863$ (0.106) $\mu = -0.293$ (0.358)	–410.73	0.019	0.135	0.038 (0.993)
LN3 (α, β, γ)	$\alpha = 0.877$ (0.090) $\beta = 1.925$ (0.111) $\mu = -0.623$ (0.372)	–410.36	0.017	0.115	0.029 (0.998)
G3 (α, β, γ)	$\alpha = 1.098$ (0.134) $\beta = 8.424$ (1.238) $\mu = 0.075$ (0.018)	–412.98	0.125	0.778	0.067 (0.618)
W3 (α, β, γ)	$\alpha = 1.031$ (0.072) $\beta = 9.743$ (0.908) $\mu = 0.077$ (0.013)	–413.33	0.134	0.839	0.080 (0.387)
W2 (α, β)	$\alpha = 1.049$ (0.068) $k = 9.576$ (0.854)	–414.08	0.131	0.784	0.071 (0.545)
BXII (α, β)	$\alpha = 2.342$ (0.356) $k = 0.233$ (0.040)	–453.48	0.752	4.564	0.251 (<0.005)
LL2 (α, β)	$\alpha = 0.578$ (0.043) $k = 1.805$ (0.088)	–411.47	0.043	0.310	0.041 (0.984)

3.6.2.1 TTT Plot

The TTT and box plots of the data set are presented in Figure 3.14. These plots indicate that the empirical hazard rate function of the 1st data set is a reverse bathtub shape.

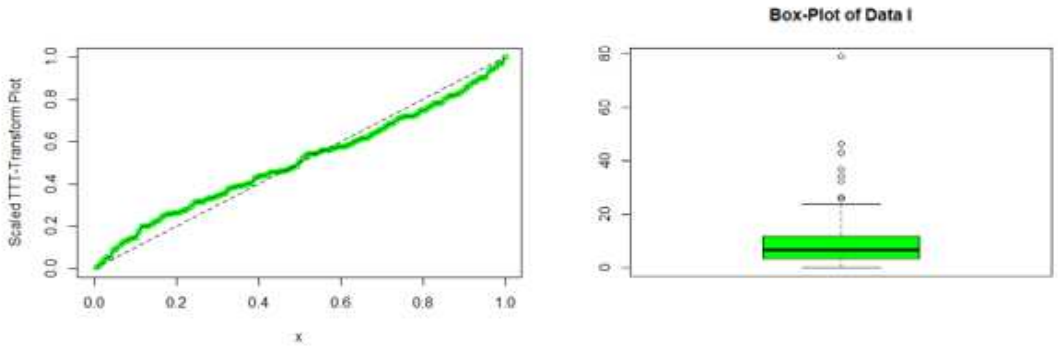

Figure 3.14: TTT and Box plots for the bladder cancer data set

The estimated fitted pdfs and CDFs of data set I for the competitive models are shown in Figure 3.15.

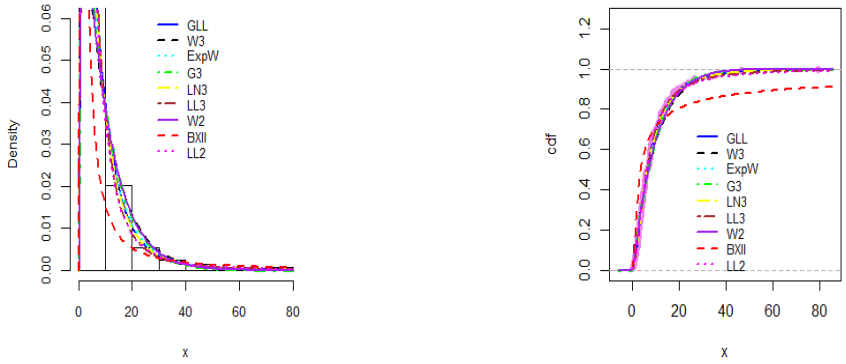

Figure 3.15: Estimated density and cumulative distribution functions for the competitive models

3.7 Bayesian Model Formulation

Given a set of data $x = (x_1, x_2, \ldots, x_n)$ from GLL (α, k, η), the likelihood function of the model is given by

$$L(\alpha, k, \eta \mid x) = (\alpha k)^n \prod_{i=1}^{n} (kx_i)^{\alpha=1} \prod_{i=1}^{n} [1 + (\eta x_i)^{\alpha}]^{-((k^n/\eta^n)+1)} . \tag{3.60}$$

The Bayesian model is built by specifying the prior distribution for the model parameters α, k and η and then multiplying with the likelihood function $L(\alpha, k, \eta \mid x)$ for the given data $x = (x_1, x_2, \ldots, x_n)$ to obtain the posterior distribution function using the Bayes theorem. The prior distribution of α, k and η is denoted as $p(\alpha, k, \eta)$. The joint posterior is

$$p(\alpha, k, \eta \mid x) \propto L(\alpha, k, \eta \mid x) p(\alpha, k, \eta). \tag{3.61}$$

3.7.1 Prior Distribution.

We assumed independent noninformative gamma priors for the parameters of the proposed model in this study due to the flexibility of gamma distributions in accom-

modating many possible shapes for the types of parameters involved in the proposed distribution. Furthermore, they enable efficient posterior calculations and the recovery of the noninformative distribution for each parameter. Many research papers in the literature consider taking these priors into account (see (Alvares et al., 2021; Lázaro Hervás, 2018; Danish and Arshad, 2017; Danish, 2016)).

For the model parameters, we assume independent gamma priors: $\alpha \sim G(a_1, b_1), k \sim G(a_2, b_2)$, and $\eta \sim G(a_3, b_3)$.

$$\begin{aligned} p(\alpha) &= \frac{b_1^{a_1}}{\Gamma(a_1)}\alpha^{a_1-1}\exp(-b_1\alpha), \quad \alpha > 0, a_1 > 0, b_1 > 0, \\ p(k) &= \frac{b_2^{a_2}}{\Gamma(a_2)}k^{a_2-1}\exp(-b_2 k), \quad \alpha > 0, a_2 > 0, b_2 > 0, \\ p(\eta) &= \frac{b_3^{a_3}}{\Gamma(a_3)}\eta^{a_3-1}\exp(-b_3\eta), \quad \eta > 0, a_3 > 0, b_3 > 0. \end{aligned} \tag{3.62}$$

Hence, we have

$$p(\alpha, k, \eta) = p(\alpha)p(k)p(\eta). \tag{3.63}$$

3.7.2 Posterior Distribution

The posterior expression can be obtained, up to proportionality, by multiplying the likelihood by the prior, and this can be written as:

$$p(\alpha, k, \eta \mid x) \propto \alpha^{a_1+n-1}k^{a_2+n-1}\eta^{a_3+n-1}e^{-(b_1\alpha+b_2k+b_3\eta)}L_1, \tag{3.64}$$

where

$$L_1 = (\alpha k)^n \prod_{i=1}^{n}(kx_i)^{\alpha-1}\prod_{i=1}^{n}[1+(\eta x_i)^{\alpha}]^{-((k^{\alpha}/n^n)+1)}. \tag{3.65}$$

The posterior is complicated, and there are no closedform inferences. As a result, we, propose using McMC techniques to simulate samples from the posterior, allowing for simple sample-based inferences.

3.8 Bayesian Analysis

In this work, we assumed the independent gamma priors for $\alpha \sim G(a_1, b_1)$, $\kappa \sim G(a_2, b_2)$, and $\eta \sim G(a_3, b_3)$ with hyperparameter values $(a_1 = b_1 = a_2 = b_2 = a_3 = b_3 = 1.0)$.

3.8.1 Convergence Diagnostics

The proposed model is built with the goal of calculating Bayesian estimates for GLL parameters using the McMC method. Due to the Ergodic property of the Markov chain, all inferences are based on the assumption that it will converge. Hence, the McMC convergence diagnostic is crucial. If the simulated sample gives an acceptable approximation for the posterior density, the inferences are correct. Several convergence diagnostic analyses are used to determine whether the chains have converged, including the following.

3.8.1.1 Geweke's Convergence Diagnostic

The Geweke's diagnostic plot for the model parameters are shown Figure 3.16 which indicates that the convergence was achieved

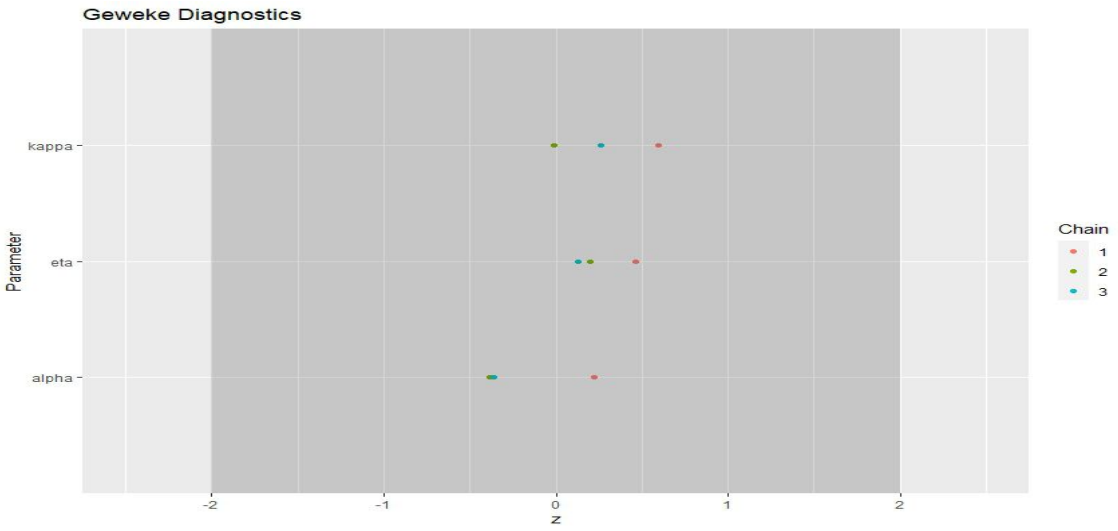

Figure 3.16: Geweke's diagnostic plot for alpha, eta, and kappa parameters

3.8.1.2 Autocorrelation Diagnostics

The autocorrelation plot for the parameters is shown in Figure 3.17.

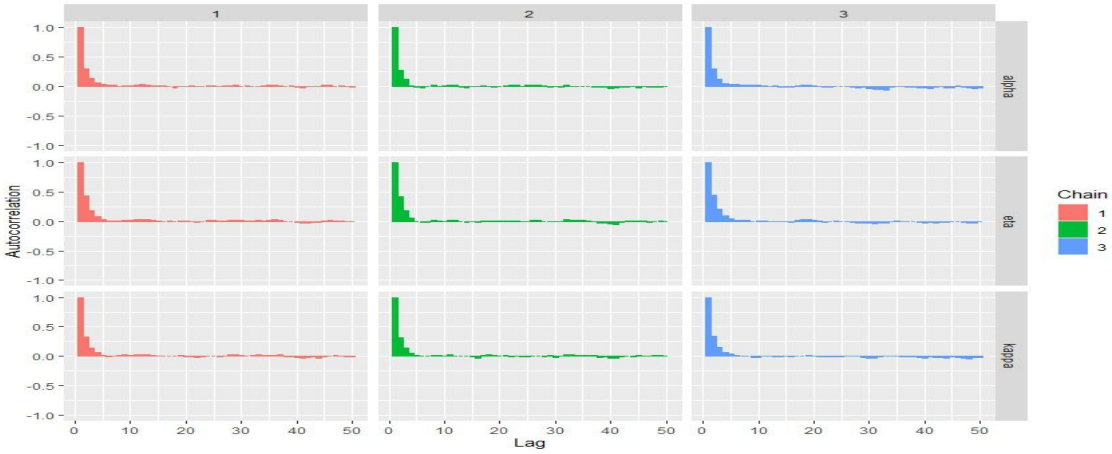

Figure 3.17: Autocorrelation plot for the alpha, eta, and kappa parameters

3.8.1.3 Brooks-Gelman-Rubin (BGR) Convergence Diagnostic

The fact that the lines for all of the parameters are close to 1 indicates convergence from BGR plots as shown in Figure 3.18.

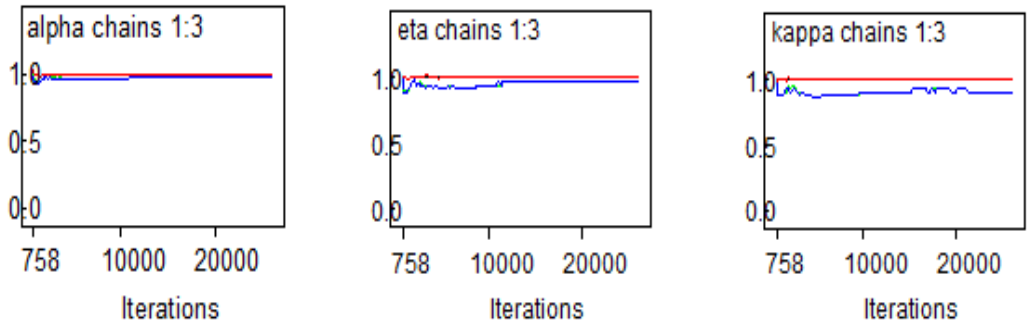

Figure 3.18: BGR plots for the alpha, eta, and kappa parameters

3.8.1.4 Some Common Statistical Convergence Diagnostic Tests

In this section, a summary of some common statistical convergence diagnostics tests is provided in Table 3.10.

Table 3.10: Summary of some statistical convergence diagnostic tests

Parameter	Geweke's diagnostic Pr $>\|z\|$	Raftery and Lewis Total no. of samp.	p-value	Heidelberger-Welch Stationarity test	Halfwidth test
alpha	-1.1992	3823	0.072	Passed	Passed
eta	-0.5711	4338	0.690	Passed	Passed
kappa	0.4144	4106	0.980	Passed	Passed

3.8.1.5 Ergodic Mean Plot

Figure 3.19 shows a time-series graph of each parameter and it displays the running mean (or ergodic mean) plots for the three parameters of the GLL distribution. The running mean plots of alpha, eta, and kappa show that the chains converge to the values in Table 3.11 after N iterations.

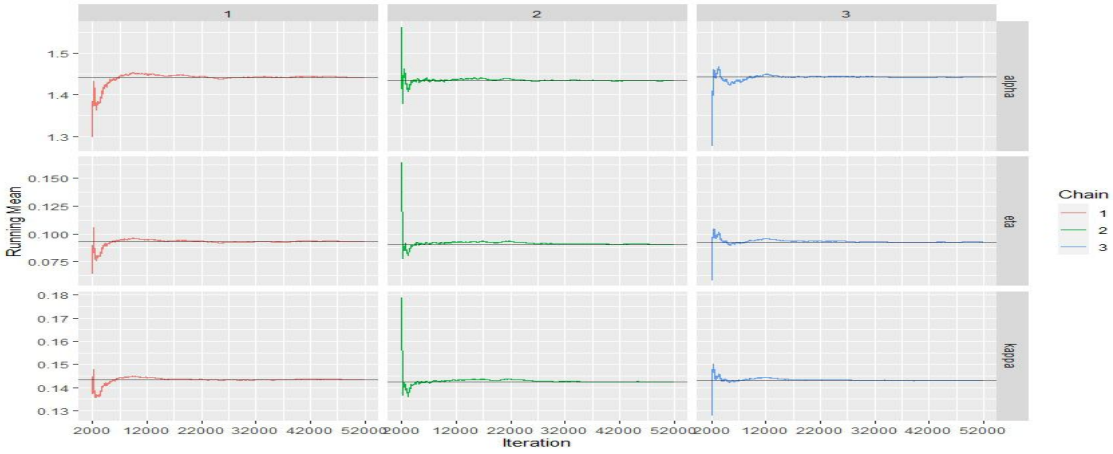

Figure 3.19: Ergodic mean plots for the alpha, eta, and kappa parameters

3.8.2 Posterior Analysis

In this section, we present numerical and visual summaries of the posterior distribution for each of the three chains. The joint posterior distribution for the proposed model was estimated using the JAGS software (Plummer et al., 2019). For each proposed model, we ran three parallel chains with 50,000 iterations and a burn-in of 5,000 . Chains were thinned by storing every fifth iteration to reduce autocorrelation in the sample. The use of various convergence diagnostic tools ensured convergence to the joint posterior.

3.8.2.1 Numerical Summary

We have considered different quantities of interest and their numeric data based on an McMC sample of posterior properties for generalized log-logistic distribution. The McMC simulation results include the results of of the posterior mean, posterior standard deviation, naïve standard error, time-series standard error, Markov chain error, the posterior five-point summary statistics (minimum, lower quartile (Q1), median (Q2), upper quartile (Q3), and maximum), the posterior skewness, posterior kurtosis, 2.5th percentile, 97.5 th percentile, and the credible interval followed by the highest probability density (HPD).

The naive standard error is defined as a measure of simulation error in the mean rather than posterior uncertainty.

$$\text{naive SE} = \frac{\text{posterior SD}}{\sqrt{n}}. \tag{3.66}$$

The time-series SE adjusts the "naïve" SE for autocorrelation. The posterior properties of the model parameters are summarized in Table 3.11

3.8.2.2 Visual Summary

In this subsection, we have considered different graphs for a visual summary of the posterior properties; those include the box plot, density strip plots, histogram, and trace plots for the parameters. These graphs and plots provide a nearly complete picture of the parameters' posterior uncertainty (Fernández-i Marín, 2016). We applied the posterior sample $\left(\alpha^{(j)}, k^{(j)} \text{ and } \eta^{(j)}\right), j = 1, \ldots, 15000$, to draw these graphs.

(1) Box Plots. The boxes in Figure 3.20 represent interquartile ranges, and the line in the middle of each box is the median; the arms of each box extend to encompass the central 95 percent of the distribution, and their ends thus correspond to the 2.5

Table 3.11: Numerical summaries of posterior properties for the GLL model with gamma priors based on an McMC sample.

Characteristics	Chain 1			Chain 2			Chain 3		
	α	η	k	α	η	k	α	η	k
Mean	1.444	0.094	0.144	1.437	0.093	0.144	1.441	0.093	0.143
SD	0.175	0.041	0.018	0.172	0.040	0.018	0.174	0.040	0.018
Naïve SE	0.002	0.001	0.0002	0.002	0.001	0.0002	0.002	0.001	0.0002
Time-series SE	0.003	0.0003	0.0003	0.003	0.0003	0.0004	0.003	0.001	0.0003
MC error	0.001	0.0004	0.0001	0.001	0.0003	0.0002	0.002	0.002	0.0001
Minimum	0.967	0.003	0.090	0.933	0.001	0.090	0.965	0.001	0.088
2.5th percentile	1.139	0.027	0.112	1.134	0.025	0.112	1.134	0.024	0.112
Q1	1.319	0.065	0.131	1.316	0.064	0.131	1.316	0.064	0.131
Medium (Q2)	1.432	0.090	0.142	1.425	0.088	0.143	1.425	0.090	0.142
Q3	1.550	0.118	0.155	1.548	0.117	0.155	1.553	0.118	0.155
97.5th percentile	1.825	0.184	0.182	1.820	0.180	0.181	1.820	0.180	0.183
Maximum	2.240	0.307	0.238	2.087	0.281	2.392	2.233	0.290	0.233
Mode	1.450	0.090	0.145	1.450	0.090	0.145	1.350	0.090	0.145
Variance	0.031	0.002	0.0003	1.636	0.002	0.0003	0.030	0.002	0.0003
Skewness	0.463	0.655	0.484	0.372	0.514	0.375	0.365	0.524	0.423
Kurtosis	0.372	0.821	0.612	0.153	0.344	0.301	0.044	0.420	0.412
95% credible interval	(1.139, 1.825)	(0.027, 0.184)	(0.112, 0.182)	(1.134, 1.820)	(0.025, 0.180)	(0.112, 0.181)	(1.134, 1.820)	(0.024, 0.180)	(0.112, 0.183)
95% HPD interval	(1.113, 1.784)	(0.021, 0.174)	(0.112, 0.181)	(1.104, 1.764)	(0.021, 0.171)	(0.110, 0.179)	(1.107, 1.779)	(0.019, 0.172)	(0.108, 0.178)

percent and 97.5 percent quartiles, respectively.

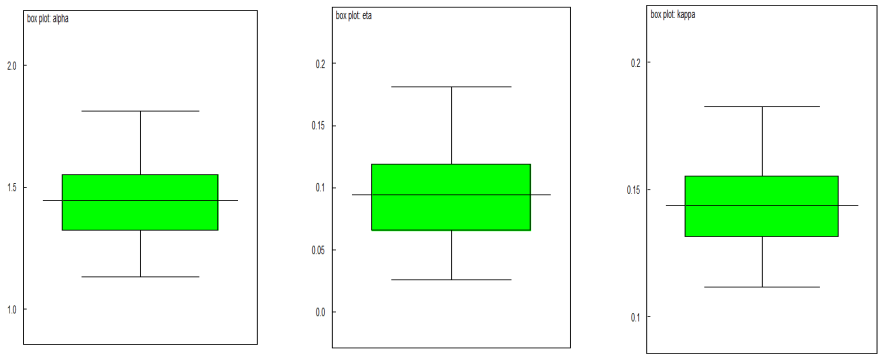

Figure 3.20: The box plots for the alpha, eta, and kappa parameters

(2) Density and Histogram Plots. Histogram can provide information about the behaviour in the tails, skewness, data outliers, and the presence of multimodal behaviour. The graphs in Figure 3.21 can provide us with a nearly complete picture of the posterior uncertainty about the GLL parameters, while the graphs in Figure 3.22 show a comparison of the full density and partial density of the parameters.

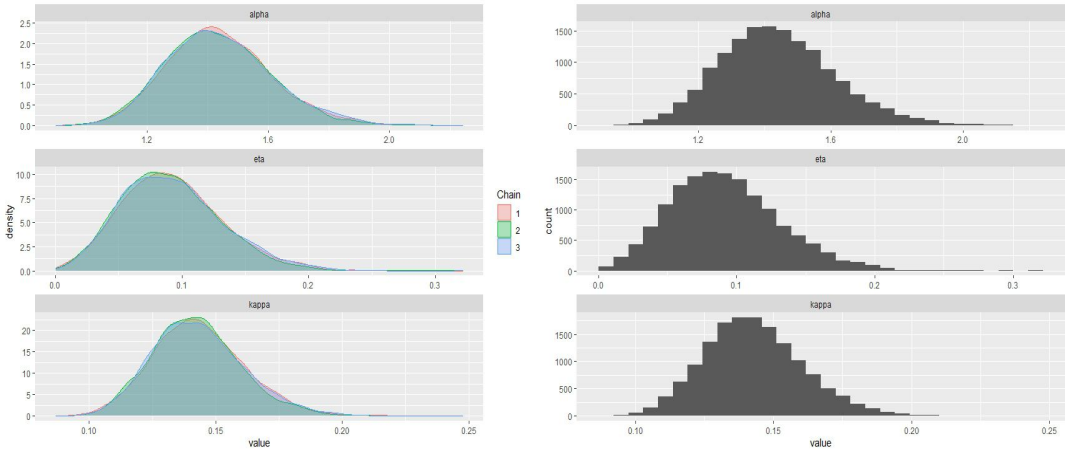

Figure 3.21: Histogram and Kernel density plots for the alpha, eta, and kappa parameters.

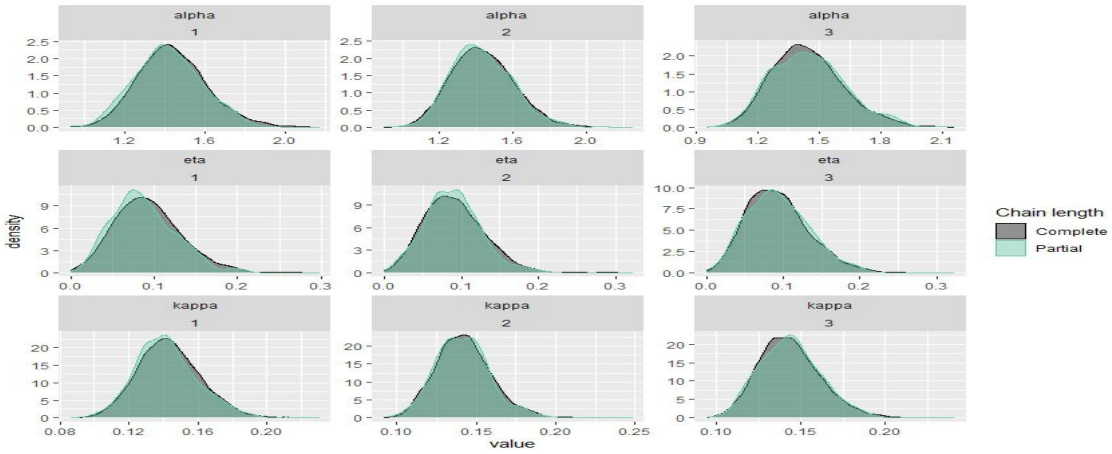

Figure 3.22: Density plots for the parameters comparing the whole chains with their last parties.

(3) Trace Plots. A trace plot, also known as "a time-series plot," is a representation of the iteration number versus the value of the parameter drawn at each iteration. Because the plots do not show long-term increasing or decreasing trends but rather resemble a horizontal band in Figure 3.23, we can conclude that the chains have converged.

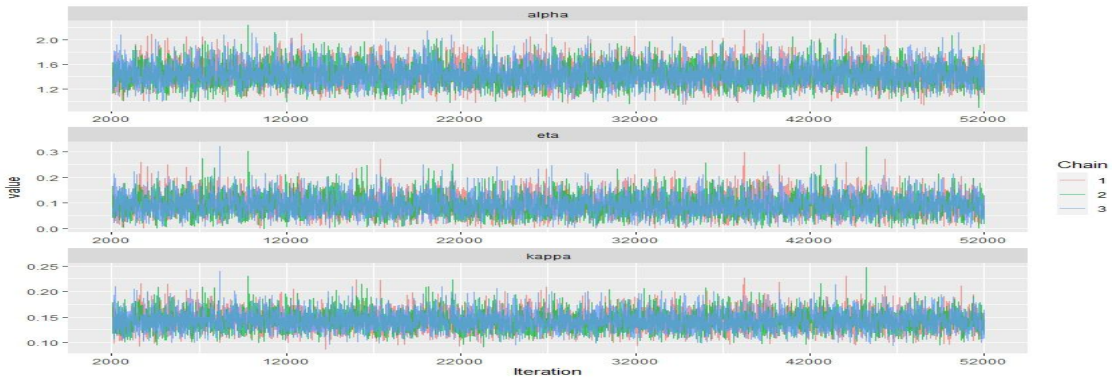

Figure 3.23: Trace plots for the alpha, eta, and kappa parameters

3.9 Conclusions

This chapter introduced and presented results on the mathematical and statistical properties of the generalized log-logistic distribution. The GLL model contains several parametric survival sub-models that could be used in a variety of statistics and probability applications. Statistical properties such as quantile function and their related results, moments and their related results, rth central moments, and residual and reversed residual life were derived. We have also considered the Bayesian and classical inference of the unknown parameters of the proposed distribution when the data is uncensored or complete. The Bayesian estimates are obtained using the Gibbs sampling method under the assumption of independent gamma priors on the shape and scale parameters. It is worth noting that when prior information is available, Bayes estimates clearly outperform maximum likelihood estimates. To assess the behaviour of the estimators, Monte Carlo simulations are run. The proposed distribution was also applied to a real-world data set and provided a better fit than its submodels and other common parametric survival distributions based on goodness-of-fit statistics, log-likelihood function, and information criterion values. As a result, we conclude that the GLL is the most appropriate model among the distributions considered and it is a very competitive model for explaining lifetime phenomena.

CHAPTER 4

A Flexible Bayesian Parametric Proportional Hazard Model: Simulation and Applications to Right-Censored Healthcare Data

In this chapter, we present our third published manuscript [1] about a flexible Bayesian parametric proportional hazard model. Note that the materials of this chapter have been reproduced from our article.

4.1 Introduction

The statistical analysis of survival data is an essential topic in many fields, including medicine, biology, environmental science, healthcare, economics, engineering, social science, and epidemiology, among others. Probability distributions serve as the foundation for survival models. The family of distributions can be parametric, semi-parametric, or non-parametric. The parametric survival models lead to more efficient and smaller standard errors of the estimates than semi-parametric and non-parametric models (Collett, 2015). If the distributional assumption is correct, to be more specific.

In analyzing survival data, parametric survival models are crucial. The benefits of using parametric survival models include: (1) handling all types of censored data (left, right, interval, double, and middle); (2) application of survival analysis in a

[1] Muse, A. H., Ngesa, O., Mwalili, S., Alshanbari, H. M., & El-Bagoury, A. A. H. (2022). A Flexible Bayesian Parametric Proportional Hazard Model: Simulation and Applications to Right-Censored Healthcare Data. Journal of Healthcare Engineering, 2022

healthcare care problem , and (3) producing better estimation when you have a theoretical expectation of the baseline hazard, also (4) they can apply random effects – frailty models, and can also be used to estimate expected lives, not only hazard ratios like the accelerated failure time models (Lawless, 2011).

The proportional hazards (PH) model, in which covariates affect the hazard rate function, and the accelerated failure time (AFT) model, in which covariates affect both the hazard rate and time scale, are the two most common methods for developing parametric regression models for survival data (Christensen et al., 2010). However, other class of models have also been proposed such as the accelerated hazard (AH) model (Rubio et al., 2019), and the proportional odds (PO) model (Legrand, 2021).

One of the first steps in using a parametric approach to model survival data is to choose a suitable baseline distribution that can capture significant features of the observations of interest. Certain probability distributions are widely used in the modelling of survival data. Only a few are closed under the proportional hazard model, and none are flexible enough to describe a wide range of survival data (Khan and Khosa, 2016). Most of the distributions closed under the PH assumption fails to model a non-monotone (i.e., bathtub and unimodal) survival data sets.

The log-logistic (LL) distribution has a wide range of applications in survival data analysis and can accommodate unimodal survival data sets. The distribution is closed under both proportionality odds (PO) and multiplication of failure time (AFT) frameworks (Lawless, 2011). It is not a PH model, but rather an AFT model. However, when the log-logistic distribution is generalized, it has the appealing feature of being a member of all classes of parametric hazard-based regression models of the survival analysis because its failure rate function is quite versatile and its cumulative hazard function (chf) has a tractable form.

Extensive efforts have been made over the last decades to extend classical distributions to use as a baseline distribution for parametric hazard-based regression

models. Many modifications to the LL distribution have been introduced to make it more adaptable to a wide range of hazard shapes (Cordeiro et al., 2020), and (Muse et al., 2021b). The generalized log-logistic distribution (GLL) is one such model, which modifies the log-logistic distribution by inducing an additional shape parameter (Muse et al., 2022a). The model is tractable and closed under the PH assumption and can account for both non-monotone and monotone hazard rates (Khan and Khosa, 2016). On the other hand, recent computational advances have advocated for the use of Bayesian techniques in the field of survival and reliability analysis.

The motivating ideas behind our work on Bayesian parametric proportional hazard (PH) model with GLL baseline hazard are as follows: (i) despite the fact that there are some classical distributions closed under the PH framework, none of which is flexible enough to incorporate both monotone and non-monotone hazard rate; (ii) Bayesian inference does not rely on asymptotic approximation for statistical inference; (iii) the availability of software makes Bayesian implementation for hazard-based complicated models relatively more straightforward and simple than classical inference [18]; (iv) parametric PH model may lead to more precise estimates than semi-parametric PH model; and, last but not least, (v) the use of generalized distributions that can capture both monotone and non-monotone hazard rate functions is what makes our work unique and more appealing to biostatisticians, epidemiologists, healthcare workers as well as other applied researchers in multiple disciplines. To the best of author's knowledge, no Bayesian inferences study has been conducted on the PH model with generalized log-logistic baseline hazard. As a result, in this paper, we consider the Bayesian inference for the generalized log-logistic proportional hazard model, beginning with the PH model formulation and assumptions, revising the generalized log-logistic distribution, and verifying that the GLL distribution is closed under the PH framework. In addition, we discuss the inferential procedures and how to obtain the classical and Bayesian estimators for the model's parameters.

We also compare the proposed model to other existing distributions closed under the PH framework, and one interesting feature of this model is that it can incorporate different hazard rate shapes. Hence, the formulation of the parametric PH model and its lifetime function, the inferential procedures using both classical and bayesian approaches, and the development of the computational algorithms to fit the proposed PH model and its competing models using rjags in r software are the novelty of this study.

The chapter is structured as follows: the PH model formulation, assumptions and its probabilistic functions are discussed in Section 4.2. Section 4.3 revises the most common probability distributions closed under the PH model. The GLL distribution under the PH model is presented in Section 4.4. Section 4.5 discusses the inferential procedures of the proposed model. In Section 4.6, we present an McMC simulation study to assess the performance of the proposed model. Section 4.7 presents the application of the proposed model to two right-censored cancer data sets with monotone and non-montone hazard rates. In addition the convergence diagnostics of the McMC techniques were discussed. The Bayesian Model selection criterion is presented in Section 4.8. Finally, in the final portion, the chapter's concluding notes are offered.

4.2 PH Model Formulation and Assumptions

The parametric PH model is given with the similar form to the Cox PH model. It is the parametric form of the Cox PH models (Collett, 2015) and is formulated as follows:

4.2.1 PH Formulation

The parametric proportional hazard (PH) models are formulated using a defined baseline hazard and a link function $\psi(\boldsymbol{x}'\boldsymbol{\beta})$ for the covariates which is defined as

follows:

i. $\psi\left(\boldsymbol{x}'\boldsymbol{\beta}\right) > 0, \forall x \neq 0.$

ii. $\psi\left(x'\boldsymbol{\beta}\right)$ is is a monotone function that has a one-to-one correspondence,

iii. $\psi(\mathbf{0}) = 1.$

The most commonly found option for the link function $\psi\left(\boldsymbol{x}'\boldsymbol{\beta}\right)$ is the exponentialexp $\left(\boldsymbol{x}'\boldsymbol{\beta}\right)$ (or log-linear) function. In this work, we define the PH model with the assumption that $\psi\left(x'\boldsymbol{\beta}\right) = \exp\left(x'\boldsymbol{\beta}\right)$.

4.2.2 PH Assumptions

The PH model assumption is that the effect of covariates is to increase or decrease the hazard rate function by a proportionate amount which does not depend on t. The assumption of the PH model can be defined as:

$$h(t; x) = h_0(t)\psi\left(\boldsymbol{x}'\boldsymbol{\beta}\right) = h_0(t)\exp\left(\boldsymbol{x}'\boldsymbol{\beta}\right) = h_0(t)e^{x'\beta} \tag{4.1}$$

where $h_0(t)$ is called the baseline hazard, simplifying we get;

$$h(t \mid x) = h_0(t)\exp\left(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p\right). \tag{4.2}$$

The main difference between the Cox PH model and the parametric PH model is that the baseline hazard function is assumed to follow a specific distribution when it is fitted to the data. Using Equation 4.1 we can see that the hazard ratio (HR) comparing any two specifications of the covariates, for example ($\boldsymbol{x}$ and $\boldsymbol{x}^*$), is

$$HR\left(\boldsymbol{x}, \boldsymbol{x}*, h_0, \boldsymbol{\beta}\right) = \frac{h(t \mid \boldsymbol{x}, \boldsymbol{\beta})}{h\left(t \mid \boldsymbol{x}^*, \boldsymbol{\beta}\right)} = \frac{h_0(t \mid \boldsymbol{x})\exp\left(\boldsymbol{\beta}\boldsymbol{x}'\right)}{h_0(t \mid \boldsymbol{x})\exp\left(\boldsymbol{\beta}\boldsymbol{x}^{*'}\right)} = \exp\left[\left(\boldsymbol{x}' - \boldsymbol{x}^{*'}\right)^T \boldsymbol{\beta}\right] \tag{4.3}$$

The above equation shows us that the baseline hazards cancel each other from this ratio, so the hazard rate for one individual is proportional to the hazard rate for any other individual. On the other hand, the proportionality constant is independent of time which makes the main assumption of this model (Collett, 2015). As a result, the model is known as the proportional hazard (PH) model in the literature.
Unlike most parametric regression models including accelerated failure time (AFT) models, PH models does not include an intercept (Christensen et al., 2010). More properly, the vector $\boldsymbol{X}$ in the PH model is not assumed to have $x \equiv 1$. An intercept would get confounded with the baseline hazard function (h_0).

4.2.3 Functions Describing PH Model

The five frequent representatives of a lifetime distribution function that are used to characterize the PH model are addressed in this subsection.

4.2.3.1 Hazard Rate Function of the PH Model

The hrf of the PH model is of the form

$$h(t;x) = h_0(t)\psi\left(\boldsymbol{x}'\boldsymbol{\beta}\right) = h_0(t)\exp\left(\boldsymbol{x}'\boldsymbol{\beta}\right) = h_0(t)e^{x'\beta} \tag{4.4}$$

4.2.3.2 Cumulative Hazard Function of the PH Model

The chf of a PH model takes the following form:

$$H(t \mid x) = \int_0^t h(s;x)ds = e^{x'\beta}\int_0^t h_0(s)ds = e^{x'\beta}H_0(t) \tag{4.5}$$

4.2.3.3 Survival Function of the PH Model

The survivor function (sf) for a PH model can be derived using the following relationship between survival function and the hazard rate function. Hazard function

is given by:

$$h(t \mid x) = \frac{f(t \mid x)}{S(t \mid x)} \tag{4.6}$$

Cumulative hazard function:

$$\begin{aligned} H(t \mid x) &= \int_0^t h(u)\mathrm{du} = \int_0^t \frac{f(u)}{S(u)}\mathrm{du} = \int_0^t \frac{-\mathrm{dS}(u)}{S(u)}\mathrm{du} \\ &= -\log\{S(t)\}, \\ f(t \mid x) &= h(t \mid x)S(t \mid x) = h(t \mid x)\exp\{-H(t \mid x)\}. \end{aligned} \tag{4.7}$$

Using the above expressions, we can easily find

$$\begin{aligned} S(t \mid x) &= \exp\{-H(t \mid x)\}, \\ S(t \mid x) &= \exp\{-H(t \mid x)\} = \exp\left\{-\int_0^t \psi(x)h_0(t)dt\right\}, \\ &= \exp\left\{-\psi(x)\int_0^t h_0(t)dt\right\}, \\ &= \left[\exp\left\{-\int_0^t h_0(t)dt\right\}\right]^{\psi(x)}, \\ &= [S_0(t)]^{\psi(x)}, \psi(x) > 0. \end{aligned} \tag{4.8}$$

4.2.3.4 Cumulative Distribution Function of the PH Model

The cdf of the PH model, also known as the lifetime distribution function, is given by

$$\begin{aligned} F(t) &= 1 - S(t) = 1 - \exp\{-H(t)\}, \\ F(t) &= 1 - [S_0(t)]^{\psi(x)}. \end{aligned} \tag{4.9}$$

4.2.3.5 Probability Density Function of the PH Model

The pdf or the failure density function of the PH model is defined as

$$f(t) = f_0(t)\psi(x)\left[S_0(t)\right]^{\psi(x)-1}. \tag{4.10}$$

The five representatives used here were chosen for their special meaning for lifetime data, their intuitive appeal, their utility in survival data analysis, and, last but not the least, their popularity in probability theory and statistics.

The PH model can be formulated without assuming a probability distribution for survival times, and this leads to the well-known Cox PH model (Cox, 1972). On the other hand, assuming a probability distribution for survival times leads to the fully parametric PH model. The most common parametric survival models used are as follows: exponential, Weibull, Gompertz, log-logistic, log-normal, gamma, and the generalized gamma distributions. Only the exponential, Weibull, and Gompertz distributions are used for the PH model. The log-logistic and the log-normal distributions are not closed under the PH framework. Weibull distribution is the only one that is closed under both parametric AFT and PH models.

4.3 Distributions Closed under PH Framework

In this section, we present most common parametric distributions that are closed under the PH framework and are used to analyze survival data. These distributions have been studied and used in various contexts in the literature (Lawless, 2011).

4.3.1 Exponential PH Model.

For the PH model, the exponential baseline hazard is

$$h(t) = k. \tag{4.11}$$

So, according to the formulation of the PH framework in Equation (4.1), the hazard rate for an individual with covariate vector x and link function $\psi(x)$ is

$$h(t) = h_0(t)\psi(x) = k \cdot \psi(x) \tag{4.12}$$

Applying the log-linear function $\psi(\mathbf{x}'\beta) = \exp(\mathbf{x}'\beta)$, we can simplify into $h_{EPH}(t) = k \cdot \exp(\mathbf{x}\boldsymbol{\beta}) = k \cdot \exp(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p)$. In this equation, the hrf has the exponential distribution with scale parameter k. $\exp(\mathbf{x}'\beta)$ which indicates that the PH assumption is satisfied with the exponential distribution. It is worth mentioning that the exponential distribution is often found to be inadequate to describe survival data. This makes the applicability of this distribution fairly limited. The other lifetime distributions of the exponential PH model are as follows.

The survival function of the exponential PH model is

$$S_{\text{EPH}}(t) = [\exp\{-kt\}]^{\exp(\mathbf{x}'\beta)}.$$

The pdf of the exponential PH model is

$$f_{\text{EPH}}(t) = k \exp\{-kt\} \exp(\mathbf{x}'\beta) [\exp\{-kt\}]^{\exp(\mathbf{x}'\beta)-1}.$$

The cdf of the exponential PH model is

$$F_{\text{EPH}}(t) = 1 - [\exp\{-kt\}]^{\exp(\mathbf{x}'\beta)}.$$

The chf of the exponential PH model is:

$$H_{\text{EPH}}(t) = kt \exp(\mathbf{x}'\beta).$$

4.3.2 Gompertz PH Model.

For the PH model, the Gompertz baseline hazard rate function is given by

$$h(t) = \alpha \text{ke}^{tk} \tag{4.13}$$

So, according to the formulation of the PH framework in Equation (4.1),, the hazard rate for an individual with covariate vector x and link function $\psi(x)$ is

$$h(t) = h_0(t)\psi(x) = \alpha \mathrm{k} \cdot e^{tk} \cdot \psi(x). \tag{4.14}$$

Applying the log-linear function $\psi(\mathbf{x}'\beta) = \exp(\mathbf{x}'\beta)$, we can simplify into

$$h_{\mathrm{GoPH}}(t) = \alpha k.e^{tk} \cdot \exp(\mathbf{x}'\beta) = \alpha k.e^{tk}.\exp(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p). \tag{4.15}$$

In the above equation, it is straightforward that the PH property is satisfied. However, the Gompertz PH model is rarely used in the real-life applications.
The other lifetime distributions of the Gompertz PH model are as follows:
The survival function of the Gompertz PH model is

$$S_{\mathrm{GoPH}}(t) = \left[\exp\left\{-\alpha\left(e^{tk}-1\right\}\right]^{\exp(\mathbf{x}'\beta)}. \tag{4.16}$$

The pdf of the Gompertz PH model is

$$f_{\mathrm{GoPH}}(t) = \alpha k \cdot e^{tk} \exp\left\{-\alpha\left(e^{tk}-1\right\} \exp(\mathbf{x}'\beta) \left[\exp\left\{-\alpha\left(e^{tk}-1\right\}\right]^{\exp(x'\beta)-1}. \tag{4.17}$$

The cdf of the Gompertz PH model is

$$F_{\mathrm{GoPH}}(t) = 1 - \left[\exp\left\{-\alpha\left(e^{tk}-1\right\}\right]^{\exp(\mathbf{x}'\beta)} \tag{4.18}$$

The chf of the Gompertz PH model is

$$H_{\mathrm{GoPH}}(t) = \exp(\mathbf{x}'\beta)\,\alpha\left(e^{tk}-1\right). \tag{4.19}$$

4.3.3 Weibull PH Model.

For the PH model the Weibull baseline hazard is

$$h(t) = \alpha k(kt)^{\alpha-1}. \tag{4.20}$$

So, according to the formulation of the PH framework in Equation (4.1), the hazard rate for an individual with covariate vector x and link function $\psi(x)$ is

$$h(t) = h_0(t)\psi(x) = \alpha k(kt)^{\alpha-1}\psi(x). \tag{4.21}$$

Applying the log-linear function $\psi(\mathbf{x}'\beta) = \exp(\mathbf{x}'\beta)$, we can simplify into

$$\begin{aligned} h_{\text{WPH}}(t) &= \alpha k(kt)^{\alpha-1}\exp(\mathbf{x}'\beta) \\ &= \alpha k(kt)^{\alpha-1}\exp(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p) \end{aligned} \tag{4.22}$$

In this equation, the model has the Weibull distribution with rate parameter k. $\exp(\mathbf{x}'\beta)$ and shape parameter α which indicates that the PH assumption is satisfied with the Weibull distribution with constant α.
The other lifetime distributions of the PH Weibull model are as follows: the survival function of the Weibull PH model is

$$S_{\text{WPH}}(t) = [\exp\{-(kt)^{\alpha}\}]^{\exp(\mathbf{x}'\beta)} \tag{4.23}$$

The pdf of the Weibull PH model is

$$f_{WPH}(t) = \alpha k(kt)^{\alpha-1}\exp\{-(kt)^{\alpha}\}\exp(\mathbf{x}'\beta)[\exp\{-(kt)^{\alpha}\}]^{\exp(\mathbf{x}'\beta)-1}. \tag{4.24}$$

The cdf of the Weibull PH model is

$$F_{\text{WPH}}(t) = 1 - \left[\exp\left\{-(kt)^{\alpha}\right\}\right]^{\exp(\mathbf{x}'\beta)}. \tag{4.25}$$

The chf of the Weibull PH model is

$$H_{\text{WPH}}(t) = \exp(\mathbf{x}'\beta)(kt)^{\alpha} \tag{4.26}$$

4.4 The Proposed PH Model

For the PH model, the generalized log-logistic baseline hazard is

$$h(t) = \frac{\alpha k(kt)^{\alpha-1}}{[1+(\eta t)^{\alpha}]}. \tag{4.27}$$

So, according to 4.1, the hazard rate for an individual with covariate vector x and link function $\psi(x)$ is

$$h(t) = h_0(t)\psi(x) = \frac{\alpha k(kt)^{\alpha-1}}{[1+(\eta t)^{\alpha}]}\psi(x). \tag{4.28}$$

Applying the log-linear function $\psi(\mathbf{x}'\beta) = \exp(\mathbf{x}'\beta)$, we can simplify into

$$\begin{aligned} h_{\text{GLLPH}}(t) &= \frac{\alpha k(kt)^{\alpha-1}}{[1+(\eta t)^{\alpha}]}\exp(\mathbf{x}'\beta) \\ &= \frac{\alpha k^{\alpha}t^{\alpha-1}}{[1+(\eta t)^{\alpha}]}\exp(\mathbf{x}'\beta) \\ &= \frac{\alpha\left(k\cdot\exp(\mathbf{x}'\beta)^{1/\alpha}\right)^{\alpha}t^{\alpha-1}}{[1+(\eta t)^{\alpha}]} \\ &= \frac{\alpha k^{*\alpha}t^{\alpha-1}}{[1+(\eta t)^{\alpha}]} \end{aligned} \tag{4.29}$$

In this equation, the hrf can be recognized as a generalized log-logistic distribution as well, but contrary to (4.27), the rate parameter is $k^* = k.\exp(\mathbf{x}'\beta)^{1/\alpha}$ and shape

parameters are α and η which indicates that the PH assumption is satisfied with the GLL distribution and the proposed model is closed under the PH framework. The other lifetime distribution functions for the GLL PH model are as follows: the survivor function of the GLL PH model is:

$$S_{\text{GLLPH}}(t) = \left[[1 + (\eta t)^{\alpha}]^{-k^{\alpha}/\eta^{\alpha}}\right] \exp(\mathbf{x}'\beta) \tag{4.30}$$

The pdf of the GLL PH model is

$$f_{\text{GLLPH}}(t) = \frac{\alpha k (kt)^{\alpha-1}}{[1 + (\eta t)^{\alpha}]^{\frac{k^{\alpha}}{\eta^{\alpha}}+1}} \exp(\mathbf{x}'\beta) \left[[1 + (\eta t)^{\alpha}]^{k^{\alpha}/\eta^{\beta}}\right]^{\exp(\mathbf{x}'\beta)-1}. \tag{4.31}$$

The cdf of the GLL PH model is

$$F_{\text{GLLPH}}(t) = 1 - \left[[1 + (\eta t)^{\alpha}]^{-k^{\alpha}/\eta^{\alpha}}\right] \exp(\mathbf{x}'\beta). \tag{4.32}$$

The chf of the GLL PH model is

$$H_{\text{GLLPH}}(t) = \exp(\mathbf{x}'\beta) \frac{k^{\alpha}}{\eta^{\alpha}} \log[1 + (\eta t)^{\alpha}] \tag{4.33}$$

4.5 Model Inference

We discuss the classical approach (using MLE) and Bayesian approach (assuming non-informative priors) estimation techniques for the proposed parametric PH model parameters in this section.

4.5.1 MLE for Right-Censored Survival Data

We examine the challenge of estimating the proposed model's distributional parameters and regression coefficients for right-censored survival data in this sub-section. Because of its appealing qualities, such as consistency, asymptotic efficiency, asymp-

totic unbiasedness, and asymptotic normality, MLE is one of the most common strategies for estimating the parameters of hazard-based regression models. Let there be n individuals with lifetimes represented by $T_1, T_2, \ldots, T_n$. Assuming that the data are subject to right censoring, we observe $t_i = \min(T_i, C_i)$, where $C_i > 0$ corresponds to a potential censoring time for individual i. Allowing $\delta_i = \mathrm{I}(T_i, C_i)$ that equals 1 if $T_i \leq C_i$ and 0 otherwise.

Suppose that a right-censored random sample with data $D = (t_i, \delta_i, \boldsymbol{x}_i), i = 1, 2, \ldots, n$, is available, where t_i is a censoring time or a survival time according to whether $\delta_i = 0$ or 1 , respectively and $x_i = x_1, x_2, \ldots, x_n$ is an $n \times 1$ column vector of external covariates for the i$^{\text{th}}$ individual, $\boldsymbol{\vartheta}$ is the vector of parameters associated with the baseline distribution, and β is the vector of regerssion coefficients. When parametric PH model is considered the censored likelihood function can be expressed as:

$$\begin{aligned}
L(\boldsymbol{\vartheta}, \boldsymbol{\beta} \mid \boldsymbol{D}) &= \prod_{i=1}^{n} \left[f\left(t_i \mid \boldsymbol{\vartheta}, \boldsymbol{\beta}, \boldsymbol{x}\right)\right]^{\delta_i} \left[s\left(t_i \mid \boldsymbol{\vartheta}, \boldsymbol{\beta}, \boldsymbol{x}\right)\right]^{1-\delta_i} \\
&= \prod_{i=1}^{n} \left[h\left(t_i \mid \boldsymbol{\vartheta}, \boldsymbol{\beta}, \boldsymbol{x}\right).S\left(t_i \mid \boldsymbol{\vartheta}, \boldsymbol{\beta}, \boldsymbol{x}\right)\right]^{\delta_i} \left[s\left(t_i \mid \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{x}\right)\right]^{1-\delta_i} \\
&= \prod_{i=1}^{n} \left[h\left(t_i \mid \boldsymbol{\vartheta}, \boldsymbol{\beta}, \boldsymbol{x}\right)\right]^{\delta_i} \left[s\left(t_i \mid \boldsymbol{\vartheta}, \boldsymbol{\beta}, \boldsymbol{x}\right)\right] \\
&= \prod_{i=1}^{n} \left[h\left(t_i \mid \boldsymbol{\vartheta}, \boldsymbol{\beta}, \boldsymbol{x}\right)\right]^{\delta_i} \exp\left[-\int_0^t h(u)du\right] \\
&= \prod_{i=1}^{n} \left[h\left(t_i \mid \boldsymbol{\vartheta}\right) \exp\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\right]^{\delta_i} \exp\left[-\left[H\left(t_i \mid \boldsymbol{\vartheta}\right) \exp\left(\boldsymbol{x}'\boldsymbol{\beta}\right)\right]\right]
\end{aligned} \tag{4.34}$$

An iterative optimization procedure (e.g., Newton-Raphson algorithm) can be used to obtain the maximum likelihood estimation $\widehat{\boldsymbol{\vartheta}}_{\text{of}}\, \boldsymbol{\vartheta}$. Hypothesis testing and interval estimations of model parameters are possible due to the MLEs' approaching normality [7]. The natural logarithm of the likelihood function, so-called log-likelihood

function can be written as follows:

$$\ell(\boldsymbol{\vartheta}, \boldsymbol{\beta} \mid \boldsymbol{D}) = \sum_{i=1}^{n} \delta_i \log \left[h_0 \left(t_i \mid \boldsymbol{\vartheta} \right) + \boldsymbol{x}_i' \boldsymbol{\beta} \right] - \sum_{i=1}^{n} H_0 \left(t_i \mid \boldsymbol{\vartheta} \right) \exp \left(\boldsymbol{x}_i' \boldsymbol{\beta} \right), \tag{4.35}$$

where; $\boldsymbol{\beta}$ is a vector of the regression coefficients, and $\boldsymbol{\vartheta}' = (k, \alpha, \eta)$ is the vector of the baseline distributional parameters In our case; if we assume that $a = \sum_{i=1}^{n} \delta_i, p_i = \exp\left(\boldsymbol{x}_i' \boldsymbol{\beta}\right)$ and $q_i = (\eta t_i)^k$. Using Eqn. (50) for the $h_0(.)$ and noting that $H_0(t; \boldsymbol{\theta}) = \int_0^t h(u) du$ is the baseline cumulative hazard rate function as given by Eqn. (51). The full loglikelihood function of the GLL PH model can be expressed as follow:

$$\begin{aligned} \ell(\boldsymbol{\vartheta} \mid t) = & \operatorname{alog} \alpha + \operatorname{a} \alpha \log k + (\alpha - 1) \sum_{i=1}^{n} \delta_i \log t_i \\ & - \sum_{i=1}^{n} \delta_i \log \left(1 + q_i\right) + a \log p_i \\ & - \left(\frac{\kappa}{\eta} \right)^{\alpha} \sum_{i=1}^{n} p_i \log \left(1 + q_i\right) \end{aligned} \tag{4.36}$$

To obtain the MLE's of $\boldsymbol{\theta}' = (k, \alpha, \eta)$ and $\boldsymbol{\beta}'$, we can maximize (58) directly with respect to (k, α, η) and $\boldsymbol{\beta}'$. or we can solve the non-linear equations below or the 1st derivative of the log-likelihood function. The 1st derivatives of the log-likelihood function are

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\tau} \mid t)}{\partial \alpha} = & \frac{a}{\alpha} + a \log p_i + \sum_{i=1}^{n} \delta_i \log t_i - \frac{1}{\alpha} \sum_{i=1}^{n} \delta_i q_i \left[\frac{\log q_i}{(1 + q_i)} \right] \\ & - \left(\frac{k}{\eta} \right)^{\alpha} \left(\frac{1}{\alpha} \right) \sum_{i=1}^{n} p_i q_i \left[\frac{\log q_i}{(1 + q_i)} \right] \\ & - \left(\frac{k}{\eta} \right)^{\alpha} \log \left(\frac{k}{\eta} \right) \sum_{i=1}^{n} p_i \log \left(1 + q_i\right) \end{aligned} \tag{4.37}$$

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\tau} \mid t)}{\partial \eta} = & - \left(\frac{\alpha}{\eta} \right) \sum_{i=1}^{n} \delta_i \left[\frac{q_i}{1 + q_i} \right] - \left(\frac{\alpha}{\eta} \right) \left(\frac{k}{\eta} \right)^{\alpha} \sum_{i=1}^{n} p_i \left[\frac{q_i}{1 + q_i} \right] \\ & - \left(\frac{\alpha}{\eta} \right) \left(\frac{k}{\eta} \right)^{\alpha} \sum_{i=1}^{n} p_i \log \left(1 - \left[\frac{q_i}{1 + q_i} \right] \right) \end{aligned} \tag{4.38}$$

$$\frac{\partial \ell(\boldsymbol{\tau} \mid t)}{\partial k} = \frac{a\alpha}{k} - \left(\frac{\alpha}{k}\right)\left(\frac{k}{\eta}\right)^{\alpha} \sum_{i=1}^{n} p_i \log\left(1+q_i\right) \tag{4.39}$$

$$\frac{\partial \ell(\tau \mid t)}{\partial \beta_j} = \sum_{i=1}^{n} \delta_i Z_{ij} - \left(\frac{k}{\eta}\right)^{\alpha} \sum_{i=1}^{n} p_i \log\left(1+q_i\right) Z_{ij} \quad \text{for } j = 1, 2, \ldots, p \tag{4.40}$$

To maximize log-likelihood functions, many software packages are available including proven optimization algorithms.

4.5.2 Bayesian Inference

In this sub-section, Bayesian inference was used to estimate distributional parameters and regression coefficients using objective (or non-informative) priors to obtain proper posterior distributions.

1. Prior Distribution

The specification of a prior distribution is a crucial aspect of any Bayesian inference. In parametric survival regression models, this is especially true. As a result, the prior scenario is built in this study using a non-informative independent prior for the parameters. The marginal prior distribution for every regression coefficient $\beta_m, m = 1, \ldots, 5$, is prompted as a normal distribution centred at zero and with a small precision, $N(0, 0.001)$; on the other hand, a gamma distribution, Gamma $(10, 10)$, is chosen as the marginal prior distribution for the parameters of the GLL PH model due to the versatility of gamma distribution that include the non-informative priors (uniform) on the shape parameters. Many research publications in the literature, such as Danish and Arshad (2017) considered the assumption of the gamma priors for the baseline hazard parameters of PH models. Alvares et al. (2021) take the assumption of independent gamma priors for the baseline hazard parameters of eight different parametric survival models, Muse et al. (2022a) used the assumption of independent gamma priors for the baseline hazard parameters of the of the generalized loglogistic AFT model, and other researchers take these priors into account.

For the baseline parameters of the GLL-PH model, we assume independent gamma priors.

$$p(\alpha) \sim G\left(a_1, b_1\right) = \frac{b_1^{a_1}}{\Gamma\left(a_1\right)} \alpha^{a_1-1} e^{-b_1 \alpha}; a_1, b_1, \alpha > 0 \tag{4.41}$$

$$p(\eta) \sim G\left(a_2, b_2\right) = \frac{b_2^{a_2}}{\Gamma\left(a_2\right)} \eta^{a_2-1} e^{-b_2 \eta}; a_2, b_2, \eta > 0 \tag{4.42}$$

$$p(\kappa) \sim G\left(a_3, b_3\right) = \frac{b_3^{a_3}}{\Gamma\left(a_3\right)} k^{a_3-1} e^{-b_3 k}; a_3, b_3, k > 0 \tag{4.43}$$

Prior to that, we had the regression coefficients (assuming a normal distribution).

$$p\left(\boldsymbol{\beta}'\right) \sim N\left(a_4, b_4\right) \tag{4.44}$$

The density function of the combined prior distribution of all unknown parameters and the regression coefficients are given us:

$$p\left(\alpha, k, \eta, \boldsymbol{\beta}'\right) = p(\alpha) p(\eta) p(k) p\left(\boldsymbol{\beta}'\right) \tag{4.45}$$

2. Likelihood Function

Unfortunately, the likelihood function of this generalized model is not implemented in BUGS and JAGS syntax. To generate the likelihood function we use the "Zero's trick" method that become popular in survival analysis and relies on Poisson modelling of expanded or reconstructed data (Alvares et al., 2021). The zero's trick approach works on the assumption that perhaps the contribution of a Poisson (λ) observable of zero is $\exp(-\lambda)$; if we set $\lambda = -\log\left(f\left(t_i \mid \boldsymbol{\vartheta}, \boldsymbol{\beta}, \boldsymbol{x}\right)\right)$ with observable data as a vector of $0's$, we receive the right contributions of the proposed model (Christensen et al., 2010).

3. Posterior Distribution

The joint posterior density function is equal to the multiplication of the prior distribution $p\left(\alpha, k, \eta, \boldsymbol{\beta}'\right)$ and the likelihood function the joint posterior density function

of the parameters α, k, η, and $\boldsymbol{\beta}'$ of GLL PH model given the data can be expressed using Bayes' Theorem as:

$$\begin{aligned} p\left(\alpha, k, \eta, \boldsymbol{\beta}' \mid x\right) &\propto p\left(\alpha, k, \eta, \boldsymbol{\beta}'\right) L\left(\alpha, k, \eta, \boldsymbol{\beta}'\right) \\ p\left(\alpha, k, \eta, \boldsymbol{\beta}' \mid x\right) &\propto p(\alpha) p(\eta) p(k) p\left(\boldsymbol{\beta}'\right) L\left(\alpha, k, \eta, \boldsymbol{\beta}'\right) \end{aligned} \tag{4.46}$$

where; the first four terms on the equation represent the prior specification for the unknown parameters and are assumed to be independent, the $L\left(\alpha, \kappa, \eta, \boldsymbol{\beta}'\right)$ is the likelihood function expressed as follows:

$$L\left(\alpha, k, \eta, \boldsymbol{\beta}'\right) = \prod_{i=1}^{n} \left[\frac{\alpha k (kx)^{\alpha-1}}{\left[1+(\eta t)^{\alpha}\right]} \exp\left(x'\boldsymbol{\beta}\right)\right]^{\delta_i} \left[\exp\left(\boldsymbol{x}'\boldsymbol{\beta}\right) \frac{k^{\alpha}}{\lambda^{\alpha}} \log\left[1+(\lambda x)^{\alpha}\right]\right] \tag{4.47}$$

$$p\left(\alpha, k, \eta, \boldsymbol{\beta}' \mid x\right) \propto \left\{\prod_{j=0}^{p} \boldsymbol{\pi}\left(\boldsymbol{\beta}_j\right)\right\} \alpha^{a_1+n-1} \eta^{a_2+n-1} k^{a_3+n-1} e^{-(b_1\alpha + b_2\eta + b_3 k)} L\left(\alpha, k, \eta, \boldsymbol{\beta}'\right) \tag{4.48}$$

The marginal distributions of the model parameters and the normalising joint posterior density function are difficult to calculate analytically, requiring high-dimensional integration and no close form inferences. To obtain estimates, we use McMC simulation methods, which involve sampling from the posterior distribution through using the Metropolis-Hastings Algorithm.

4.6 Simulation Study

In this section, we undertake an extensive simulation investigation to demonstrate the proposed parametric proportional hazard model's good Bayesian features. The parameter values are chosen to construct situations that mimic cancer population studies using a cancer that is severe (with a lower five-year survival rate), such as lung cancer (Rubio et al., 2019). We demonstrate parame-ter estimation, the effect of censoring proportions and sample sizes on inference in more detail.

4.6.1 Generating Survival Data from PH Model

To simulate survival data for the GLL PH model, we use the inversion technique (Bender et al., 2005; Austin, 2012) to generate survival data. This strategy is based on the link between a survival random variable's cumulative hazard rate function and a standard uniform random variable. When the cumulative hazard rate function has a closed form expression, it may be immediately applied, inverted, and readily implemented with R (Team, 2019). The censoring rates were estimated using administrative censorship at (1) Tc = 5 years, which resulted in around 20% censoring in all sets, and (2) Tc = 3 years, which resulted in about 30% censoring in all sets. For the purposes of this simulation, we assume that survival times are distributed using the generalized log-logistic distribution $GLL(\alpha, \eta, k)$. Using the reverse chf given in Eqn [48], lifetimes data can be simulated as follows:

$$\mathrm{T} = \mathrm{H}_0^{-1}\left\{\frac{\left(\mathrm{e}^{\left(\frac{\eta}{k}\right)^{\alpha}\left[\frac{-\log(1-\mathrm{U})}{\mathrm{e}^{\beta x_i}}\right]}-1\right)^{\frac{1}{\alpha}}}{\eta}\right\} \tag{4.49}$$

Where; α, η, and $k > 0$.

4.6.2 Simulation Design

The simulation analysis was carried out by conducting a series of simulations with different sample sizes (n = 100, and 300) sets and censoring proportions (Tc = 20 and 30 percentages), all based on the PH model in equation (4.29). The GLL PH model's true parameter vector is set as follows:(1) Set I: distributional parameter values ($\alpha = 1.5, k = 0.75$, and $\eta = 1.25$), and covariates $\beta = (0.75, -0.75, 0.5)$, (2) Set II: distributional parameter values ($\alpha = 1.5, k = 0.95$, and $\eta = 1.5$), and Covariates $\beta = (0.75, -0.75, 0.5)$.

The values of the covariates were simulated as follows: (1) combination of uniform

distributions with 0.25 probability on $(30, 65)$, 0.35 probability on $(65, 75)$, and 0.40 probability on $(75, 85)$ years old was used to simulate the continuous covariate "age", and (2) the binary covariates "treatment" and "gender" were both simulated using a 0.5 binomial distribution. We recommend that the reader go to for further details (Rubio et al., 2019).

4.6.3 Posterior Analysis of the Simulated Data

We fitted the proposed PH model with GLL baseline hazard to assess its Bayesian properties in the simulation sets. With all censoring rates and different sample sizes, each simulation set was used to estimate the proposed PH model. JAGS software (Plummer et al., 2003) was used to approximate posterior distributions using three parallel chains with 40,000 iterations each plus another 3,000 for the burn-in period. To minimize autocorrelation in the sequences, the chains were thinned further by storing every 10th draw.

4.6.4 Measures of Performance

The actual mean, standard deviation (SD), Naive standard error, bias, percentage of bias, coverage probability (CP), potential scale reduction factor ($\hat{R}$), and the effective number of separate simulation draws were used to test the posterior distribution stability for the suggested PH model.

4.6.4.1 Evaluating the Performance of the Estimators

We calculate the bias of the estimators using:

$$\text{Bias}(\widehat{\theta}) = \frac{1}{N}\sum_{i=1}^{N}(\widehat{\theta} - \theta). \tag{4.50}$$

An underestimation is indicated by a negative bias, whereas an overestimation is shown by a positive bias.

4.6.5 Accuracy of the Estimators

The mean square error (MSE) is a good indicator of overall accuracy and is calculated as follows:

$$\text{MSE}(\widehat{\theta}) = \frac{1}{N}\sum_{i=1}^{N}(\widehat{\theta} - \theta)^2 \tag{4.51}$$

This metric determines how accurate the estimates are as follows. The lower the MSE, the more accurate the estimations of impacts.

The Naive standard error, which is calculated by dividing the posterior standard deviation by the square root of the sample size, is another accuracy metric. As a result, the smaller the standard error, the larger the sample size. The Naïve SE incorporates simulation error rather than posterior uncertainty.

$$\text{Naive SE} = \frac{\text{posterior SD}}{\sqrt{n}}. \tag{4.52}$$

4.6.5.1 Coverage

The 95 percent coverage probability (CP) is the percentage of N simulated data sets in which the true estimates were included in the 95 percent confidence interval. The more precise the estimations are, the closer the outcome is to a 95 percent coverage probability. The following is how CP is expressed:

$$CP = \widehat{\theta} \mp 1.96 \times SE(\widehat{\theta}) \tag{4.53}$$

4.6.5.2 Convergence Diagnostics

Quantitatively, Gelman (1996) recommended that the acceptable limit of multivariate potential scale reduction factor (MPSRF) and potential scale reduction factor

(PSRF) be near $1\hat{R} < 1.1$, and the effective number of sample size simulation draws be greater than or equal to 100 for checking the convergence of McMC simulations. It is clear from the summary characteristics in Tables 4.1 to 4.4 that the PSRF is less than 1.1, that number of sample size simulation draws is larger than 100, and that Naive SE is smaller than the standard deviations (SD) for all of the distributional parameters and regression coefficients, as expected, indicating that the McMC algorithm has converged to the posterior distribution. Trace plots, auto-correlation plots, and Gelman plot diagnostics are the most common ways to judge the convergence of a McMC simulation graphically (Basharat, 2019). The McMC simulation has been achieved as evidenced by the trace plot, density plot, autocorrelation plot, and Gelman diagnostic plots for each distributional parameter and regression coefficients. That is, the McMC simulation for the GLL PH model explores the target posterior distribution appropriately.

4.6.6 Simulation Results

Tables 4.1-4.4 shows the simulation results for the posterior mean, bias, Naive standard error (SE), mean square error (MSE), coverage probability (CP), Gelman-Rubin diagnostic ($\hat{R}$), and the number of sample size simulation draws (No. of Eff) of the proposed PH model, and Figure 2-Figure 5 shows the visual summary for the convergence diagnostics.

Table 4.1: Simulation results from a GLL PH framework with distributional parameters ($\alpha = 1.5, k = 0.75$, and $\eta = 1.25$), covariates $\beta = (0.75, -0.75, 0.5)$, and $n = 100$

	Posterior properties						
True ($\hat{\theta}$) Value	Posterior mean ($\hat{\theta}$)	Bias	Naïve SE	MSE	CP	$\widehat{R}$	No. of eff
			$C = 20\%$				
$\alpha = 1.50$	1.506	0.006	0.001	0.032	0.032	1.000	3782
$\beta_1 = 0.75$	0.872	0.122	0.002	0.072	0.880	1.000	3823
$\beta_2 = -0.75$	– 0.727	0.023	0.001	0.008	0.945	1.001	3761
$\beta_3 = 0.50$	0.501	0.001	0.002	0.060	0.997	1.000	3700
$\eta = 1.25$	1.575	0.325	0.002	0.193	0.851	1.002	3865
$k = 0.75$	0.567	– 0.183	0.001	0.045	0.911	1.000	4084
			$C = 30\%$				
$\alpha = 1.50$	1.463	– 0.037	0.001	0.029	0.935	1.000	3802
$\beta_1 = 0.75$	0.726	−0.024	0.001	0.023	0.954	1.000	3660
$\beta_2 = -0.75$	−0.752	0.002	0.000	0.004	0.975	1.001	3802
$\beta_3 = 0.50$	0.445	−0.055	0.001	0.023	0.937	1.001	3802
$\eta = 1.25$	1.437	0.187	0.002	0.123	0.911	1.002	4434
$k = 0.75$	0.847	0.097	0.001	0.023	0.907	1.003	4792

Table 4.2: Simulation results from a GLL PH framework with distributional parameters ($\alpha = 1.5, k = 0.75$, and $\eta = 1.25$), covariates $\beta = (0.75, -0.75, 0.5)$, and $n = 300$

	Posterior properties						
True ($\hat{\theta}$) Value	Posterior mean ($\hat{\theta}$)	Bias	Naïve SE	MSE	CP	$\widehat{R}$	No. of eff
			$C = 20\%$				
$\alpha = 1.50$	1.449	−0.001	0.001	0.017	0.991	1.000	4017
$\beta_1 = 0.75$	0.712	−0.038	0.001	0.019	0.946	1.000	3700
$\beta_2 = -0.75$	−0.723	0.027	0.000	0.003	0.956	1.000	3761
$\beta_3 = 0.50$	0.483	−0.017	0.001	0.016	0.962	1.000	3782
$\eta = 1.25$	1.309	0.059	0.002	0.070	0.923	1.001	4609
$k = 0.75$	0.731	−0.019	0.001	0.012	0.941	1.001	4609
			$C = 30\%$				
$\alpha = 1.50$	1.527	0.027	0.001	0.021	0.945	1.000	4174
$\beta_1 = 0.75$	0.726	−0.024	0.001	0.023	0.954	1.000	3660
$\beta_2 = -0.75$	−0.752	0.002	0.000	0.004	0.975	1.001	3802
$\beta_3 = 0.50$	0.445	−0.055	0.001	0.023	0.937	1.001	3802
$\eta = 1.25$	1.437	0.187	0.002	0.123	0.911	1.002	4434
$k = 0.75$	0.847	0.097	0.001	0.023	0.907	1.003	4792

Table 4.3: Simulation results from a GLL PH framework with distributional parameters ($\alpha = 1.75, k = 0.95$, and $\eta = 1.5$), covariates $\beta = (0.75, -0.75, 0.5)$, and $n = 100$

		Posterior properties					
True ($\hat{\theta}$) Value	Posterior mean ($\hat{\theta}$)	Bias	Naïve SE	MSE	CP	$\widehat{R}$	No. of eff
			$C = 20\%$				
$\alpha = 1.75$	1.718	– 0.032	0.002	0.038	0.942	1.000	3865
$\beta_1 = 0.50$	0.523	0.023	0.002	0.051	0.955	1.000	3823
$\beta_2 = –0.85$	– 0.817	– 0.033	0.001	0.010	0.946	1.000	3720
$\beta_3 = 0.50$	0.489	– 0.011	0.002	0.050	0.981	1.000	37402
$\eta = 1.50$	1.441	– 0.059	0.002	0.068	0.931	1.000	4084
$k = 0.95$	0.828	– 0.122	0.001	0.147	0.925	1.001	4084
			$C = 30\%$				
$\alpha = 1.75$	1.717	– 0.033	0.002	0.044	0.939	1.000	3802
$\beta_1 = 0.50$	0.577	0.077	0.002	0.063	0.943	1.000	3823
$\beta_2 = –0.85$	– 0.833	– 0.017	0.001	0.009	0.971	1.000	3761
$\beta_3 = 0.50$	0.474	– 0.026	0.002	0.058	0.952	1.000	3700
$\eta = 1.50$	1.625	0.125	0.002	0.143	0.919	1.002	3865
$k = 0.95$	0.778	– 0.172	0.001	0.213	0.908	1.001	4084

Table 4.4: Simulation results from a GLL PH framework with distributional parameters ($\alpha = 1.75, k = 0.95$, and $\eta = 1.5$), covariates $\beta = (0.75, -0.75, 0.5)$, and $n = 300$

		Posterior properties					
True ($\hat{\theta}$) Value	Posterior mean ($\hat{\theta}$)	Bias	Naïve SE	MSE	CP	$\widehat{R}$	No. of eff
			$C = 20\%$				
$\alpha = 1.75$	1.756	0.006	0.001	0.023	0.978	1.000	3951
β_1=0.50	0.503	0.003	0.001	0.040	0.991	1.000	3761
$\beta_2 = –0.85$	– 0.827	– 0.023	0.000	0.003	0.963	1.000	3761
$\beta_3 = 0.50$	0.505	0.005	0.000	0.045	0.987	1.000	3740
$\eta = 1.50$	1.519	0.019	0.002	0.107	0.942	1.000	4458
$k = 0.95$	0.973	0.023	0.001	0.013	0.941	1.001	4458
			$C = 30\%$				
$\alpha = 1.75$	1.811	0.061	0.001	0.091	0.935	1.000	4011
$\beta_1 = 0.50$	0.612	0.112	0.001	0.129	0.880	1.000	3978
$\beta_2 = –0.85$	– 0.815	– 0.035	0.000	0.004	0.945	1.000	4011
$\beta_3 = 0.50$	0.521	0.021	0.001	0.063	0.997	1.000	3789
$\eta = 1.50$	1.531	0.031	0.002	0.171	0.851	1.001	4458
$k = 0.95$	0.990	0.040	0.002	0.145	0.911	1.002	4565

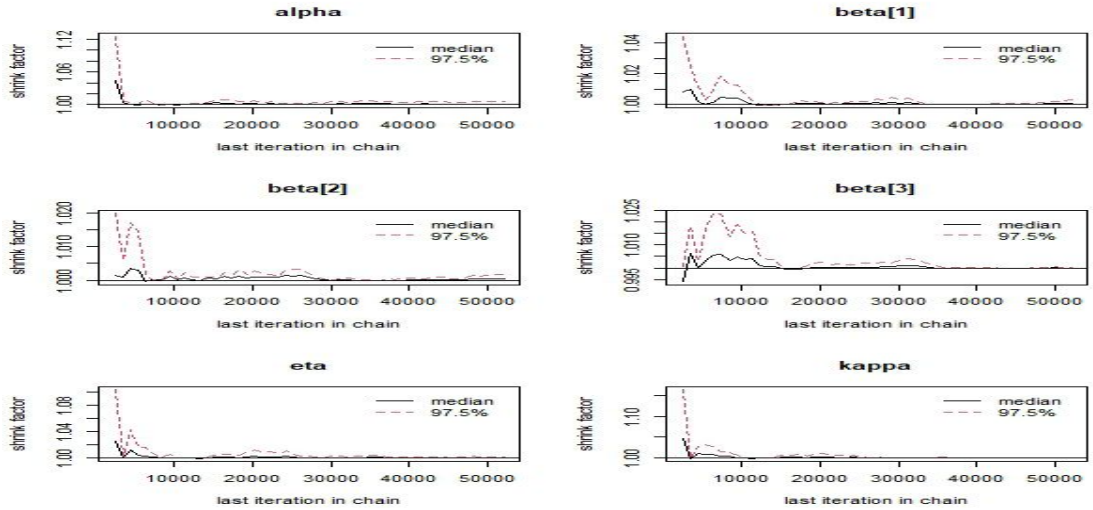

Figure 4.1: Gelman diagnostics from a GLL PH framework with distributional parameters $(\alpha = 1.5, \kappa = 0.75, \text{and} \eta = 1.25)$**, and covariates** $\beta = (0.75, -0.75, 0.5)$**,and** $n = 300$**, and censoring proportion for** 20 **percentage.**

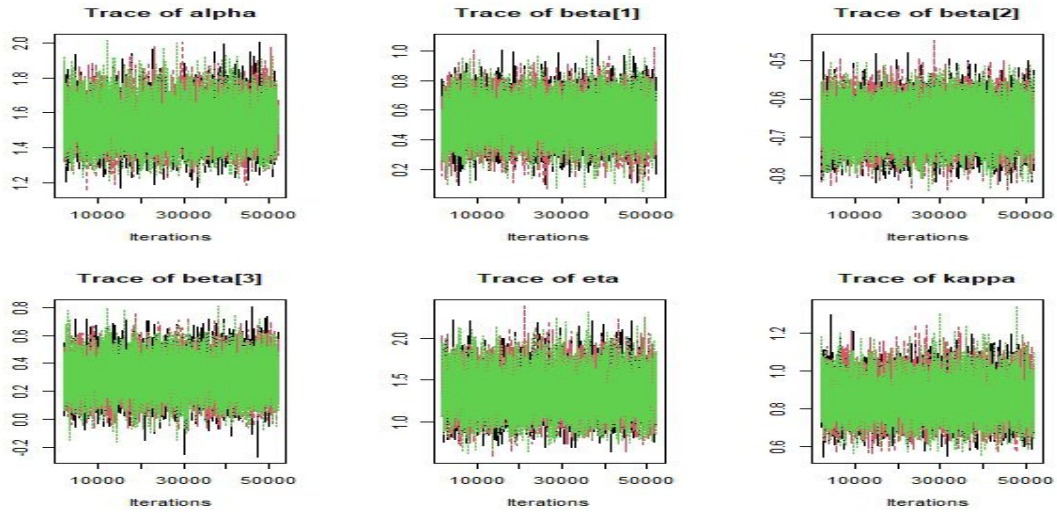

Figure 4.2: Trace plots from a GLL PH framework with distributional parameters $(\alpha = 1.5, \kappa = 0.75, \text{and} \eta = 1.25)$**, and covariates** $\beta = (0.75, -0.75, 0.5)$**,and** $n = 300$**, and censoring proportion for** 20 **percentage.**

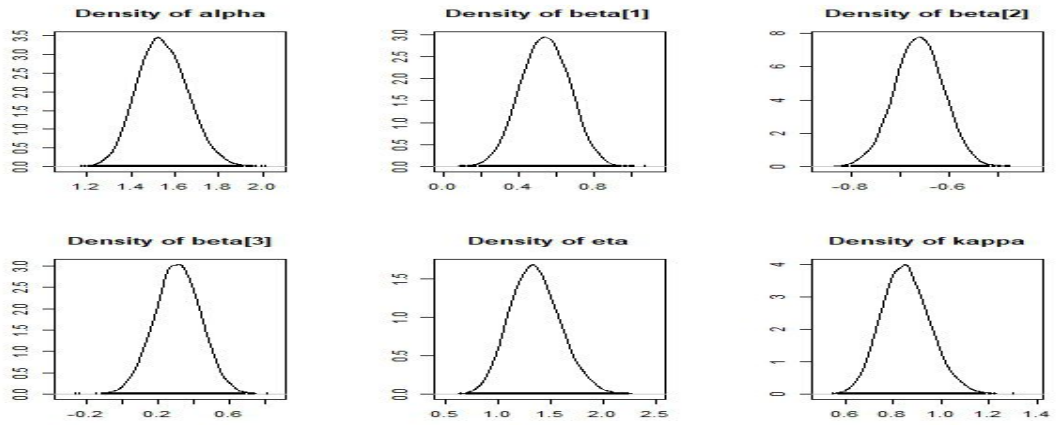

Figure 4.3: Kernel density plots from a GLL PH framework with distributional parameters $(\alpha = 1.5, \kappa = 0.75, \text{and} \eta = 1.25)$**, and covariates** $\beta = (0.75, -0.75, 0.5)$**,and** $n = 300$**, and censoring proportion for** 20 **percentage.**

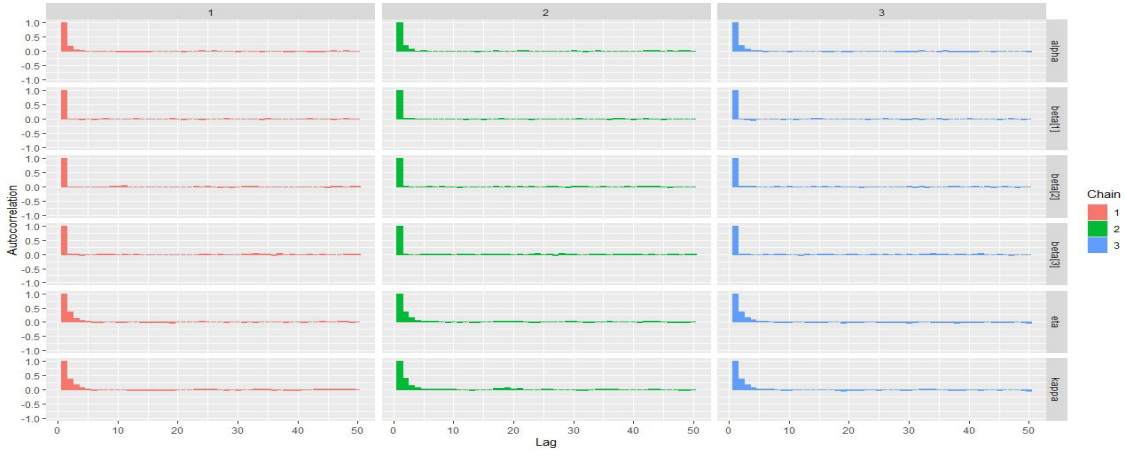

Figure 4.4: Autocorrelation plots from a GLL PH framework with distributional parameters $(\alpha = 1.5, \kappa = 0.75, \text{and} \eta = 1.25)$**, and covariates** $\beta = (0.75, -0.75, 0.5)$**,and** $n = 300$**, and censoring proportion for** 20 **percentage.**

Based on these findings, we may deduce that as the sample size grows, the biases and MSE of the estimators decrease; also, the censoring proportion impacts the bias and MSE of the estimators, with larger censoring rates increasing the bias and MSE. The Gelman-Rubin diagnostic, on the other hand, as well as the number of efficiency sample size draws, show that convergence has been attained. While the estimators' coverage probability was close to 95%.

4.7 Practical Illustrations

In this section, two real-life survival data sets dealing with right-censored cancer data were considered to demonstrate the flexibility and applicability of the proposed GLL PH in modelling different survival data sets with different hazard rate shapes.

4.7.1 Data Set I: Lung Cancer Data

4.7.1.1 Data Description

In this subsection, we reanalyse the data set reported in (Kalbfleisch and Prentice, 2011) which is available in the R package survival. The Veterans Administration Lung Cancer Study Group followed up on n=137 patients in this data set. For this clinical investigation, the censorship rate is around 6.5 percent (9 observations out of 137 were censored). The response and exploratory factors in this clinical trial are the time until death (in days), the number of months from diagnosis to study enrolment (diagt), age (in years), a history of previous lung cancer therapy (prior), and the trt= (treatment = conventional chemotherapy).

4.7.1.2 Hazard Rate Shape

The hazard rate function appears to be unimodal or decreasing in Figure 1 based on the TTT plot (careful inspection reveals a slight indication of unimodality). The data could be evaluated with a model like the log-logistic distribution, which can accommodate decreasing or unimodal hazard rate forms. However, because the classical LL distribution is not closed under the PH framework, we employ the GLL distribution, which is closed and can encompass various hazard rate shapes. The box plot, histogram and TTT plots are shown in Figure 4.5.

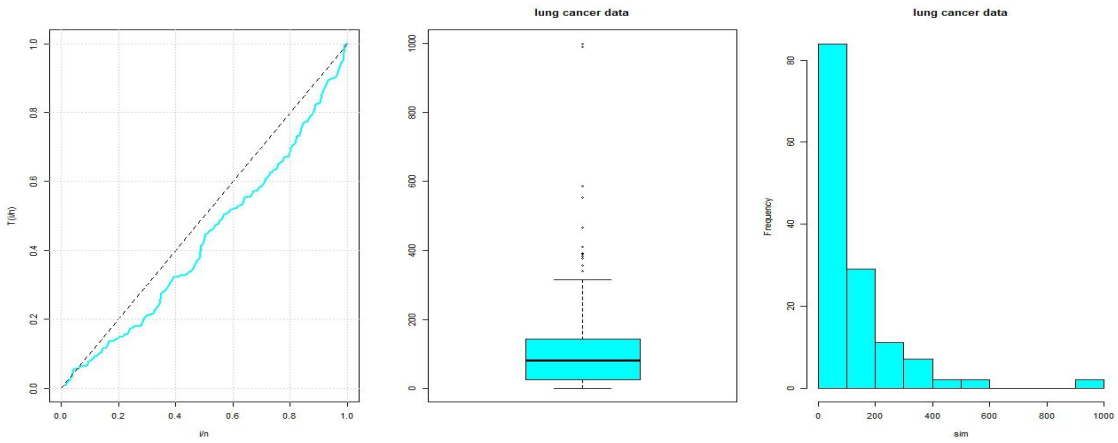

Figure 4.5: TTT plot, box plot, and the histogram for the survival times of the lung cancer data sets

4.7.1.3 Proportionality Assumption

There are two widely used methods for assessing the PH assumption: (1) graphical diagnostics based on (a) time-dependent variables, and (b) standardized Schoenfeld residuals, and (2) statistical tests (Khosa, 2019). The Standardized Schoenfeld residuals are used in this section to evaluate the PH assumption of the Cox model for each covariate included in the model. Based on Figure 4.6 and the significance threshold of 5%, there is no evidence to reject the proportional hazards assumption. As a result, we anticipate that the GLL PH model will provide a good fit when compared to the other existing parametric PH model employed in this study.

4.7.1.4 Posterior Analysis

In this paper, we assume the noninformative independent framework with a normal prior $N(0, 0.001)$ for β/s (regression coefficients) and an independent gamma prior for the distributional parameters $\alpha \sim G(a_1, b_1), \eta \sim G(a_2, b_2)$, and $k \sim G(a_3, b_3)$ with hyperparameter values $(a_1 = b_1 = a_2 = b_2 = a_3 = b_3 = 10)$.

(1) Numerical Summary. We looked at various quantities of interest and their numerical values using the McMC sample of posterior properties for the generalized

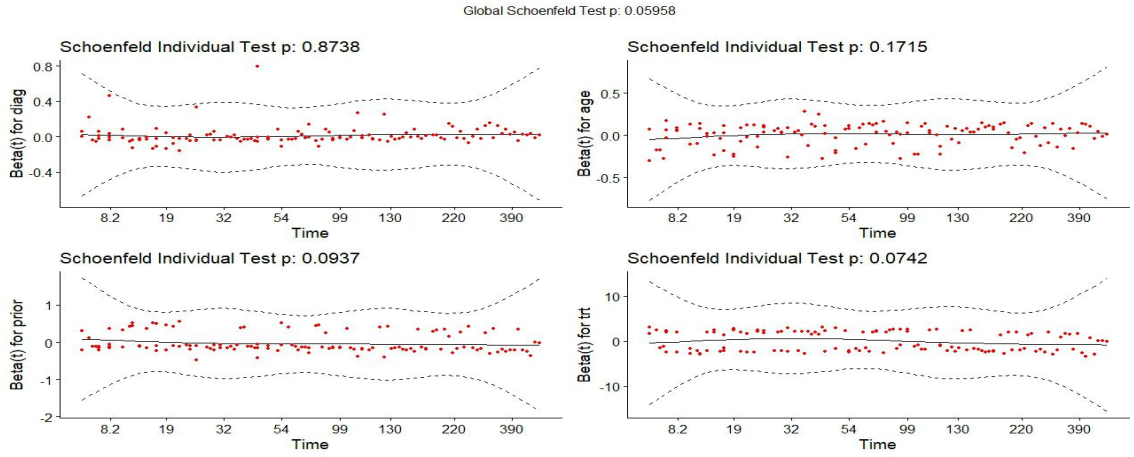

Figure 4.6: The Standardized Schoenfeld residuals from the data I - lung cancer data set, taking the test p-value for each covariate into ac-count.

log-logistic proportional hazard model using the lung cancer data in this section. The posterior summaries for the generalized log-logistic PH model parameters using Veterans lung cancer data sets are illustrated in Table 4.5 . The probability that the corresponding parameter is $+ve$ is given in the last row of Table 4.5. A posterior probability of 0.5 indicates that a positive parameter value is as likely as a negative one. Once we've saved the posterior sample for each model parameter, we can compute the posterior probability, for example, for β_1, using mean $(\beta_1 > 0)$.

(2) Visual Summary. We looked at trace plots, Gelman-Rubin diagnostic plots, and Ergodic mean diagnostic plots in this section to get a visual description of the posterior properties. These plots and graphs provide a nearly comprehensive representation of the parameters' posterior uncertainty for the application of the lung cancer data sets. Figure 4.7 shows that the McMC sampling process converges to the joint posterior distribution with no periodicity. As a result, we can say that the chains have converged. Figure 4.8 shows us that both PSRF and MPSRF are less than 1.1.

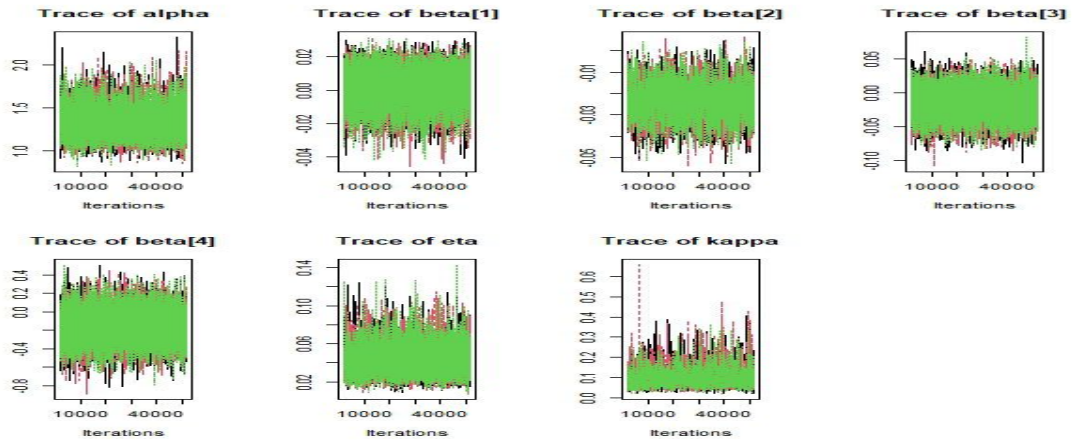

Figure 4.7: The time series plots for the baseline hazard parameters and the regression coefficients for the Veterans lung cancer data..

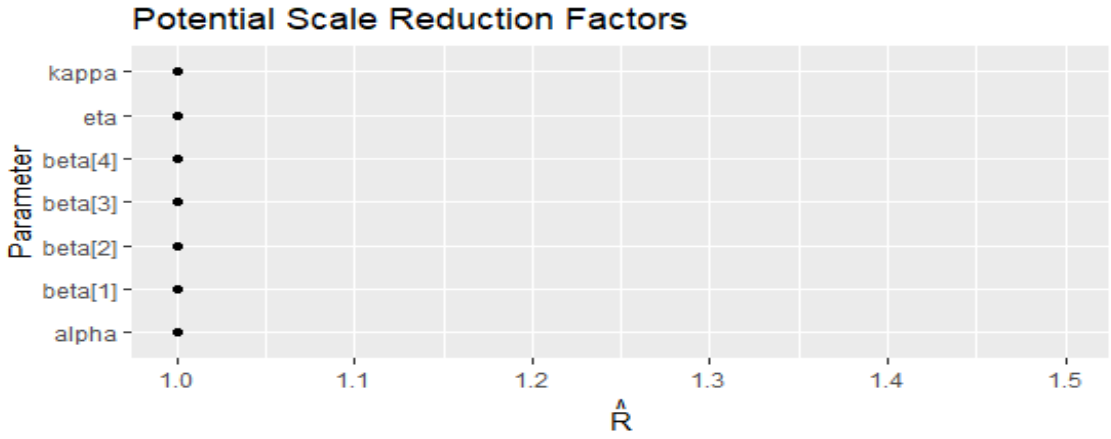

Figure 4.8: PSRF of the baseline hazard parameters and the regression coefficients for the Veterans lung cancer data.

Figure 4.9 shows us the Ergodic mean plots for the regression coefficients and the baseline hazard parameters. It is quite clear from the running mean time-series plots that the chains converge after N iterations to their mean values. However, these plots display only at the mean of the baseline hazard parameters and the regression coefficients and therefore are inadequate.

Table 4.5: Numerical summaries of posterior characteristics based on McMC sample of the GLL-PH model for the lung cancer data set.

	Pars						
Characteristics	Alpha	β_1 (diagt)	β_2(age)	β_3 (prior)	β_4(trt)	Eta	Kappa
Mean	1.317	0.002	−0.024	−0.015	−0.151	0.042	0.103
SD	0.173	0.010	0.008	0.021	0.178	0.015	0.049
Naïve SE	0.001	0.0001	0.0001	0.0002	0.001	0.0001	0.0004
Time series SE	0.003	0.0001	0.0001	0.0001	0.002	0.0002	0.001
Minimum	0.813	−0.046	−0.054	−0.109	−0.890	0.007	0.019
2.5 th percentile	1.023	−0.020	−0.040	−0.057	−0.500	0.019	0.040
Q1	1.194	−0.005	−0.029	−0.029	−0.271	0.031	0.068
Medium (Q2)	1.302	0.003	−0.024	−0.015	−0.150	0.040	0.092
Q3	1.422	0.010	−0.018	−0.0003	−0.029	0.051	0.125
97.5 th percentile	1.697	0.021	−0.007	0.027	0.193	0.078	0.231
Maximum	2.324	0.032	0.006	0.082	0.511	0.143	0.658
Mode	1.250	0.003	−0.028	−0.015	−0.150	0.035	0.075
Variance	0.030	0.0001	0.0001	0.001	0.032	0.0002	0.002
Skewness	0.550	−0.361	0.082	−0.058	−0.027	0.957	1.656
Kurtosis	0.558	0.152	0.011	0.001	−0.009	1.510	4.992
95% credible interval	1.023, 1.697)	−0.020.0.021)	(−0.040, −0.007)	(−0.057, 0.027)	(−0.500, 0.193)	(0.019, 0.078)	(0.040, 0.231)
$P(.>0\|data)$	1.000	0.598	0.003	0.244	0.199	1.000	1.000

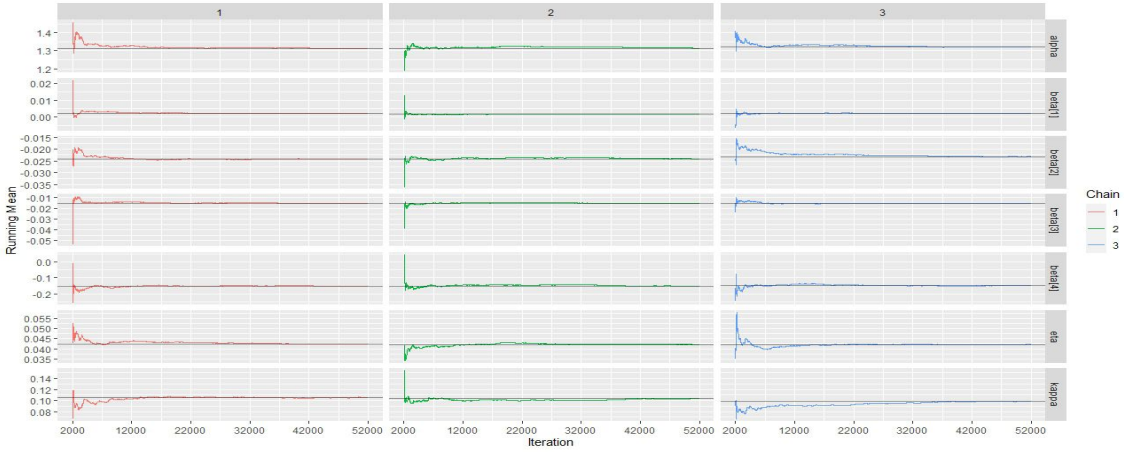

Figure 4.9: The running mean plots for the baseline hazard parameters and regression coefficients for the Veterans lung cancer data set.

4.7.2 Data Set II: Larynx Cancer Data

4.7.2.1 Data Description

Lifetimes for 90 patients with larynx-cancer, according to the stage of cancer tumour (stages 1-4) are given in Table 7. The study time or time to death are recorded in months (where, * shows us the censored time). Alvares et al. (2021); Christensen et al. (2010); Wang et al. (2018b) discussed the data from different aspects under different hazard-based regression models, and the data were first reported by (Kar-

Table 4.6: Survival times (in months) of patients with larynx cancer according to stages of tumour (1–4).

Stages	Survival time (* = indicating censoring)
Stage I (33 patients)	$0.6, 1.3, 2.4, 2.5^*, 3.2, 3.3^*, 3.5, 3.5, 4.0,$
	$4.0, 4.3, 4.5^*, 4.5^*, 5.3, 5.5^*, 5.9^*, 5.9^*, 6.0^*, 6.1^*, 6.2^*,$
Stage II (17 patients)	$6.4, 6.5, 6.5^*, 6.7^*, 7.0^*, 7.4, 7.4^*, 8.1^*, 8.1^*,$
	$9.6^*, 10.7^*$
Stage III (patients)	$0.2, 1.8, 2.0, 2.2^*, 2.6^*, 3.3^*, 3.6, 4.0^*, 4.3,$
	$4.3^*, 5.0^*, 6.2, 7.0, 7.5^*, 7.6^*, 9.3^*$
Stage IV (13 patients)	$0.3, 0.3, 0.5, 0.7, 0.8, 1.0, 1.3, 1.6, 1.8, 1.9, 1.9, 3.2,$
	$3.5, 3.7^*, 4.5^*, 4.8^*, 4.8^*, 5.0^*, 5.0^*, 5.1^*, 6.3, 6.4, 6.5$

daun, 1983). The survival times (in months) of patients is illustrated in Table 4.6. The other covariates of the data are as follows: (1) age (in years) at diagnosis and (2) the year of diagnosis. One goal of this study was to see if the age, year of diagnosis, and stage of cancer were associated with the death of patients with laryngeal cancer.

4.7.2.2 Hazard Rate Shape

Based on the TTT plot in Figure 4.10, the hazard rate function is an increasing hazard in Figure 13. The data could be analyzed using a model such as the Weibull distribution, which can handle monotone hazard rate forms. We adopt the GLL distribution, which would be represented by the PH framework and can accommodate a variety of hazard rate shapes to see its applicability of the monotone (increasing) hazard rates. Figure 4.10 shows the box plot, histogram, and TTT plots.

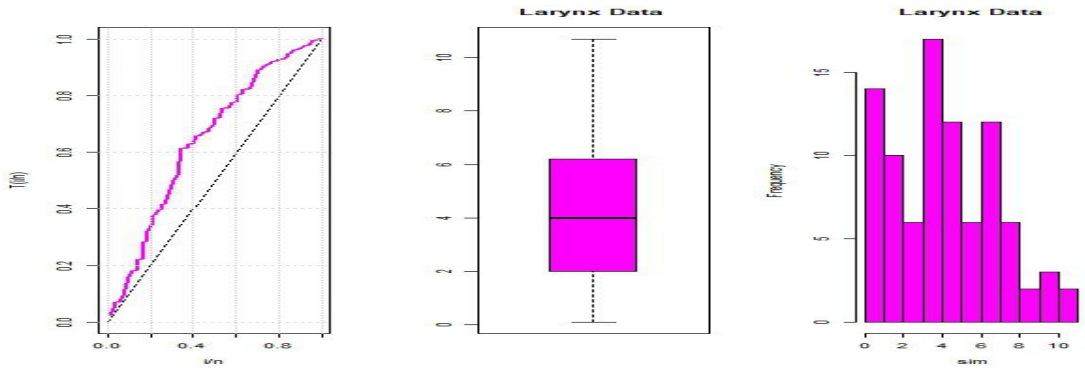

Figure 4.10: TTT plot, box plot, and the histogram for the survival times of the larynx cancer data set.

4.7.2.3 Proportionality Assumption

We investigated if the proportional hazards model could be used with this data set. The underlying assumption of the Cox model for each explanatory variable utilized in the model is depicted in Figure 4.11. With a significance level of 5%, there is no evidence to reject the PH assumption. As a result, we anticipate that the parametric PH model will provide a strong fit.

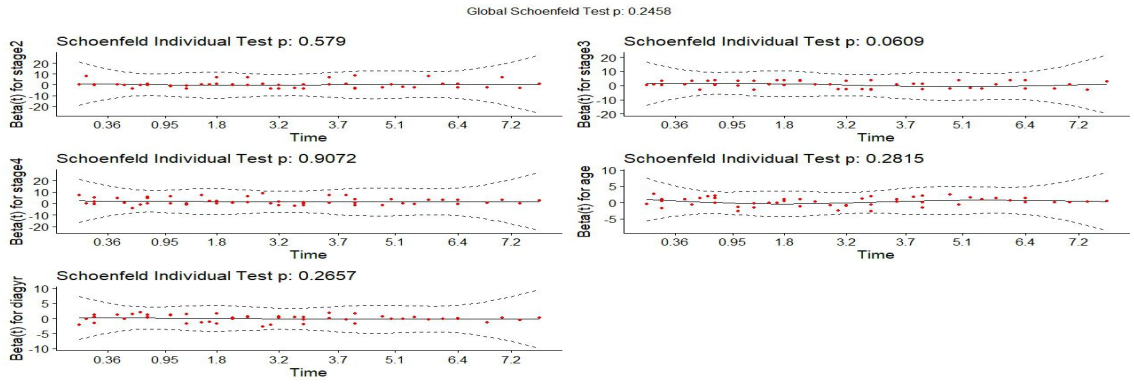

Figure 4.11: The standardized Schoenfeld residuals from the data II larynx cancer data set, taking the test p value for each covariate into account.

4.7.3 Posterior Analysis

In this chapter, we assume the noninformative independent framework with $N(0, 0.001)$ for β 's (regression coefficients) and an independent gamma prior for the distribu-

Table 4.7: Numerical summaries of posterior characteristics based on McMC sample for GLL PH model for the larynx cancer data.

Characteristics	Pars							
	Alpha	β_1 (stage 2)	β_2 (stage 3)	β_3 (stage 4)	β_4 (age)	β_5 (diagyr)	Eta	Kappa
Mean	1.539	-0.182	0.376	1.222	0.187	-0.111	0.869	0.336
SD	0.215	0.454	0.337	0.411	0.144	0.149	0.247	0.077
Naïve SE	0.002	0.004	0.003	0.004	0.001	0.001	0.002	0.001
Time series SE	0.002	0.004	0.003	0.004	0.001	0.001	0.003	0.001
Minimum	0.847	-1.975	-0.902	-0.531	-0.373	-0.730	0.197	0.112
2.5 th percentile	1.157	-1.108	-0.289	0.396	-0.091	-0.403	0.457	0.207
Q1	1.389	-0.480	0.152	0.952	0.089	-0.212	0.691	0.282
Medium (Q2)	1.524	-0.170	0.377	1.230	0.187	-0.112	0.846	0.328
Q3	1.668	0.128	0.605	1.498	0.284	-0.012	1.020	0.382
97.5 th percentile	2.005	0.667	1.030	2.010	0.476	0.181	1.412	0.507
Maximum	2.701	1.648	1.770	2.848	0.817	0.509	2.131	0.763
Mode	1.550	-0.100	0.300	1.300	0.150	-0.150	0.850	0.325
Variance	0.046	0.207	0.113	0.169	0.021	0.022	0.061	0.006
Skewness	0.447	-0.173	-0.041	-0.086	0.081	0.023	0.595	0.604
Kurtosis	0.511	0.068	0.010	0.070	0.102	0.027	0.514	0.656
95% Credible interval	(1.157, 2.005)	(-1.108, 0.667)	(-0.289, 1.030)	(0.396, 2.010)	(-0.091, 0.476)	(-0.730, 0.181)	(0.197, 1.412)	(112, 0.507)
P	1.000	0.352	0.870	0.998	0.906	0.227	1.000	1.000

tional parameters $\alpha \sim G(a_1, b_1)$, $\eta \sim G(a_2, b_2)$, and $k \sim G(a_3, b_3)$ with hyperparameter values $(a_1 = b_1 = a_2 = b_2 = a_3 = b_3 = 10)$.

4.7.3.1 Numerical Summary.

We looked at various quantities of importance as well as their numerical values using the McMC sample of posterior properties for the generalized log-logistic proportional hazard model considering the larynx data in this section.

The posterior summaries for the GLL-PH model parameters using larynx cancer data are illustrated in Table 4.7. The probability that the corresponding parameter is $+ve$ is given in the last row of Table 4.7.

4.7.3.2 Visual Summary

We looked at autocorrelation plots in Figure 4.12, and Ergodic mean plots in Figure 4.13 for the proposed model parameters which shown us that the convergence was attained.

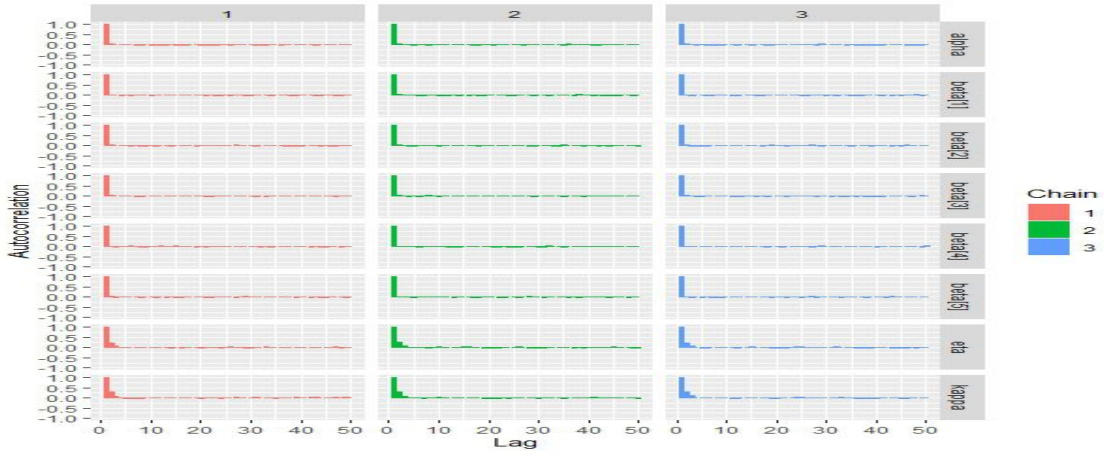

Figure 4.12: Autocorrelation plots for all the baseline hazard parameters and regression coefficients for the larynx cancer data.

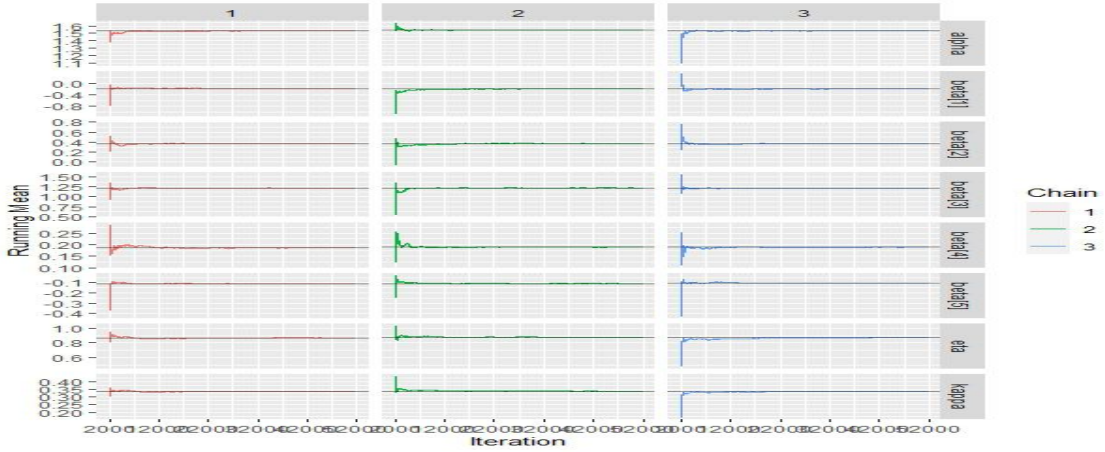

Figure 4.13: The Ergodic mean plots for the baseline hazard parameters and regression coefficients for the larynx cancer data..

4.7.4 Hazard Ratio

One of the most intriguing aspects of PH models is that the regression coefficients can be interpreted using the hazard ratio, which is preferred by many clinicians.

A key feature for PH models is the hazard ratio (HR), also known as the relative risk, between two individuals with covariate vectors $\mathbf{x}_1$ and $\mathbf{x}_2$. The HR is defined as:

$$HR\left(\mathbf{x}_1, \mathbf{x}_2, h_0, \boldsymbol{\beta}\right) = \frac{h\left(t \mid \mathbf{x}_1, h_0, \boldsymbol{\beta}\right)}{h\left(t \mid \mathbf{x}_2, h_0, \boldsymbol{\beta}\right)} = \exp\left[\left(\mathbf{x}_1 - \mathbf{x}_2\right)^{\mathrm{T}} \boldsymbol{\beta}\right] \tag{4.54}$$

Table 4.8: Posterior characteristics of the hazard ratio between two men of the same age and diagnosis year but in different tumour stages

Posterior Properties	Stages 3 and 4	Stages 2 and 4	Stages 2 and 3
Mean	0.467	0.280	0.638
Standard deviation	0.203	0.149	0.298
Naïve SE	0.001	0.001	0.002
Time series SE	0.002	0.001	0.002
2.5% percentile	0.197	0.088	0.218
Q1	0.326	0.175	0.423
Medium (Q2)	0.425	0.250	0.585
Q3	0.564	0.349	0.791
97.5 percentiles	0.967	0.648	1.366

which does not depend on time t. the hazard function in the numerator is equal to this constant HR times the hazard in the denominator, i.e.,

$$h\left(t \mid \mathbf{x}_1, h_0, \beta\right) = HRxh\left(t \mid \mathbf{x}_2, h_0, \beta\right)$$

Hence, the name "proportional hazards model" (Christensen et al., 2010). For example, the posterior distributions of the HR between two individuals of the same age and diagyr (year of diagnosis) but in different stages can be easily summarized. Table 4.8 depicts the posterior characteristics of the hazard ratio between two men of the same age and diagnosis year (diagyr) but in different stages.

4.8 Bayesian Model Selection

In this study, we used the deviance information criterion (DIC) to distinguish between the proposed models.

4.8.1 Data Set I

Table 4.9 displays some posterior characteristics for the three PH models (generalized log-logistic, Gompertz, and Weibull). Even though the estimates of the regression coefficient are significant compared, the flexibility provided by the GLL distribution's additional shape parameter contributes to its ultimate superiority over the Gompertz and Weibull models and the DIC shows us its goodness-of-fit and versatility comparing to the competing parametric PH models.

Table 4.9: Posterior properties summaries and the information criterion values for the considered GLL PH model and its competing models for the lung cancer data.

Summaries		Posterior characteristics			
Parametric competitive models	Parameter(s)	Posterior mean	Posterior SD	Pr (>|0|data)	HPD interval (95%)
GLL-PH model (DIC = 1505.165)					
	Alpha	1.317	0.173	1.000	(1.001, 1.661)
	β_1 (diagt)	0.002	0.010	0.598	(−0.019, 0.021)
	β_2 (age)	−0.024	0.008	0.003	(−0.039, −0.007)
	β_3 (prior)	−0.015	0.021	0.244	(−0.057, 0.026)
	β_4 (trt)	−0.151	0.178	0.199	(−0.505, 0.186)
	Eta	0.042	0.015	1.000	(0.016, 0.073)
	Kappa	0.103	0.049	1.000	(0.029, 0.200)
Weibull-PH model (DIC = 1521.310)					
	Alpha	0.744	0.048	1.000	(0.654, 0.842)
	β_1 (diagt)	0.005	0.010	0.648	(−0.018, 0.024)
	β_2 (age)	−0.025	0.007	0.001	(−0.039, −0.010)
	β_3 (prior)	−1.027	0.021	0.102	(−0.068, 0.015)
	β_4 (trt)	−0.252	0.180	0.080	(−0.593, 0.108)
	Kappa	0.206	0.090	1.000	(0.060, 0.388)
Gompertz-PH model (DIC = 1556.407)					
	Alpha	1.134	0.311	1.000	(0.567, 1.746)
	β_1 (diagt)	0.021	0.009	0.984	(0.003, 0.039)
	β_2 (age)	0.027	0.006	1.000	(0.014, 0.039)
	β_3 (prior)	−0.056	0.023	0.006	(−0.099, −0.012)
	β_4 (trt)	−0.136	0.182	0.228	(−0.494, −0.211)
	Kappa	0.001	0.0002	1.000	(0.001, 0.002)

4.8.2 Data Set II

Table 4.10 displays some posterior characteristics for the three PH models (generalized log-logistic, Gompertz, and Weibull). Even though the estimates of the regression coefficient are significant compared, the flexibility provided by the GLL distribution's additional shape parameter contributes to its ultimate superiority over the Gompertz and Weibull models and the DIC demonstrates us ts goodness-of-fit and versatility comparing to the competing parametric PH models.

Table 4.10: Posterior properties summaries and the information criterion values for the considered GLL PH model and its competing models for the larynx cancer data..

Summaries		Posterior characteristics			
Parametric competitive models	Parameter(s)	Posterior mean	Posterior SD	Pr (>|0|data)	HPD interval (95%)
GLL-PH model (DIC = 294.412)					
	Alpha	1.539	0.215	1.000	(1.157, 2.005)
	β_1 (stage 2)	−0.182	0.454	0.352	(−1.108, 0.667)
	β_2 (stage 3)	0.376	0.337	0.870	(−0.289, 1.030)
	β_3 (stage 4)	1.222	0.411	0.998	(0.396, 2.010)
	β_4 (age)	0.187	0.144	0.906	(−0.091, 0.476)
	β_5 (diagyr)	−0.111	0.149	0.227	−0.730, 0.181)
	Eta	0.869	0.247	1.000	(0.197, 1.412)
	Kappa	0.336	0.077	1.000	(0.112, 0.507)
Weibull-PH model (DIC = 296.776)					
	Alpha	0.908	0.105	1.000	(0.713, 1.118)
	β_1 (stage 2)	−0.380	0.446	0.198	(−1.275, 0.468)
	β_2 (stage 3)	0.174	0.318	0.711	(−0.483, 0.781)
	β_3 (stage 4)	1.095	0.393	0.997	(0.329, 1.857)
	β_4 (age)	0.176	0.141	0.899	(−0.092, 0.461)
	β_5 (diagyr)	−0.012	0.146	0.468	(−0.294, 0.274)
	Kappa	0.154	0.041	1.000	(0.081, 0.236)
Gompertz-PH model (WAIC = 297.560)					
	Alpha	0.134	0.031	1.000	(0.076, 0.196)
	β_1 (stage 2)	−0.138	0.455	0.392	(−1.040, 0.737)
	β_2 (stage 3)	0.393	0.328	0.886	(−0.252, 1.041)
	β_3 (stage 4)	1.544	0.397	1.000	(0.776, 2.308)
	β_4 (age)	0.206	0.149	0.919	(−0.084, −0.501)
	β_5 (diagyr)	0.075	0.155	0.685	(−0.230, 0.374)
	Kappa	0.552	0.186	1.000	(0.227, 0.919)

4.9 Concluding Remarks

In this chapter, we explored how to derive Bayesian estimates of the baseline hazard parameters and the regression coefficients of the parametric proportional hazard model with generalized log-logistic baseline hazard using right-censored survival data utilizing McMC approaches. The McMC techniques offer an alternative technique for estimating the parameters of the proposed model that is more flexible than frequentist techniques such as maximum likelihood estimation. Bayesian Inference was performed with a variety of priors, and the convergence pattern was investigated using various diagnostic procedures.

To test the performance of the proposed parametric PH model, a comprehensive McMC simulation study was conducted. According to the simulation results, the PH model produces better results, with fewer absolute biases and MSEs for most regression coefficients and baseline distributional parameters. The behavior of the

PH model in a generic PH regression situation comprising numerous covariates was also examined using synthetic right-censored data sets. Our findings indicate that the PH model performs well when handling with multiple factors.The paper's final analysis focused on a real-world application involving two well-known right-censored survival data sets for lung cancer and laryngeal cancer patients. In conclusion, the findings of the proposed parametric PH model show that it performs better and is superior to the other competing PH model, as well as indicating significant distributional parameters and regression coefficients.

Furthermore, for both simulation and real-data analysis, the regression coefficients were assumed to have a normal prior, and the baseline distribution parameters were assumed to have an independent gamma prior to compute the quantities of importance derived from the proposed model's posterior distribution. It has been attempted to create a visual summary and other essential graphs to aid in the interpretation of results and decision making. Finally, we hope that this paper will be an extension of the work of Khan and Khosa (2016) and will encourage researchers who employ parametric hazard-based regression models to conduct their analyses using the Bayesian approach from the BUGs codes with the help of the R software's rjags package.

CHAPTER 5

Bayesian and Frequentist Approaches for the Generalized Log-logistic Accelerated Failure Time Model

In this chapter, we present our fourth published manuscript [1] about the Bayesian and frequentist approaches for the generalized log-logistic accelerated failure time model. Note that the materials of this chapter have been reproduced from our article.

5.1 Introduction

Certain continuous probability distributions are extensively applied to model survival data. The most frequently used parametric survival models are the log-normal, log-logistic, gamma and Weibull distributions. The acceptance of these distributions in time-to-event analysis is largely due to their (i) tendency to adequately model data which are frequently encountered in time-to-event analysis, (ii) model closeness, (iii) readily available statistical software package, (iv) common structure, since all of these distributions are of the log-location-scale class, (v) straight forwardness of the approach (Khosa, 2019). The log-logistic and the log-normal families are used to model non-monotone (or unimodal) hazard rates, whereas the Weibull family is widely applied to model monotone hazard functions (Lawless, 2011).

The log-logistic distribution is by far the most popular used model for analyzing

[1] Muse, A. H., Mwalili, S., Ngesa, O., Alshanbari, H. M., Khosa, S. K., and Hussam, E. (2022). Bayesian and frequentist approach for the generalized log-logistic accelerated failure time model with applications to larynx-cancer patients. Alexandria Engineering Journal, 61(10), 7953-7978.

and modeling survival times with unimodal hazard shapes. Its major weakness is its inability to incorporate monotone failure rates (i.e., in particular increasing failure rates). This has led to the need to seek modifications and generalizations of the log-logistic distribution. Due to the restriction of the log-logistic distribution in analyzing data with monotone hazard rate function. Muse et al. (2021c) proposed a new two-parameter LL model that accommodate for both monotone and unimodal hazard shapes using the tangent function; Aldahlan (2020) proposed the Alpha power transformed LL distribution; Adeyinka and Olapade (2019) introduced the transmuted four-parameters generalized LL distribution; Lima and Cordeiro (2017) extended the LL distribution by applying the exponentiated generalized generator approach; Mendoza et al. (2016) developed an exponentiated log-logistic geometric distribution, among others. The failure rate function allows for incorporating constant, monotonically decreasing, monotonically increasing, bath-tub shaped and unimodal hazard rates when the classical LL model was modified by adding extra shape parameters (Cordeiro et al., 2020). More details about the extensions and recent modifications of the LL distribution can be found in (Muse et al., 2021b).

The applications of the generalized distributions in the context of parametric survival models have been widespread. Many generalizations and modifications of the log-logistic, Weibull and even other common classical distributions have been developed to make these more flexible in fitting a range of failure rate shapes (Cordeiro et al., 2020; Muse et al., 2021c; Nasiru et al., 2019; Ahmad et al., 2019; Muse et al., 2021b; Tahir and Nadarajah, 2015; Tahir and Cordeiro, 2016; Dey et al., 2021). One such model is the generalized log-logistic distribution which extends the log-logistic by adding an extra shape parameter Khan and Khosa (2016). The model is flexible and sparse in the sense that it fitted to data sets whose failure rate function is monotone and non-monotone and has some models including the Weibull, Burr XII, log-logistic and exponential distributions, see (Muse et al., 2021a). It could also be used to examine the goodness of fit of sub-models such as log-logistic, Burr XII,

exponential and Weibull .

Parametric hazard-based regression models focus on the association between hazard rates or survival time random variables and the explanatory variables (risk factors, or predictor variables, or covariates). There are two popular approaches for parametric regression of survival data: proportional hazard (PH) models and accelerated failure time (AFT) models (Wang et al., 2018b). There are other two approaches for parameteric survival models: accelerated hazard model (Chen and Wang, 2000), and the proportional odds model (Bennett, 1983a). AFT model is a more realistic and valuable alternative to the parametric PH model in some situations and also under some probability distributions and their generalizations (Khan and Basharat, 2021).

The underlying shapes that commonly employed parametric AFT models may capture are limited (Crowther, 2014). Many researchers are currently focusing on the construction of flexible parametric AFT models utilizing various methodologies (Pang, 2020).The development of new modified distributions as a baseline hazard function for an AFT model is one of the most prevalent methods. Khan (2018) introduced the exponentiated Weibull AFT model, which can handle monotone and non-monotone hazard rate forms. Using Bayesian inference, Ashraf-Ul-Alam and Khan (2021) developed the generalized Topp-Leone Weibull AFT model. These distributions frequently have a small number of sub-models. Furthermore, our proposed baseline hazard distribution includes both Weibull and Log-logistic AFT models as a special cases, as well as it has the ability to accommodate various hazard rate forms. Hence, in this study, we extend the generalized log-logistic PH model (Khan and Khosa, 2016) into a generalized log-logistic AFT model to be a convenient alternative to the classical Weibull distribution since it can accommodate for both monotone and non-monotone hazard rates, as well as it is closed under both PH and AFT frameworks.

Weibull distribution is the only probability distribution that is closed under both

PH and AFT families (Kalbfleisch and Prentice, 2011). The LL distribution has wide applications in analyzing time-to-event data. The model is closed under both accelerated failure time (AFT) and proportionality of odds (PO) but it is not closed under the proportional hazard (PH) framework. However, when it is generalized it has the attractive feature to of being a member of all hazard-based regression models because its cumulative distribution function (cdf) has a closed form and its hazard rate function is quite flexible. Therefore, Khan and Khosa (2016) developed a generalized log-logistic that is closed under the PH framework.

Hence, the main goal of this chapter is to develop a generalized log-logistic AFT model. The Bayesian and maximum likelihood estimation approaches for estimating the model parameters are considered to estimate the distributional parameters and covariates of the proposed model. We evaluated the performance of the generalized log-logistic AFT model in comparison with the most frequently used lifetime AFT models, namely log-logistic, and Weibull. Furthermore, convergence diagnostic techniques based on the McMC approach were used.

As a result, in light of the above discussion, and the growing need to include explanatory varaibles in the analysis of survival data we primarily focus on parametric AFT models which need a distributional assumption for T given a vector of explanatory variables X. In particular, we present AFT approach based on the generalized log-logistic distribution. The most appealing aspects of this technique may be its flexibility in accepting both monotone and non-monotone failure rate functions, as well as the fact that it can do so at the low cost of estimating one additional parameter compared to other frequently used parameteric regression models (e.g., log-normal, Weibull, and log-logistic) and also the distribution is closed under the PH framework.

Hence, proposing and showing that the generalized log-logistic is also closed under the AFT framework is a great contribution to the field of survival analysis. Estimating the model's parameters using both Bayesian and maximum likelihood estimation

techniques and developing an extensive simulation study to assess the model's performance by comparing the proposed model's fit to the fits of other frequently used parametric regression models including log-logistic and Weibull model in analyzing various types of survival data with monotone and non-monotone failure rate shapes under different censoring proportions is the novelty of this work.

The rest of the chapter is classified as follows. Section 5.2, presents the AFT model formulation and its assumptions. In Section 5.3, we show that the generalized log-logistic distribution is closed under the AFT framework. Based on this outcome, we then derive the common lifetime distribution functions of the generalized log-logistic AFT model. The common sub-models closed under the proposed model are also discussed. The inferential procedures of the proposed model including the maximum likelihood estimation and the Bayesian approach for inference is also discussed in Section 5.4. An extensitve simulation study is presented in Section 5.5. In Section 5.6, the generalized log-logistic AFT model is applied into a real-life survival data relatig to larynx-cancer patients and the posterior characteristics summaries and graphs of the parameters and the regression coefficients are provided with the support of the ggmcmc package (Fernández-i Marín, 2016). The model comparison is presented in Section 5.7 by comparing its sub-models Weibull, and log-logistic. In Section 5.8, Markov chain Monte Carlo (McMC) algorithm (Gelman et al., 1995) for the proposed AFT model is demonstrated using BUGS syntax (Gilks et al., 1994), with the assistance of JAGS (Plummer et al., 2003) and the rjags package (version 4-10) (Plummer et al., 2019; Qi et al., 2020), for the R language (version 4.0.2) (Team, 2019) and the convergence diagnostics were discussed briefly. Some remarkable conclusion of our findings in Section 5.9.

5.2 The AFT Model Formulation and Its Assumptions

The Formulation of the parametric AFT model using a defined baseline distribution and a link function for covariates is defined as follows:

$$h(t;x) = h_0(t)\psi(x'\beta) \tag{5.1}$$

where the h_0 is the baseline hazard and the most popular natural choice for the link function $\psi(\boldsymbol{x}'\boldsymbol{\beta})$ is the log-linear (or exponential) function $\exp(\boldsymbol{x}'\boldsymbol{\beta})$. In this chapter, we describe the AFT model assuming that $\psi(\boldsymbol{x}'\boldsymbol{\beta}) = \exp(\boldsymbol{x}'\boldsymbol{\beta})$

5.2.1 Assumptions

The assumption of the AFT model is that the effect of covariates acts multiplicative on the survival time. The assumption of the AFT model can be defined as,

$$S(t \mid x) = S_0[t\psi(x)] \quad \text{for } t \geq 0 \tag{5.2}$$

where; $S_0[t\psi(x)]$ is the baseline survival function (i.e., survival function for an individual with $\mathbf{x} = \mathbf{0}$), $S(t \mid x)$ is the survival function at the time t, and the $\psi(x)$ is the link function. The covariates are linked to the lifetime by $\psi(x)$, satisfying

$$\psi(0) = 1 \text{ and } \psi(x) > 0 \forall x \neq 0$$

with these attributes of $\psi(x), x = 0$ implies $S(t \mid x) = S_0(t \mid x)$.

$$\psi(x) = \exp\left(\beta^{/} x\right) \tag{5.3}$$

The vector β represents regression coefficients and x is a vector of non-random regressors. With 5.3 the covariates accelerate ($\beta' x > 0$) or decrease ($\beta' x < 0$) the rate at which a unit moves through time with respect to the baseline case.

5.2.2 Lifetime Distribution Functions for the AFT Model

In this sub-section, we derive the common probabilistic functions of the accelerated failure time model. The AFT model can be expressed as:

$$S(t \mid x) = S_0 \left(te^{\boldsymbol{x}^{*'}\boldsymbol{\beta}^*} \right) for t \geq 0 \tag{5.4}$$

Other, less popular choices for the link function are

$$\psi(x) = \beta' x \text{ and } \psi(x) = (\beta' x)^{-1}$$

With these two choices it may happen that $\psi(x) < 0$ for some value of β resulting into a negative lifetime.

The other four common life time distribution representatives for an AFT model with 5.4 can be expressed as: The cumulative distribution function of the AFT model is given by:

$$F(t) = 1 - S(t) = 1 - S_0[t\psi(x)] = 1 - S_0 \left(\exp\left(\beta^{/} x\right) t \right) for t \geq 0 \tag{5.5}$$

The probability density function of the AFT model can be obtained by using:

$$f(t) = \psi(x) f_0[t\psi(x)] = e^{x^{*'}\beta^*} f_0 \left[te^{x^{*'}\beta^*} \right] for t \geq 0 \tag{5.6}$$

The hazard rate function of the AFT model is given by:

$$h(t) = \frac{f(t)}{S(t)} = \psi(x) \frac{f_0[t\psi(x)]}{S_0[t\psi(x)]} = \psi(x) h_0[t\psi(x)] = e^{x^{*'}\beta^*} h_0 \left[te^{\boldsymbol{x}'\boldsymbol{\beta}^*} \right], \tag{5.7}$$

The cumulative Hazard rate function of the AFT model is given by:

$$H(t) = -\log\left(S_0[t\psi(x)]\right) = H_0[t\psi(x)] = H_0\left[te^{x^{*'}\beta^*}\right], \tag{5.8}$$

where; $F_0(.), f_0(.), h(.)$ and $H_0(.)$ are the baseline cdf, pdf, hrf and cumulative hazard functions, respectively.

From the above equations 5.2 up to 5.8 we can see that the explanatory variables act multiplicatively on survival time so that there is to decelarate or accelerate the time to failure, hence the name of the model. For an AFT model we are interesting to measure the direct effect of the covariates on the survival time.

5.3 Generalized Log-logistic Accelerated Failure Time Model

The parameteric survival regression models can be expressed in a variety of ways. One such method involves the AFT model formulation.

Using $\psi(x) = \exp(-\beta' x)$ in Equation (5.2), the GLL regression model can be expressed in an AFT model. Let us $\boldsymbol{\beta}^* = \left(\beta_1^*, \beta_2^*, \cdots, \beta_p^*\right)'$ be the column vector of regression coefficients, and $\boldsymbol{x}^* = (x_1, x_2, \cdots, x_p)'$ be the corresponding vector of p covariates. Under the accelerated failure time assumption, the probability of an individual (with explanatory variables $\boldsymbol{x}^*$) surviving to time t is the same with the probability of a reference individual (i.e., $\boldsymbol{x}^* = 0$) surviving to time $te^{x^{x'}\beta^*}$ (Lawless, 2011). The failure rate function under the AFT assumption is given by:

$$h(t) = e^{x'\beta} h_0\left[te^{x'\beta}\right]$$

This is reference to the notion that explanatory variables work multiplicatively on time, causing the time to failure to accelerate (or decelerate). Starting with a gener-

alized loglogistic baseline survivor function. Suppose that T has a GLL distribution with parameters α, η, and k. The hazard function with covariate vector x is witten as follow:

$$\begin{aligned} h(t;x) &= h_0\left[te^{x'\beta}\right]e^{x^{x'\beta}} \\ &= \left(\frac{\alpha k\left(kte^{x'\beta}\right)^{\alpha-1}}{[1+(\eta te^{x'\beta})^{\alpha}]}\right)e^{x'\beta} \\ &= \alpha k^{\alpha}t^{\alpha-1}\left(\frac{\left(e^{x'\beta'}\right)^{\alpha-1}}{[1+(\eta te^{x'\beta})^{\alpha}]}\right)e^{x'\beta} \\ &= \alpha k^{\alpha}t^{\alpha-1}\left(\frac{\left(e^{x'\beta'}\right)^{\alpha}}{[1+(\eta te^{x'\beta})^{\alpha}]}\right). \\ &= \frac{\alpha t^{\alpha-1}\left(ke^{x'\beta}\right)^{\alpha}}{[1+(t\eta e^{x'\beta})^{\alpha}]} \\ &= \frac{\alpha t^{\alpha-1}\left(k^*\right)^{\alpha}}{[1+(t\eta^*)^{\alpha}]}, \end{aligned} \tag{5.9}$$

which is again the generalized log-logistic hazard function with $k^* = ke^{x'\beta}$ and $\eta^* = \eta e^{x'\beta}$.

Furthermore, using Equation (5.4), $S\left(t;\boldsymbol{x}^*\right) = S_0\left(te^{x'\beta'}\right)$, where; $S_0(.)$ is the baseline survivor function. We get

$$S_0\left(t;x\right) = \left[1+\left(\eta te^{x'\boldsymbol{\beta}^*}\right)^{\alpha}\right]^{-\frac{\kappa^{\alpha}}{\eta^{\alpha}}} = \left[1+(\eta^* t)^{\alpha}\right]^{-\frac{\kappa^{\alpha}}{\eta^{\alpha}}} \tag{5.10}$$

Which is again the generalized log-logistic survivor function with $\eta^* = \eta e^{r'\beta^*}$. This verifies that the generalized log-logistic is closed under the AFT family.

The other lifetime distribution functions of the generalized log-logistic AFT model are given by:

The cumulative distribution function for the GLL-AFT model is given by:

$$F\left(t;x\right) = 1 - S_0\left(te^{x'\boldsymbol{\beta}^*}\right) = 1 - \left[1+\left(\eta te^{x'\boldsymbol{\beta}}\right)^{x^x}\right]^{-\frac{\kappa^{\alpha}}{\eta^{\alpha}}} \tag{5.11}$$

The probability density function for the GLL-AFT model can be obtained by using:

$$f\left(t;x\right)=e^{x'\boldsymbol{\beta}^{*}}f_{0}\left[te^{x'\boldsymbol{\beta}^{*}}\right] \tag{5.12}$$

The cumulative hazard rate function for the GLL-AFT model is given by:

$$\begin{aligned} H\left(t;x\right) &= -\log\left(S_{0}\left[te^{x'\beta^{*}}\right]\right)=H_{0}\left[te^{x'\beta^{*}}\right] \\ &= \frac{k^{\alpha}}{\eta^{\alpha}}\log\left[1+\left(\eta te^{x'\beta^{*}}\right)\right]. \end{aligned} \tag{5.13}$$

Note that, using similar arguments, the Weibull, loglogistic, and log-normal distributions are all belong to the AFT family.

5.3.1 Sub-Models

The proposed model has two sub-models that are also closed under the AFT framework.

5.3.1.1 Sub-model I: $\eta = k$

When we put $\eta = k$ in Equations (5.9), then it shall be referring to the hazard rate of the loglogistic AFT model, the mathematical forms are expressed as:

$$h(t;x)=\frac{\alpha t^{\alpha-1}\left(ke^{x^{c'}\beta^{*}}\right)^{\alpha}}{\left[1+\left(t\eta e^{x^{x'}\beta^{*}}\right)^{\alpha}\right]}=\frac{\alpha t^{\alpha-1}\left(ke^{x^{x'}\boldsymbol{\beta}^{*}}\right)^{\alpha}}{\left[1+\left\{t\left(ke^{x^{x'}\boldsymbol{\beta}^{*}}\right)\right\}^{\alpha}\right]} \tag{5.14}$$

Which is again the log-logistic hazard function with $k^{*}=ke^{x'\beta^{*}}$.

5.3.1.2 Sub-model II: $\eta^{\alpha} \to 0$.

If we replace $\eta^{\alpha} \to 0$ in the Equation (5.9), then the derived hazard rate function will stand for Weibull AFT model, the mathematical forms are described as:

$$h(t;x) = \frac{\alpha t^{x^{\alpha-1}} \left(k e^{x^{x'}\beta^*}\right)^{\alpha}}{\left[1 + \left(t\eta e^{x^{t'}\beta^*}\right)^{\alpha}\right]} = \frac{\alpha t^{\alpha-1} \left(k e^{x'^{\beta^*}}\right)^{\alpha}}{[1 + (t e^{x'\beta})^{\alpha} \cdot 0]} = \alpha t^{\alpha-1} \left(k e^{x'\beta^*}\right)^{\alpha} \tag{5.15}$$

which is the Weibull hazard with $k^* = k e^{x^{x'}\beta'}$. This verifies that the generalized log-logistic and its sub-models are closed under the AFT family.

5.4 Inferential Procedures

In this section, we discuss the frequentist (via maximum likelihood estimation) and Bayesian Inference (applying independent gamma priors for the distributional parameters and normal prior for the regression coefficients) estimation procedures for the generalized log-logistic accelerated failure time model parameters.

5.4.1 MLE for Right-Censored Survival Data

In this section, we consider the problem of estimating the parameters of the proposed model for right-censored sample measurements. Suppose there be n individuals with survival times denoted by $T_1, T_2, \cdots, T_n$. Let that the data are subject to right censoring, we observe $t_i = \min(T_i, C_i)$, where $C_i > 0$ refers to a potential censoring time for individual i. Letting $\delta_i = \mathrm{I}(T_i, C_i)$ that equals 1 if $T_i \leq C_i$ and 0 otherwise. Let that a right-censored random sample consisting of data $(t_i, \delta_i, \boldsymbol{x}_i)$, $i = 1, 2, \cdots, n$, is available, where t_i is a censoring time or a lifetime according to whether $\delta_i = 0$ or 1 , respectively and $\boldsymbol{x_i} = x_1, x_2, \cdots, x_n$ is an $n \times 1$ column vector of external covariates for the i^{th} individual. We assume non-informative censoring, in which the survival time distribution offers no information about the censoring

time distirbution, and vice versa.

The assumption non-informative censoring is justifiable when censoring is independent (i.e., censoring is assumed random within any subgroup of interest) and/or random (i.e., hazard rates for censored and uncensored observations who remain in the risk set are equal); For more details about the non-informative censoring [1]. Under non-informative censoring, t_i, and δ_i are random variables with $P\left(t_i = c_i, \delta_i = 0\right) = P\left(T_i > C_i\right) = S\left(t_i\right)$ and $P\left(t_i, \delta_i = 1\right) = f\left(t_i\right)$. Then, the joint probability density function of t_i and δ_i is $[f(t)]^{\delta_i}[S(t)]^{1-\delta_i}$, which is the contribution of the i^{th} individual to the likelihood function. Thus, individual i contributes $f\left(t_i\right)$ to the likelihood function if an event occurs at time t_i, and contributes $S\left(t_i\right)$ if the individual is censored at t_i. Combining the information from the censored observations, we obtain the likelihood function.

$$L(\boldsymbol{\vartheta}) = \prod_{i=1}^{n}[f(t)]^{\delta_i}[S(t)]^{1-\delta_i} \tag{5.16}$$

where; ϑ is a vector of parameters characterizing the distribution of T_i. Using the definition of the probability density function in terms of hazard rate and survival functions as below:

$$f(t) = h(t)S(t) \tag{5.17}$$

The likelihood function can also be defined as:

$$L(\boldsymbol{\vartheta}) = \prod_{i=1}^{n}[h(t)S(t)]^{\delta_i}[S(t)]^{1-\delta_i}, \tag{5.18}$$

By simplifying, we get

$$L(\boldsymbol{\vartheta}) = \prod_{i=1}^{n}[h(t)]^{\delta_i}[S(t)] \tag{5.19}$$

If a parametric AFT model is considered, the log-likelihood function can be written

using (5.19) as follows:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \delta_i \log\left[h\left(t; \boldsymbol{x}_i\right)\right] + \sum_{i=1}^{n} \log\left[S\left(t; \boldsymbol{x}_i\right)\right] \tag{5.20}$$

where; $\boldsymbol{\theta} = (\boldsymbol{\vartheta}, \boldsymbol{\beta}), \boldsymbol{\vartheta}$ is a vector of the distributional parameters, and $\boldsymbol{\beta}$ is a vector of the regression coefficients. This can be maximized directly using the Newton-Raphson optimization algorithm and the hypothesis testing and interval estimates of the model parameters can proceed under the approximate normality of the MLE estimators (Lawless, 2011).

In GLL AFT model, the likelihood function is given by

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \left[\frac{\alpha t^{\alpha-1}\left(ke^{x^{x'}\beta^*}\right)^{\alpha}}{\left[1+\left(t\eta e^{x^{x'}\beta^*}\right)^{\alpha}\right]}\right]^{\delta_i} \prod_{i=1}^{n} \left[\left[1+\left(t\eta e^{x^{x'}\beta^x}\right)^{\alpha}\right]^{-\frac{k^x}{\eta^x}}\right] \tag{5.21}$$

In our case; if we assume that $d = \sum_{i=1}^{n} \delta_i$. Applying the likelihood for the generalized log-logistic AFT model, the full log-likelihood function can be expressed as:

$$\begin{aligned}\ell(\boldsymbol{\theta}) &= d\ln\alpha + (\alpha-1)d\ln t_i + \alpha d\ln\left(ke^{x^{x'}\beta^*}\right) \\ &\quad - d\ln\left(1+\left(t\eta e^{x^{x'}\beta^*}\right)^{\alpha}\right) - \left(\frac{k}{\eta}\right)^{\alpha}\sum_{i=1}^{n}\ln\left(1+\left(t\eta e^{x^{x'}\beta^*}\right)^{\alpha}\right)\end{aligned} \tag{5.22}$$

where; $\boldsymbol{\theta}$ denotes all the model parameters collectively. For our model, includes the distributional parameters $\vartheta' = (\alpha, \eta, k)$ that characterize the baseline hazard and survival functions $h_0(t)$ and $S_0(t)$ and the regression coefficients $\boldsymbol{\beta}$, leading to $\boldsymbol{\theta} = (\boldsymbol{\vartheta}', \boldsymbol{\beta}')$

To obtain the MLE's of $\boldsymbol{\theta}' = (\alpha, \eta, k)$ and $\boldsymbol{\beta}'$, we can maximize (29) directly with respect to (α, η, k) and $\boldsymbol{\beta}'$. or we can solve the first derivative of the log-likelihood

function (non-linear equations below):

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha} &= \frac{d}{\alpha} + d\ln t_i + d\ln\left(ke^{x^{x'}\boldsymbol{\beta}^*}\right) \\ &\quad - \frac{d\left(t_i\eta e^{x^{x'}\beta^*}\right)^{\alpha}\ln\left[\left(t_i\eta e^{x^{x'}\beta^*}\right)\right]}{1+\left(t_i\eta e^{x^{x'}\beta^*}\right)^{\alpha}} - \left(\frac{k}{\eta}\right)^{\alpha} \\ &\quad \ln\left(\frac{k}{\eta}\right)\sum_{i=1}^{n}\ln\left(1+\left(t_i\eta e^{\alpha^{x'}\beta^*}\right)^{\alpha}\right) \\ &\quad - \frac{\left(\frac{k}{\eta}\right)^{\alpha}\sum_{i=1}^{n}\left(t_i\eta e^{x'\beta'}\right)^{\alpha}\ln\left[\left(t_i\eta e^{x^{x'}\beta^*}\right)\right]}{1+\left(t_i\eta e^{x'\beta^*}\right)^{\alpha}}\end{aligned} \tag{5.23}$$

$$\frac{\partial \ell(\theta)}{\partial k} = \frac{\alpha d}{k} - \frac{\left(\frac{k}{\eta}\right)^{\alpha}\alpha\sum_{i=1}^{n}\ln\left(1+\left(t_i\eta e^{x^{x'}\beta^*}\right)^{\alpha}\right)}{k}, \tag{5.24}$$

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\theta})}{\partial \eta} &= -\left(\frac{\alpha}{\eta}\right)d\left(\frac{\left(t_i\eta e^{\boldsymbol{x}^{x'}\beta^*}\right)^{\alpha}}{1+\left(t_i\eta e^{x^{x'}\beta^*}\right)^{\alpha}}\right) \\ &\quad + \left(\frac{\alpha}{\eta}\right)\left(\frac{k}{\eta}\right)^{\alpha}\sum_{i=1}^{n}\ln\left(1+\left(t_i\eta e^{x^{x'}\beta^*}\right)^{\alpha}\right) \\ &\quad - \left(\frac{\alpha}{\eta}\right)\left(\frac{k}{\eta}\right)^{\alpha}\sum_{i=1}^{n}\left(\frac{\left(t_i\eta e^{x'\beta^*}\right)^{\alpha}}{1+\left(t_i\eta e^{x'\beta^*}\right)^{\alpha}}\right),\end{aligned} \tag{5.25}$$

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_m} &= \alpha d x_{im} - d\alpha x_{im}\left(\frac{\left(t_i\eta e^{x^{x'}\beta^*}\right)^{\alpha}}{1+\left(t_i\eta e^{x^{x'}\beta^*}\right)^{\alpha}}\right) - \left(\frac{k}{\eta}\right)^{\alpha}\alpha x_{im} \\ &\quad \times\sum_{i=1}^{n}\frac{\left(t_i\eta e^{x^{x'}\beta'}\right)^{\alpha}}{1+\left(t_i\eta e^{x^{x'}\beta'}\right)^{\alpha}}, m \\ &\quad = 1, 2, \cdots, r\end{aligned} \tag{5.26}$$

It is important to know that the $\theta' = (\alpha, \eta, \kappa, \beta')$ cannot be solved analytically. In order, to solve these equations, numerical integration methods, such us the Newton-Raphson algorithm are used.

5.4.2 Bayesian Approach

In this sub-section, we assume non-informative priors in setting up the Bayesian framework for estimating the parameters. For each regression coefficients $\beta_k, \mathrm{k} = 1, 2, \ldots, \mathrm{k}$ the marginal prior distribution is assumed as a normal distribution with mean zero and a minimal precision, $N(0, 0.001)$. Gamma priors are chosen as the marginal prior distributions for the generalized log-logistic parameters $\alpha : G(a_1, b_1)$, $\kappa : G(a_2, b_2)$, and $\eta : G(a_3, b_3)$ because of the flexibility of the gamma priors and its popularity under a non-informative prior framework.

The Bayesian model is implemented by specifying a prior distribution with their corresponding parameters and regression coefficients of the model and then multiplying with the likelihood function.

5.4.2.1 The likelihood function

The likelihood function is an important part of the inferential process. The construction of survival data requires special attention in the area of survival data analysis because it is dependent on the type of censoring observations. Assuming that lifetimes and censoring are independent, the likelihood of the model's parameters can be written by incorporating elements such as: (i) When the exact lifetime is unknown, the pdf of the survival time at the observed time t, f(t); (ii) In the case of right-censored observation; the sf at the censoring time, S(t); (iii) In the case of a left-censored observation, the cdf at the interval-censored observation, the difference between the ff at times left-censoring and right-censoring. The likelihood function

for our proposed model is given us:

$$
\begin{aligned}
L(\boldsymbol{\vartheta} \mid \boldsymbol{D}) &= \prod_{i=1}^{n} \left[f\left(t_i \mid \alpha, \eta, k, \beta, x\right)\right]^{\delta_i} \left[S\left(t_i \mid \alpha, \eta, k, \beta, x\right)\right]^{1-\delta_i} \\
&= \prod_{i=1}^{n} \left[\frac{f\left(t_i \mid \alpha, \eta, k, \beta, x\right)}{S\left(t_i \mid \alpha, \eta, k, \beta, x\right)}\right]^{\delta_i} \left[S\left(t_i \mid \alpha, \eta, k, \beta, x\right)\right] \\
&= \prod_{i=1}^{n} \left[h\left(t_i \mid \alpha, \eta, k, \beta, x\right)\right]^{\delta_i} \left[S\left(t_i \mid \alpha, \eta, k, \beta, x\right)\right] \\
&= \prod_{i=1}^{n} \left[f\left(t_i \mid \boldsymbol{\vartheta}, x\right)\right]^{\delta_i} \left[S\left(t_i \mid \boldsymbol{\vartheta}, x\right)\right]^{1-\delta_i}
\end{aligned}
\tag{5.27}
$$

where; $f\left(t_i \mid \alpha, \eta, k, \beta, x\right), h\left(t_i \mid \alpha, \eta, k, \beta, x\right),)$, and $S\left(t_i \mid \alpha, \eta, k, \beta, x\right)$ of GLL AFT model are given in Eqs. (15), (16), (18), $\boldsymbol{\vartheta} = (\alpha, \eta, k, \beta)$, and $\boldsymbol{D} = (t, \delta, \boldsymbol{X})$, where δ is the censoring indicator ($\delta = 1$, iftheobservationisfailedand $\delta = 0$ iftheobservationiscensored $), t$ is the survivor time, and $\boldsymbol{X}$ is the matrix of covariates which is formulated as desing or model matrix.

The likelihood function of our proposed model is not implemented in JAGS, so the "zeros trick" method of specifying it indirectly using a Poisson distribution has been utilized (Christensen et al., 2010; Alvares et al., 2021; Alvares and Rubio, 2021; Lázaro Hervás, 2018). The idea behind this method is that the contribution of a Poisson (φ) observation of zero is $\exp(-\varphi)$; if we set $\varphi_i = -\log f\left(t_i \mid \alpha, \eta, k, \beta, x\right), i = 1, 2, \cdots, n$, with observed data as vector of 0 's, than we get the correct contributions.

5.4.2.2 Prior distribution

In this sub-section, we have assumed normal priors for the covariates and independent gamma priors for the distributional parameters. because of the flexibility of gamma distribution. These priors are considered in many research papers in the literature such as (Alvares et al., 2021; Lázaro Hervás, 2018). Let us propose the

independent gamma priors for the parameters

$$p(\alpha)G\left(a_1,b_1\right)=\frac{b_1^{a_1}}{\Gamma\left(a_1\right)}\alpha^{a_1-1}e^{-b_1\alpha};a_1,b_1,\alpha>0 \tag{5.28}$$

$$p(k)G\left(a_2,b_2\right)=\frac{b_2^{a_2}}{\Gamma\left(a_2\right)}k^{a_2-1}e^{-b_2 i};a_2,b_2,k>0 \tag{5.29}$$

$$p(\eta)G\left(a_3,b_3\right)=\frac{b_3^{a_3}}{\Gamma\left(a_2\right)}\eta^{a_3-1}e^{-b_3\eta};a_3,b_3,\eta>0 \tag{5.30}$$

The regression coefficients (taking as a normal distribution) we have

$$p\left(\boldsymbol{\beta}_m\right)N\left(\mu,\sigma^2\right) \tag{5.31}$$

The joint prior distribution of the distributioanl parameters and regression coefficients is given by

$$p\left(\alpha,k,\eta,\boldsymbol{\beta}_m\right)=p(\alpha)p(\eta)p(k)p\left(\boldsymbol{\beta}_m\right) \tag{5.32}$$

5.4.2.3 Posterior distribution

The joint posterior density function of α,k,η, and β ' given data is given by

$$\begin{aligned}
&p\left(\alpha,k,\eta,\boldsymbol{\beta}_m\mid x\right)\propto p\left(\alpha,k,\eta,\boldsymbol{\beta}'\right)L\left(\alpha,k,\eta,\boldsymbol{\beta}_m\right)\\
&p\left(\alpha,k,\eta,\boldsymbol{\beta}_m\mid x\right)\propto p(\alpha)p(k)p(\eta)\left\{\prod_{i=0}^{m}p\left(\boldsymbol{\beta}_m\right)\right\}L\left(\alpha,k,\eta,\boldsymbol{\beta}_m\right)
\end{aligned} \tag{5.33}$$

where; the first four terms on the right hand side of Equation (5.33) corresponds to the prior specifications for α,k,η, and $\boldsymbol{\beta}_m$ and $L\left(\alpha,k,\eta,\boldsymbol{\beta}_m\right)$ is the likelihood function for the generalized log-logistic AFT model.

The posterior density function is complicated and no close form inferences are available. Therefore, the Markov chain Monte Carlo (McMC) techniques are used to do Bayesian approach, in which we sample from the posterior distribution using the Metropolis within Gibbs algorithm [53]. In our implementation each component

of the distributional parameters and regression coefficients is updated applying the independence sampler (Smith and Roberts, 1993).

5.5 Simulation Study

We conduct a thorough simulation study in this section to demonstrate the proposed model's inferential properties. In particular, we demonstrate parameter estimation, the tendency to recover baseline hazard shapes using the Akaike information criterion (AIC) to choose models that accuratley reflect the underlying hazard rate shape, and the effect of censoring proportions on inferencial properties of the model.

5.5.1 Data Generation and Simulation Designs

Breifly, in the design of the simulation study, we simulated N = 1000 data sets assuming the AFT regression framework given in equation (5.8).

In the case of covariates, four covariates are considered in the simulation study. Two continuous covariates (x_1 and x_2) were generated using the standard normal distribution, while two binary covariates (x_3, and x_4) were generated using the Bernoulli (0.5) distribution. The values for the AFT regression coefficients are chosen to be $(-2, 0.5, -0.5, 0.75, -0.75)$ corresponding to the covariate vector $x = (1, x_1, x_2, x_3, x_4)^T$.

In the lifetime and censored data generation, we used the exponentiated Weibull (EW) distribution to simulate lifetime data from the AFT framework using the inverse transform method (Leemis et al., 1990), and assuming effects of the covariates and the intercept in the regression equation. Additionally, for censored data we assume an adminstrative censoring.

The probability density function (pdf) of the exponentiated Weibull distribution is

expressed as follows:

$$f(t;\alpha,k,\eta)=\alpha k\eta(kt)^{\alpha-1}\left(1-\exp\left\{-(kt)^{\alpha}\right\}\right)^{\eta-1}\exp\left\{-(kt)^{\alpha}\right\}, \tag{5.34}$$

The exponentiated Weibull (EW) distribution can incorporate all five basic shapes of the hazard rate function including; constant, decreasing, increasing, unimodal and bathtub shapes. The EW distribution is also closed under the AFT regression framework in equation (5.4).

The chf for the exponentiated Weibull model are given as:

$$H_0(t;\alpha,k,\eta)=-\log\left\{1-\left(1-e^{-(kt)^{\alpha}}\right)^{\eta}\right\}, \tag{5.35}$$

The reversed chf is expressed as follows:

$$H_0{}^{-1}(u;\alpha,k,\eta)=-\frac{\log\left[\left(e^{-u}-1\right)^{1/\eta}-1\right]^{\frac{1}{\alpha}}}{k}, \tag{5.36}$$

where, $t>0$, and $\alpha,k,\eta>0$ are the parameters.

5.5.2 Simulation Algorithm

1. First we specify the initial values for the model parameters
2. Generate the lifetimes data using the inverse transform method by inverting the cumulative hazard rate function for the proposed model.
3. Evaluate the values of the estimates using the different estimates
4. Evaluate the inferential properties of the estimates including Bias and MSEs.
5. Selecting the best superior model using AIC criterion

5.5.3 Simulation Scenarios

To assess the performance of generalized log-logistic AFT model in comparison with its sub-models log-logistic and Weibull AFT models, as well as the exponentiated Weibull distribution, we conducted three simulation scenarios based on the shape of the failure rate function (non-monotone (unimodal), monotone (increasing) and decreasing) to explore the affect of the baseline hazard rate shape specification in the infrential properties of the AFT model.

5.5.3.1 Scenario 1: non-monotone (or unimodal) hazard rate function

In the Scenario 1, the lifetimes data were generated from the exponentiated Weibull model using the parameter values for ($\alpha = 1.75, k = 0.70$, and $\eta = 2.5$) and the censored data were generated from assuming adminstrative censoring (a) $Tc = 5$, which induced approximaley 20% censoring, (b) $Tc = 8$, which induced approximately 30% senting from light to heavy censoirng cases.

5.5.3.2 Scenario 2: monotone increasing hazard rate function

In the Scenario 2, the lifetimes data were generated from the exponentiated Weibull model using the parameter values for ($\alpha = 1.0, k = 2.50$, and $\eta = 1.0$) and the censored data were generated from assuming adminstrative censoirng (a) $Tc = 14$, which induced approximaley 20% censoring, (b) $Tc = 10$, which induced approximately 30% censoirng, representing from light to heavy censoirng cases.

5.5.3.3 Scenario 3: monotone decreasing hazard rate function

In the Scenario 3, the lifetimes data were generated from the exponentiated Weibull model using the parameter values for ($\alpha = 0.50, k = 0.50$, and $\eta = 0.85$) and the censored data were generated from assuming adminstrative censoring (a) $Tc = 8.0$,

which induced approximaley 20% censoring, (b) $Tc = 5$, which induced approximately 30% censoirng, representing from light to heavy censoirng cases.

5.5.4 Simulated Data Analysis

To evaluate the inferential properties of the proposed models in all simulation scenarios, we fitted the AFT model with GLL baseline hazard (Generalized log-logistsic AFT), as compared to the corresponding true generating model from the exponentiated Weibull AFT model. Furthermore, the stability of the estimators for the regression coefficients of each model were evaluated by using the average bias (AB), the mean square error (MSE) and the relative bias (RB). We also fitted the submodels in each scenario.

$$AB = \frac{1}{N}\sum_{i=1}^{N}(\widehat{\theta_i} - \theta) \tag{5.37}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\widehat{\theta_i} - \theta)^2 \tag{5.38}$$

Relative Bias:

$$RB = \frac{\frac{1}{N}\sum_{i=1}^{N}(\widehat{\theta_i} - \theta),}{\theta} \tag{5.39}$$

where $\theta = \alpha, \eta, \kappa, \text{and} \beta^T$ Since our goal was to analyze the estimator's qualities rather than the characteristics of the optimization process, we applied the values of the parameters of the generating model as starting points in the optimization step in all situations.The analysis was carried out using the R programming language.The optimization stage was carried out with the help of the R program "nlminb()."

5.5.5 Measures of Performance

In this study, the flexibility of the models for the covariates were assessed by using the measures including; the mean, the bias, the mean (estimated) standard error (MSE), and the letative bias. In addition, for model comparion, we used the Akaike

Information Criterion (AIC). The AIC is expresesd as:

$$AIC = 2(k + p) - 2l, \tag{5.40}$$

where; k is the number of parameters for the model, p is the number of covariates, and l represents the log-likelihood function computes as the MLEs.
In general, the best model has the lowest AIC values. A good rule of thumb is that if $\Delta_{AIC} = AIC_M - AIC_{\min} > 2$, then Model M has significantly less support than the model with the lowest AIC [56-58].

5.5.6 Simulation Results

The simulation results are presented in Tables 5.1 to 5.6.

5.5.6.1 Scenario 1 Results

For Scenario 1, the adequacy of a model's fit appears to be dependent on the amount of censoring. Overall, the GLL model outperforms the LL, and the Weibull model, while it is closer to the exponentiated Weibull models in terms of AIC criterion. As shown in Table 5.1, the three main competing models are the GLL, LL and the exponentiated weibull model, since the Weibull doesnot incorportate unimodal failure rate functinons theoretically and it has the highest AIC value, as expected.. The same happens in Table 5.2 where the censoring proportions increased to 30% censoring case. In terms of MSEs, for continous covariates (x_1 and x_2), the GLL model had the smallest MSE, whereas for binary covariates (x_3 and x_4), GLL, LL and exponentiated Weibull were roughly equivalent, and all are superior to Weibull. Furthermore, in terms of the intercept, LL and Weibull models produce estimates with lower bias and MSE values than the GLL and the true generating (exponentiated Weibull) models, and both are superior to GLL and exponentiated Weibull models. Furthermore, the results were roughly equivalent to the MSE results dis-

Table 5.1: Simulation results for Scenario 1 (n = 1000) with about 20% censoring for model comparison and performance.

Models	Parameters	(True value)	Mean	AB	MSE	RB
			30% Censoring			
GLL-AFT	β_0	-2.0	-4.122	1.108	-3.204	-0.554
(**AIC = 1842.540**)	β_1	0.50	0.611	0.111	0.123	0.222
	β_2	-0.50	-0.436	0.064	-0.060	-0.128
	β_3	0.75	0.763	0.013	0.020	0.017
	β_4	-0.75	-0.765	-0.015	0.185	0.153
	α	1.75	1.428			
	κ	0.70	0.478			
	η	2.50	2.223			
EW-AFT	β_0	-2.0	2.507	4.508	2.289	-2.254
(**AIC = 1841.818**)	β_1	0.50	0.663	0.163	0.189	0.326
	β_2	-0.50	-0.505	-0.005	0.005	-0.010
	β_3	0.75	0.863	0.113	0.182	0.151
	β_4	-0.75	-0.960	-0.210	0.359	0.280
	α	1.75	0.181			
	κ	0.70	0.187			
	η	2.50	4.193			
W-AFT	β_0	-2.0	-2.230	-0.230	0.975	0.116
(AIC = 1863.423)	β_1	0.50	0.723	0.223	0.273	0.446
	β_2	-0.50	-0.613	-0.113	0.125	0.226
	β_3	0.75	0.881	0.131	0.214	0.175
	β_4	- 0.75	-1.067	-0.317	0.576	0.423
	α	1.75	2.401			
	κ	0.70	0.554			
LL-AFT	β_0	-2.0	-2.200	0.200	-0.760	-0.100
(AIC = 1854.482)	β_1	0.50	534	0.035	0.036	0.070
	β_2	-0.50	-0.498	0.002	-0.002	-0.004
	β_3	0.75	0.853	0.103	166	0.137
	β_4	-0.75	-0.659	0.091	-0.128	-0.121
	α	1.75	2.294			
	κ	0.70	1.192			

cussed above in terms of bias and relative bias. Finally, when it comes to censoring, the GLL model outperforms the other competing models except the true generating model.

5.5.6.2 Scenario 2 Results

In the AIC values for Scenario 2, the Weibull, exponentiated Weibull, and GLL models were roughly close to each other and all were superior to the Log-logistic model as shown in Tables 5.3 to 5.4. Because LL cannot theoretically accommodate increasing failure rates, it has the highest AIC value, as expected. Except for the intercept variable, where the Weibull and exponentiated Weibull perform better than the other competing models in terms of bias, MSE, and relative bias, all of the models produce small values and are roughly close to each other. Overall, the GLL

Table 5.2: Simulation results for Scenario 1 ($n = 1000$) with about 30% censoring for model comparison and performance.

Models	Parameters	(True value)	Mean 30% Censoring	AB	MSE	RB
GLL-AFT	β_0	-2.0	-0.487	1.513	3.763	-0.756
(AIC = 2671.365)	β_1	0.50	0.616	0.116	0.129	0.232
	β_2	-0.50	-0.437	0.063	0.059	-0.126
	β_3	0.75	0.800	0.050	0.078	0.067
	β_4	-0.75	-0.912	-0.162	0.270	0.216
	α	1.75	1.428			
	κ	0.70	0.478			
	η	2.50	2.223			
EW-AFT	β_0	-2.0	8.4587	10.458	67.573	-5.229
(AIC = 2670.107)	β_1	0.50	0.713	0.214	0.259	0.428
	β_2	-0.50	-0.543	-0.043	0.045	0.086
	β_3	0.75	0.990	0.240	0.418	0.320
	β_4	-0.75	-1.041	-0.291	0.521	0.388
	α	1.75	0.113			
	κ	0.70	0.121			
	η	2.50	4.493			
W-AFT	β_0	-2.0	-2.317	-0.317	1.369	0.158
(AIC = 2705.119)	β_1	0.50	0.826	0.326	0.432	0.652
	β_2	-0.50	-0.685	-0.185	0.219	0.370
	β_3	0.75	1.082	0.332	0.608	0.443
	β_4	- 0.75	-1.202	-0.452	0.881	0.603
	α	1.75	0.773			
	κ	0.70	0.543			
LL-AFT	β_0	-2.0	-1.800	0.200	0.760	-0.100
(AIC = 2686.816)	β_1	0.50	557	0.057	0.060	0.114
	β_2	-0.50	-0.511	0.011	0.011	-0.022
	β_3	0.75	0.861	0.111	0.179	0.148
	β_4	-0.75	-0.683	0.082	0.116	-0.109
	α	1.75	0.000			
	κ	0.70	0.953			

model outperforms the LL, Weibull, and exponentiated Weibull models in terms of AIC criterion for both cases of censoring.

5.5.6.3 Scenario 3 Results

In case of Scenario 3, In theory, all of the fitted models can account for decreasing failure rate shapes. In terms of regression coefficients, except for the intercept, where the GLL model produced the highest values, all four models produced estimates with very close bias and MSE values for both continuous and binary covariates as shown in Tables 5.5 to 5.6. In terms of AIC values, the GLL model outperforms the others, including the true model (exponentiated Weibull).

Finally, in all scenarios involving heavy censoring, the LL model outperforms the other competing models.As expected, LL model has a cdf with a tractable closed-form expression which makes it to be useful for censored survival data.

Table 5.3: Simulation results for Scenario 2 (n = 1000) with about 20% censoring for model comparison and performance.

Models	Parameters	(True value)	Mean	AB	MSE	RB
			30% Censoring			
GLL-AFT	β_0	-2.0	-2.905	-0.905	4.440	-0.452
(**AIC = 3955.988**)	β_1	0.50	0.584	0.084	0.091	0.168
	β_2	-0.50	-0.553	0.053	0.055	-0.106
	β_3	0.75	0.847	0.097	0.155	0.129
	β_4	-0.75	-0.881	-0.131	0.213	0.175
	α	1.00	2.337			
	κ	2.50	2.262			
	η	1.00	0.708			
EW-AFT	β_0	-2.0	-1.994	0.006	0.025	-0.003
(**AIC = 3955.517**)	β_1	0.50	0.583	0.083	0.090	0.166
	β_2	-0.50	-0.553	-0.053	0.055	0.106
	β_3	0.75	0.846	0.096	0.153	0.128
	β_4	-0.75	-0.881	-0.131	0.214	0.175
	α	1.00	1.029			
	κ	2.50	1.997			
	η	1.00	1.235			
W-AFT	β_0	-2.0	-2.055	-0.055	0.224	0.028
(AIC = 3954.917)	β_1	0.50	0.581	0.081	0.087	0.162
	β_2	-0.50	-0.554	-0.054	0.056	0.108
	β_3	0.75	843	0.093	0.149	0.124
	β_4	-0.75	-0.881	-0.131	0.213	0.175
	α	1.00	1.053			
	κ	2.50	2.264			
LL-AFT	β_0	-2.0	-1.452	0.548	1.891	-0.274
(AIC = 4013.810)	β_1	0.50	581	0.081	0.088	0.162
	β_2	-0.50	-0.534	-0.034	0.035	-0.068
	β_3	0.75	0.832	0.082	0.129	0.109
	β_4	-0.75	-0.863	-0.113	0.183	0.151
	α	1.00	0.452			
	κ	2.50	0.322			

Table 5.4: Simulation results for Scenario 2 (n = 1000) with about 30% censoring for model comparison and performance.

Models	Parameters	(True value)	Mean	AB	MSE	RB
			30% Censoring			
GLL-AFT	β_0	-2.0	-2.965	-0.965	4.792	-0.482
(**AIC = 3471.332**)	β_1	0.50	0.623	0.123	0.139	0.246
	β_2	-0.50	-0.589	-0.089	0.097	0.178
	β_3	0.75	0.908	0.158	0.263	0.211
	β_4	-0.75	-0.965	-0.215	0.369	0.287
	α	1.00	1.104			
	κ	2.50	1.085			
	η	1.00	2.605			
EW-AFT	β_0	-2.0	-2.071	-0.071	0.287	-0.035
(**AIC = 3471.261**)	β_1	0.50	0.622	0.122	0.137	0.244
	β_2	-0.50	-0.553	-0.053	0.055	0.178
	β_3	0.75	0.907	0.157	0.260	0.209
	β_4	-0.75	-0.965	-0.215	0.368	0.287
	α	1.00	1.055			
	κ	2.50	1.936			
	η	1.00	1.160			
W-AFT	β_0	-2.0	-2.077	-0.077	0.313	0.038
(AIC = 3469.869)	β_1	0.50	0.619	0.119	0.133	0.238
	β_2	-0.50	-0.589	-0.089	0.097	0.178
	β_3	0.75	0.904	0.154	0.255	0.205
	β_4	-0.75	-0.963	-0.213	0.365	0.284
	α	1.00	1.114			
	κ	2.50	2.123			
LL-AFT	β_0	-1.488	-2.077	0.512	1.786	-0.256
(AIC = 3524.605)	β_1	0.50	0.625	0.125	0.141	0.250
	β_2	-0.50	-0.575	-0.075	0.081	-0.150
	β_3	0.75	0.897	0.147	0.243	0.196
	β_4	-0.75	-0.949	-0.199	0.338	0.265
	α	1.00	0.488			
	κ	2.50	0.354			

Table 5.5: Simulation results for Scenario 3 (n = 1000) with about 20% censoring for model comparison and performance.

Models	Parameters	(True value)	Mean	AB	MSE	RB
			30% Censoring			
GLL-AFT	β_0	-2.0	-0.413	1.587	-3.829	-0.793
(**AIC = 2869.376**)	β_1	0.50	0.678	0.178	0.210	0.356
	β_2	-0.50	-0.393	-0.106	-0.095	0.212
	β_3	0.75	0.778	0.028	0.043	0.037
	β_4	-0.75	-0.912	-0.162	0.269	0.216
	α	0.50	0.532			
	κ	0.50	1.167			
	η	0.85	1.291			
EW-AFT	β_0	-2.0	2.820	4.820	3.950	-2.410
(**AIC = 2886.507**)	β_1	0.50	0.745	0.245	0.306	0.490
	β_2	-0.50	-0.486	0.014	-0.014	-0.028
	β_3	0.75	0.896	0.146	0.240	0.195
	β_4	-0.75	-1.032	-0.282	0.503	0.376
	α	0.50	0.774			
	κ	0.50	0.178			
	η	0.85	4.059			
W-AFT	β_0	-2.0	-2.123	-0.123	0.507	0.062
(AIC = 2923.193)	β_1	0.50	0.789	0.289	0.372	0.578
	β_2	-0.50	-0.581	-0.081	0.088	0.162
	β_3	0.75	0.921	0.171	0.286	0.228
	β_4	-0.75	-1.094	-0.345	0.636	0.460
	α	0.50	0.573			
	κ	0.50	0.370			
LL-AFT	β_0	-1.488	-1.500	0.500	-1.750	-0.250
(AIC = 2958.007)	β_1	0.50	0.582	0.082	0.089	0.164
	β_2	-0.50	-0.467	0.033	-0.032	-0.066
	β_3	0.75	0.992	0.242	0.422	0.323
	β_4	-0.75	-0.535	-0.215	-0.277	-0.287
	α	0.50	0.000			
	κ	0.50	1.820			

In conclusion, the simulation study demonstrated that the GLL AFT model has the potential to be a very useful parametric AFT hazard-based regression model for adequately describing different types of survival data with different hazard rate shapes and censoring proportions.

5.6 Applications to Larynx Cancer Patients Data

In this section, we considered a real-life right-censored data for larynx cancer patients to illustrate the flexibility and usefulness of the generalized log-logistic AFT model.

Table 5.6: Simulation results for Scenario 3 (n = 1000) with about 30% censoring for model comparison and performance.

Models	Parameters	(True value)	Mean	AB	MSE	RB
			30% Censoring			
GLL-AFT	β_0	-2.0	0.367	2.367	-3.865	-1.184
(**AIC = 2391.542**)	β_1	0.50	0.692	0.192	0.229	0.384
	β_2	-0.50	-0.369	-0.131	-0.114	0.262
	β_3	0.75	0.747	-0.003	-0.005	0.004
	β_4	-0.75	-0.995	-0.245	0.427	0.327
	α	0.50	0.558			
	κ	0.50	0.772			
	η	0.85	1.996			
EW-AFT	β_0	-2.0	5.188	7.188	22.902	-3.594
(**AIC = 2396.594**)	β_1	0.50	0.798	0.298	0.387	0.596
	β_2	-0.50	-0.490	0.009	-0.009	-0.018
	β_3	0.75	0.918	0.168	0.280	0.224
	β_4	-0.75	-1.155	-0.405	0.771	0.540
	α	0.50	0.113			
	κ	0.50	0.121			
	η	0.85	43.493			
W-AFT	β_0	-2.0	-2.179	-0.179	0.748	0.090
(AIC = 2471.964)	β_1	0.50	0.873	0.373	0.511	0.746
	β_2	-0.50	-0.616	-0.116	0.130	0.232
	β_3	0.75	0.954	0.204	0.347	0.272
	β_4	-0.75	-1.254	-0.504	1.010	0.672
	α	0.50	0.700			
	κ	0.50	0.343			
LL-AFT	β_0	-1.488	-1.500	0.500	-1.750	-0.250
(AIC = 2472.403)	β_1	0.50	0.618	0.118	0.132	0.236
	β_2	-0.50	-0.470	0.030	-0.029	-0.060
	β_3	0.75	0.990	0.240	0.417	0.320
	β_4	-0.75	-0.558	-0.192	-0.251	-0.256
	α	0.50	0.000			
	κ	0.50	1.890			

5.6.1 Data Set: Larynx Data

5.6.1.1 Description:

The data set contains an observation of 90 male larynx-cancer patients, who were treated and diagnosed between 1970 and 1978 and was first collected by (Kardaun, 1983), and analyzed by (Klein et al., 2014) and applied to a Bayesian AFT model by (Christensen et al., 2010). The data consist of survival times in months from diagnosis to death or censoring and covariates including; (1) The stage of the cancer $(1-4)$ the stage variable is recorded into indicators S_i for stages $1, 2, 3, 4$; (2) time to death or on-study time (in months) from the first treatment until death, or the end of study; (3) The age at diagnosis (in years) of larynx-cancer data; (4) The year of diagnosis (diagyr) of larynx cancer; and (5) the death indicator ('delta'; 0 = alive, 1 = dead). In the study, we also standardize the age and year before using them.

5.6.2 Data Analysis

In this study, we assume the independent gamma priors for the distributional parameters: $\alpha : G(a_1, b_1)$, $\kappa : G(a_2, b_2)$, and $\eta : G(a_3, b_3)$ with hyper-parameter values $(a_1 = b_1 = a_2 = b_2 = a_3 = b_3 = 10.0)$ and the marginal prior distribution for each of the regression coefficient $\beta_k, k = 1, 2, \cdots k.$ is assumed as a normal distribution $N(0, 0.001)$ and our link function is:

$$\begin{aligned}\psi(\boldsymbol{x}'\boldsymbol{\beta}) &= \exp(\boldsymbol{x}'\boldsymbol{\beta}) \\ &= \exp(\beta_1(\text{ intercept }) + \beta_2 I(\text{ stage } = 2) + \beta_3 I(\text{ Stage } = 3) \\ &+\beta_4 I(\text{ stage } = 4) + \beta_5 \text{ age } + \beta_6 \text{ year })\end{aligned}$$

where; β_1 is the intercept; (I(stage =.) is an indicator variable for stages $= 2, 3$, and 4 with regression coefficients β_2, β_3, and β_4, respectively; β_5, and β_6 are regression coefficients for the age of patient and diagnosis year. In addition, Stage = 1 is used as a reference category.

In this study, we started three parallel chains for a sufficiently large number of iterations until convergence was achieved. The convergence was achieved at 50,000 with a burn-in of 1000 . Finally, a posterior sample of size 5000 is used with a thinning interval of ten, i.e., every hundredth outcome is sorted. As a result, we have the posterior sample $(\alpha_i, k_i, \eta_i, \beta_{1i}, \beta_2, \beta_{3i}, \beta_{4i}, \beta_{5i}, \beta_{6i}), i = 1 \cdots 5000$ drawn from chain 1, chain 2, and chain 3. Since simulation-based Bayesian rize posterior distributions or compute any relevant statistical rize posterior distributions or compute any relevant statistical three Markov chains are combined using the function (as. mcmc), and the posterior samples of every parameter are derived from that function. All numerical summaries, visual summaries, and convergence diagnostics are designated for that function.

5.6.2.1 Numerical Summary

In this sub-section, we look at different quantities of interest and their numerical values for a generalized log-logistic accelerated failure time model using a McMC sample of posterior properties. The McMC simulation results of the posterior mean, posterior standard deviation, time series standard error, Naïve standard error, five-point summary statistics, 2.5% percentile, 97.5% percentile, variance, mode, skewness, kurtosis, and the credible interval following by the highest probability density (HPD) interval and the posterior probability that the associated parameter is positive. The naïve standard error is defined as the mean standard error that incorporates simulation error rather than posterior uncertainty.

$$naiveSE = \frac{\text{posterior } SD}{\sqrt{n}} \tag{5.41}$$

The time series SE adjusts the "naïve" SE for autocorrelation.

The posterior summary for the generalized log-logistic AFT model parameters based on larynx data set are shown in Table 5.7. The posterior probability that the associated parameter is positive is shown in the last row of Table 5.7. A probability of 0.5 suggests that a positive and a negative value of the parameter is equally likely. Once we have stored the posterior sample of each parameter, we can use mean (betal > 0). to calculate the posterior probability, for example for β_1.

Table 5.7: Numerical summaries of the posterior properties for GLL AFT model under non-informative priors.

Characteristics	Pars								
	alpha	betal	beta 2	beta 3	beta4	beta5	beta 6	eta	kappa
Mean	1.216	−2.312	0.148	0.882	1.918	0.192	−0.085	0.933	1.063
SD	0.167	0.461	0.507	0.418	0.530	0.174	0.183	0.275	0.298
Time series SE	0.002	0.005	0.004	0.004	0.005	0.001	0.002	0.002	0.003
Naïve SE	0.001	0.003	0.004	0.003	0.004	0.001	0.001	0.002	0.002
Maximum	2.149	−0.410	2.215	2.688	4.631	0.904	0.606	2.365	2.653
97.5th percentile	1.577	−1.388	1.144	1.731	3.017	0.547	0.259	1.525	1.732
Q3	1.318	−2.009	0.480	1.156	2.255	0.306	0.039	1.109	1.239
Medium (Q2)	1.203	−2.316	0.157	0.868	1.905	0.187	−0.079	0.909	1.034
Q1	1.101	−2.626	−0.188	0.598	1.558	0.073	−0.201	0.734	0.847
2.5th percentile	0.926	−3.202	−0.872	0.087	0.924	−0.141	−0.463	0.463	0.579
Minimum	0.724	−4.137	−2.400	−0.876	−0.017	−0.545	−0.958	0.157	0.295
Mode	1.150	−2.300	0.250	0.900	1.750	0.150	0.050	0.900	0.900
Variance	0.028	0.213	0.034	0.175	0.280	0.030	0.033	0.076	0.089
Skewness	0.507	0.050	−0.178	0.150	0.200	0.120	−0.205	0.495	0.662
Kurtosis	0.664	0.089	2.215	0.150	0.174	0.179	0.218	0.355	0.662
95% credible	(0.926,	(−3.202,	(−0.872,	(0.087,	(0.924,	(−0.141,	(−0.463,	(0.463,	(0.579,
interval	1.577)	1.388)	1.144)	1.731)	3.017)	0.547)	0.259)	1.525)	1.732)
95% HPD	(0.913,	(−3.200,	(−0.844,	(0.075,	(0.898,	(−0.141,	(−0.454,	(0.439,	(0.545,
interval	1.553)	−1.386)	1.157)	1.715)	2.982)	0.545)	0.266)	1.483)	1.676)
P(. ¿ 0 data)	1.000	0.000	0.629	0.986	0.999	0.868	0.331	1.000	1.000

5.6.3 Relative Medium (RM)

For survival data, the median survival time is an essential summary measure for time-to-event data (Christensen et al., 2010). Any accelerated failure time model may simply estimate this quantiy. The relative medium (RM) of two individuals with explanatory variables $\boldsymbol{x}_1 \boldsymbol{and} \boldsymbol{x}_2$ is the most popular effect measure obtained from the median survival time.

The RM compares the median of both individuals' survival times and is expressed as:

$$RM\left(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{\beta}\right) = \exp\left\{\left(\boldsymbol{x}_1 - \boldsymbol{x}_2\right)^T \beta\right\}.$$

For patients with same age and year of diagnosis, we are interested in the median survival times for individuals in stages 2,3 or 4 relatives to stage 1 . Table 5.8 ,

Table 5.8: Comparison between the relative medium between two men of the year of diagnosis (diagyr) and same age but in different stages.

Posterior Properties	Stages 3 and 2	Stages 4 and 3	Stages 4 and 2
Mean	2.369	3.222	6.98
Standard deviation	1.378	1.853	4.942
Naïve SE	0.011	0.015	0.040
Time series SE	0.011	0.014	0.040
2.5% percentile	0.779	1.043	1.928
Q1	1.461	2.007	3.935
Medium (Q2)	2.038	2.806	5.720
Q3	2.891	3.949	8.516
97.5 percentiles	5.886	7.957	19.243

illustrates the posterior properties of the relative medium between two men of the same year of diagnosis (diagyr) and same age but in different stages.

5.7 Model Comparison

We performed a comparison of the proposed models for both Bayesian and frequentist approaches. The comparison considered the generalizd log-logistic AFT model, and its special cases. Addtitionally, we recommend using multi-parameter hazard-based regression, rather than the traditional approach of having a single covariate-dependent (scale) -parameter, in line with our proposed flexible baseline distribution. While there are many options, we recommend using covariate-dependent choice to manage the baseline distribution selection across all distributional parameters, which covers the AFT framework.

5.7.1 Frequentist Model Comparison

In this sub-section, we used the AIC to compare the frequentist approach fits (i.e., MLE) approach for the proposed models see Table 5.9.

Table 5.9: GLL, Weibull, and LL AFT fits for the Larynx cancer data set.

Summaries			
Models	GLL-AFT (AIC = 294.737)	Weibull-AFT (AIC = 296.759)	Log-logistic-AFT (AIC = 297.231)
Beta 1 (intercept)	−1.116	−1.918	0.000
Beta 2 (stage 2)	0.157	0.162	0.182
Beta 3 (stage 3)	0.635	0.593	0.645
Beta 4 (stage 4)	1.632	1.587	1.754
Beta 5 (age)	0.184	0.186	0.255
Beta 6 (diagyr)	−0.041	−0.042	−0.035
Alpha	1.167	1.121	1.465
kappa	0.585	0.603	0.467
Eta	0.125		

5.7.2 Bayesian model selection

To give some preferenes to the alternatives. It is desirable to select an acceptable approximation model. The most frequently used tool for model comparison in Bayesian inference is Deviance Information Criterion (DIC) (Spiegelhalter et al., 2014). The DIC values for the GLL-AFT model and two of its sub-models including the LL-AFT and W-AFT are given in Table 5.10. Table 5.10 illustrates that the flexibility provided by an additional shape parameter of the GLL distribution leads to its overall superiority over the Weibull and loglogistic baseline models.

Table 5.10: Bayesian model comparison for the GLL-AFT, its special cases

Model	**DIC**
GLL-AFT	292.521
LL-AFT	296.776
W-AFT	299.560

Table 5.11: PSRF and MPSRF for the distributional parameters and the regression coefficients.

Parameter	Point estimate	Upper C.I.
Alpha	1.0	1.01
Beta 1 (intercept)	1.0	1.01
Beta 2 (stage 2)	1.0	1.0
Beta 3 (stage 3)	1.0	1.0
Beta 4 (stage 4)	1.0	1.0
Beta 5 (age)	1.0	1.0
Beta 6 (diagyr)	1.0	1.0
Eta	1.0	1.0
Kappa	1.0	1.0
Multivariate PSRF	**1.0**	

5.8 McMC Convergence Diagnostics

The McMC convergence of the model parameters are checked by using Gelman-Rubin diagnostic test as illustrated in Table 5.11and graphical tools including autocorrelation, running mean, density and trace plot as shown in Figures 5.1 to 5.4.

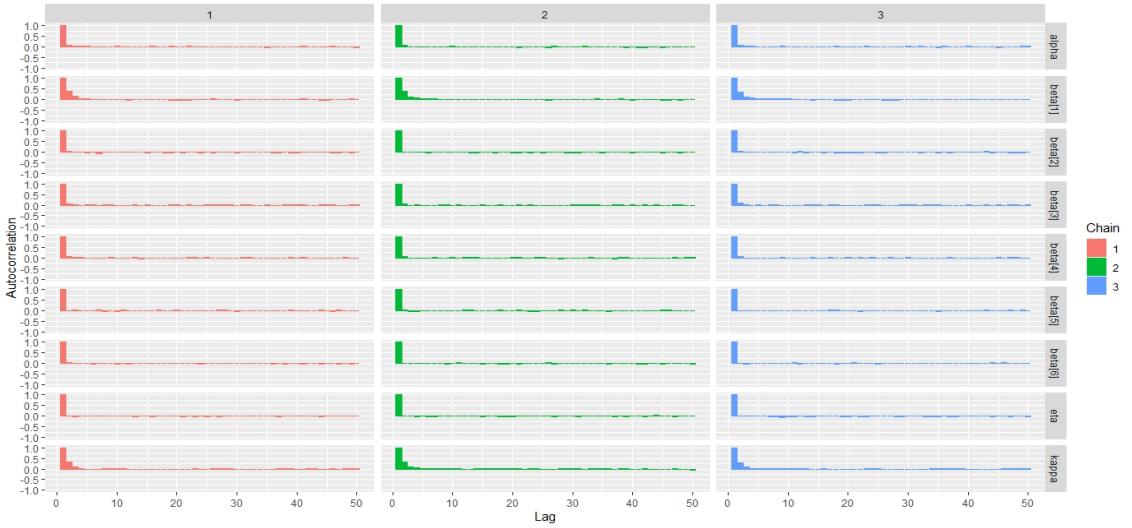

Figure 5.1: Autocorrelation plots of the regression coefficients and distributional parameters for the generalized log-logistic AFT model.

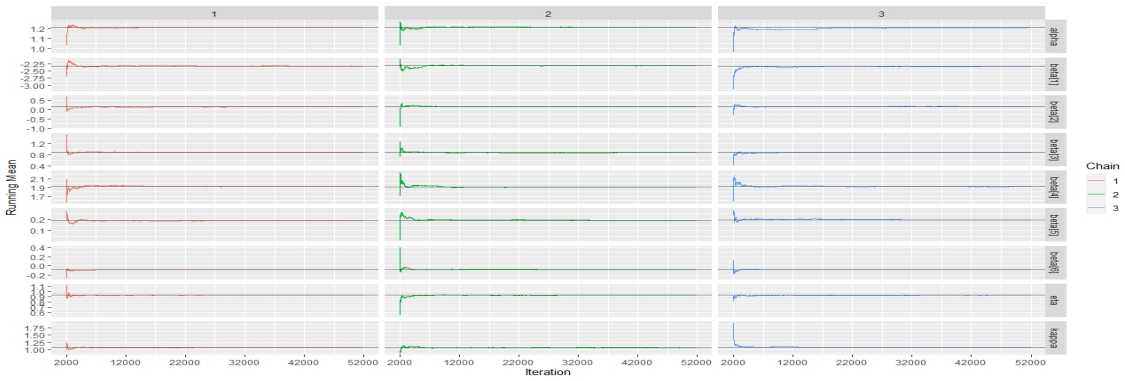

Figure 5.2: The running mean plots of the regression coefficients and distributional parameters for the generalized log-logistic AFT model.

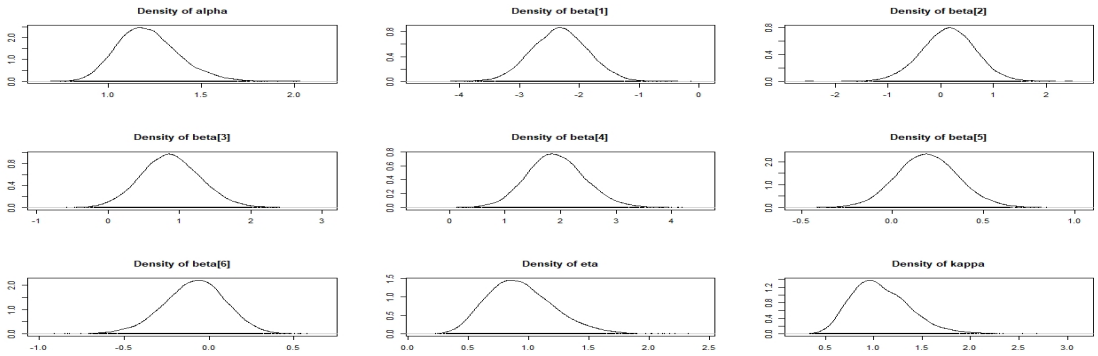

Figure 5.3: Density plots of the regression coefficients and distributional parameters for the generalized log-logistic AFT model.

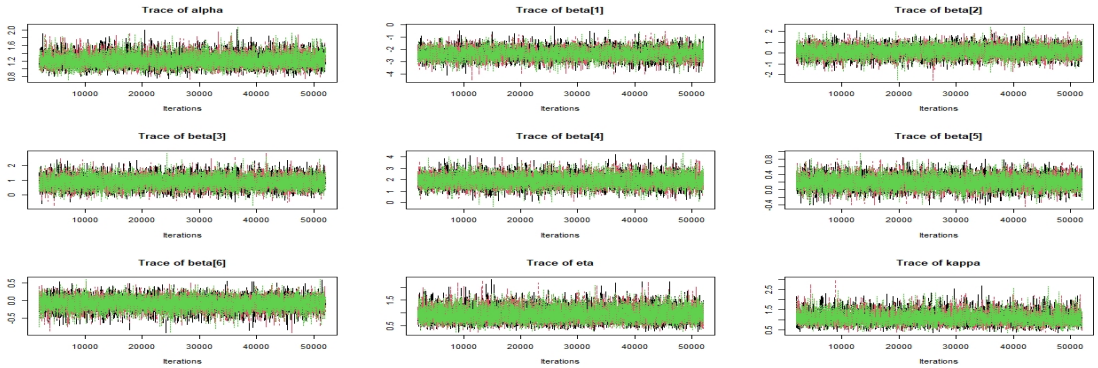

Figure 5.4: Trace plots of the regression coefficients and distributional parameters for the generalized log-logistic AFT model.

5.9 Concluding Remarks

In this chapter, we proposed the three-parameter generalized log-logistic distribution. It is a straightforward modification of the LL distribution. The proposed model modifies several classical distributions commonly used in the literature of survival analysis, such as the exponential, Weibull, LL, and Burr XII distributions. It has greater flexibility than the Weibull, LL, Burr XII, and exponentiated Weibull models. We propose a generalized LL AFT regression model that is more suitable for analyzing and modeling complete and incomplete (i.e., censored) survival data based on this proposed distribution, which is also closed under the PH regression framework. The proposed AFT model serves as an essential modification to several existing AFT models and could be a valuable addition to the literature. Hence, we developed the frequentist (using MLE) and Bayesian estimation technique for inference. Several simulation studies are performed for considering different hazard rate shapes, different parameter settings, and different censoring percentages. A comparative study was conducted to investigate the tendency of generalized log-logistic in modeling survival data, and the results show that it is a more accurate fit than its sub-models or either comparable. A real-world data relating to larynx-cancer patients showed that the proposed model could be valuable in appropriately describing various shapes of survival data and can be a good alternative to the Weibull distribution, which is also the only probability distribution that is closed under the PH and AFT regression models and the log-logistic AFT regression model. These findings led us to conclude that the generalized log-logistic AFT regression model, as introduced in this paper, has the potential to be a very helpful survival regression model in describing various forms of survival data and can be of interest in reliability and survival applications, since we can adopt goodness-of-fit tests for the several widely known AFT models as special cases.

CHAPTER 6

A Flexible Parametric Accelerated Hazard Model: An Application to Censored Lifetime Data with Crossing Survival Curves

In this chapter, we present our fifth manuscript [1] about a flexible parametric accelerated hazard model. Note that the materials of this chapter have been reproduced from our article which is under review.

6.1 Introduction

In the analysis of lifetime data, hazard-based regression models have played a pivotal role. Such models produce a much more versatile framework for modeling survival data. They also make it conceivable to easily interpret the parameters from a practical perspective. When using regression models to analyze lifetime data, the Cox proportional hazard (PH) (Cox, 1972; Kalbfleisch, 1978) model is the most widely assumed semi-parametric framework. The PH model's main assumption is that the hazard ratios are proportional over time. When such assumptions are not validated by data, alternative survival regression models, such as the accelerated failure time (AFT) (Buckley and James, 1979; Komárek and Lesaffre, 2008), and proportional odds (PO) (Bennett, 1983a) models might be applied in the analysis. However, none of them are appropriate for capturing lifetime data with crossing

[1] Muse, A. H., Chesneau, C., Ngesa, O., & Mwalili, S. (2022). Flexible Parametric Accelerated Hazard Model: Simulation and Application to Censored Lifetime Data with Crossing Survival Curves. Mathematical and Computational Applications, 27(6), 104.

survival and hazard curves (Demarqui and Mayrink, 2021).

This kind of issue is frequently associated with clinical trials, including control and treatment groups. The survival function (SF) of one group may degrade swiftly while the SF of the other group decays slowly. The curves tend to meet at some point, resulting in an inversion in terms of who is on the bottom/top. The study of this change is essential in many clinical studies because determining the crossing time reveals when the target treatment for an illness can be judged beneficial (Demarqui and Mayrink, 2021).

In practice, time-to-event data with crossing survival curves can occur for a variety of reasons. Crossing survival curves, according to Breslow et al. (1984), may occur when a treatment has an early rapid benefit and then becomes equally or worse than placebo treatment after such a time period. Additionally, as described in Diao et al. (2013), crossing survival curves may occur in clinical studies when a particular intensive treatment (i.e., surgery) may have negative consequences at first but show good results in the long term.

Several techniques have been presented in the literature to handle this crossover feature in time-to-event data. The most often used are based on regression coefficients that change over time; see, for instance, Egge and Zahl (1999), Putter et al. (2005), Shyur et al. (1999), and (Zhang et al., 2018). Two recent works considering the modeling and analysis of time-to-event data with crossing survival curves are (Demarqui and Mayrink, 2021; Demarqui et al., 2019). For this type of problem, Chen and Wang (2000) developed a semi-parametric two-sample framework. The two-sample feature refers to a scenario in which there is a control, and a treatment group, which can be readily represented by a binary variable. The AH model is an intriguing choice because it formulates similarly to the PH and AFT models. In their model, they leave the baseline hazard rate function (hrf) undefined. As an alternative to the PO or AFT models, their model relaxes the proportional hazard assumption while still allowing for the inclusion of both time-independent and

time-dependent factors.

Although they offered an exploratory visual examination of the model's suitability, they did not completely cover statistical model checking of the proposed model. Chen et al. (2003) presented the AH model and its applicability to censored survival data. They used the AH model to analyze real data from a randomized clinical study of biodegradable carmustine polymers for the treatment of brain cancer. This analysis illustrated the model's useful applications and the recommended test statistics.

The semi-parametric AH model estimators, on the other hand, include the unknown distribution in the asymptotic variance. Thus, numerically demanding approaches are required to make an inference about this parameter. As a result, Lee (2009) suggested a straightforward estimation method for the semi-parametric AH model in which estimators are asymptotically normal with a distribution-free asymptotic variance. This also yields several lack-of-fit tests. These tests are similar to Gill–Schumacher tests in that the estimating functions are assessed at two separate weight functions, generating two estimators that are close to each other. They demonstrated that the estimators and tests perform well for some weight functions using numerical experiments. For more information about the estimators and tests for the semi-parametric AH model, we refer to (Lee, 2016).

Cox (1972) pioneered the use of semi-parametric hazard-based regression models for univariate time-to-event data with the PH model. Rubio et al. (2019) and Khan (2018) presented two influential papers that propose the use of extended lifetime distributions to substitute the baseline hazard in a time-to-event analysis. The formulation of parametric hazard-based regression models is a central issue in Lawless (2011). The authors explored the benefits of using parametric hazard-based regression models. It is noticed that the baseline-modified distribution should be chosen based on its flexibility to incorporate varied failure rate shapes. A few examples include: Muse et al. (2022d), Muse et al. (2022g), Ashraf-Ul-Alam and Khan (2021),

Alvares and Rubio (2021), Muse et al. (2022a), Al-Aziz et al. (2022), and Khan and Khosa (2016).

Despite the numerous advantages of the semi-parametric AH framework, its implementation in applications appears to be restricted, owing to the technical difficulties in implementing theoretical breakthroughs. Estimation for the covariance matrices is challenging when the data are censored because the asymptotic covariance matrices for the regression estimators in this model involve the unknown baseline hrf and its derivative. However, censored data present a new technological barrier. Numerically demanding approaches, such as resampling techniques, can be used to approximate the covariance matrices. However, they are inefficient in actual settings due to their high computing cost (Chen et al., 2014).

The current study presents a fully parametric hazard-based regression model to fit the AH model to address the aforementioned concerns. The fundamental idea is to represent the baseline hazard by using a generalized log-logistic (GLL) distribution that is closed under both the AFT (Muse et al., 2022a) and PH (Muse et al., 2022g) frameworks and may incorporate various hazard rate shapes data including monotone and non-monotone shapes. Another advantage of the baseline is that it encompasses some of the most parametric distributions used in reliability and survival studies, such as log-logistic (LL), Burr XII with both 2-parameter and 3-parameter cases, Weibull, and exponential distributions. The shared tractability of parametric regression models and the adaptability of semi-parametric regression models is another appealing aspect of the suggested parametric AH model.

Thus, the main contribution of this study is to introduce and study a novel, flexible, parametric AH model to incorporate right-censored lifetime data with crossing survival curves. This is done by assuming the GLL lifetime distribution to deal with the baseline hazard in the parametric AH model. To the best of the author's knowledge, we emphasize that using the parametric AH model with GLL baseline distribution hazard to extend the original AH semi-parametric model has never been consid-

ered in the literature. The methods are studied by using the classical and Bayesian frameworks for a more comprehensive presentation of models for all statistical audiences to consider. A detailed simulation study is also being developed. This entails introducing one binary and one continuous covariate into the baseline hazard. The reader should be aware that the majority of the single covariate scenarios have been researched in prominent references, such as (Diao et al., 2013).
Additionally, the following are some significant benefits of the methodology proposed here.

i. It possesses the adaptability of parametric survival regression models.

ii. It offers a continuous sf that makes it simple to find where two survival curves overlap.

iii. It allows different shapes for the hrf and has the tractability of a parametric survival regression model.

The following is how the chapter is structured. Section 6.2 discusses the formulation of the parametric AH model and associated probabilistic functions. The proposed parametric AH model with generalized log-logistic baseline hazard and its sub-models are presented in Section 6.3. Section 6.4 discusses the model inferential procedures. Section 6.5 performs the simulation studies. Section 6.6 demonstrates a real-life, right-censored cancer dataset with crossed survival curves. Section 6.7 concludes the study with some farewell remarks and suggests future research.

6.2 Model Formulation

Let T be a non-negative random variable representing the time until the occurrence of an event of interest. In order to accommodate survival data with crossing of hazard and survival curves, Chen and Wang (2000) proposed a hazard-based regression

model known as the accelerated hazard model that is expressed as follows:

$$h(t;x) = h_0\left(t\psi\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)\right) = h_0\left(te^{x_i'\beta}\right) \tag{6.1}$$

where; $\psi\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right) = e^{x_i'\beta}$ is the link function of the covariates, $x_i = (x_1, x_2, \ldots, x_p)$ is a set of covariates, and $\beta' = (\beta_1, \beta_2, \ldots, \beta_p)$ is a vector of regression coefficients, and $h_0(t)$ corresponds to the baseline hazard function.

In this model, $e^{x_i'\beta}$ characterizes how the covariates x_i alter the time scale of the underlying hazard function. For example, $\beta < 0$ or $\beta > 0$ imply deceleration or acceleration of the time scale for the hazard, respectively. As an example, if there exists one covariate, x_i, that takes a value of 0 for a control group, and 1 for a treatment group, then $e^{\beta} = \frac{1}{2}$ means that the hazard of the treatment group progresses in half the time as those in the control group. Similarly, $e^{\beta} = 2$ means that the hazard of the treatment group progresses in twice the time as those in the control group $e^{\beta} = 1$ implies no differences between the groups.

The AH model offers some appealing and intriguing characteristics. The AH model, unlike the AFT and PH models, can handle the crossing of survivor and hazard curves (Zhang and Peng, 2009). Furthermore, the AH framework enables both the control and treatment groups' hazard curves to begin at the same time point. This is especially beneficial in randomized controlled trials, because it is more reasonable to hypothesize that the risk or hazard between groups is comparable at $t = 0$.

The inability of the AH model to handle situations where the hazard function is constant over time is a limitation that is not shared by the AFT, and PH models (e.g., exponential distribution) (Chen et al., 2014). As a result, before implementing this model, it is crucial to assess for non-constancy of the baseline function. The AH model, like the AFT and PH models, has coincidences when the baseline hazard rate is a Weibull distribution (Chen and Jewell, 2001).

Alternatively, the parametric AH model can be written in terms of the cumulative

hazard rate function as follows:

$$H(t \mid x) = H_0\left(te^{x'\beta}\right)e^{-x'\beta} \tag{6.2}$$

The other probabilistic functions for the parametric AH model, associated with Equation (6.2), can be expressed as follows:
The sf for the parametric AH model is

$$S(t;x) = \left[S_0\left(te^{x'\beta}\right)\right]^{e^{-x'\beta}}, \tag{6.3}$$

where $S_0(t)$ denotes the baseline sf. The cumulative distribution function (cdf) for the parametric AH model is

$$F(t;x) = 1 - \left[S_0\left(te^{x'\beta}\right)\right]^{e^{-x'\beta}}. \tag{6.4}$$

The probability density function (pdf) for the parametric AH model is

$$f(t;x) = f_0\left(te^{x'\beta}\right)\left[S_0\left(te^{x'\beta}\right)\right]^{e^{-x'\beta}}, \tag{6.5}$$

where $f_0(t)$ denotes the baseline pdf.

6.3 The Proposed Model

There are several approaches to expressing parametric hazard-based regression models. The AH model formulation is one such strategy. The GLL hazard-based regression model can be written in the context of the AH framework by substituting the exponential function for the link function in Equation (6.1). We recall that the hrf

under the AH framework is computed as follows:

$$h(t) = h_0\left(te^{x'\beta}\right).$$

We begin with the GLL baseline distribution hrf with parameters α, η, and k (with the AH model notations). The hrf with an explanatory variable vector x is as follows:

$$h(t;\theta,\beta,x) = h_0\left(te^{\boldsymbol{x}'\beta};\theta\right) = \frac{\alpha k\left(kt^*\right)^{\alpha-1}}{1+\left(\eta t^*\right)^{\alpha}}, \tag{6.6}$$

which is the GLL hrf with $t^* = te^{x'\beta}$ once more. In addition, the other survival probabilistic functions for the GLL–AH framework are expressed as follows.
The sf for the GLL–AH model is

$$S(t;\theta,\beta,x) = \left[S_0\left(te^{x'\beta};\theta\right)\right]^{e^{-x'\beta}} = \left[1+\left(\eta te^{x'\beta}\right)^{\alpha}\right]^{\frac{k^{\alpha}e^{-x'\beta}}{\eta^{\alpha}}}. \tag{6.7}$$

The cdf for the GLL–AH model is

$$F(t;\theta,\beta,x) = 1-\left[S_0\left(te^{x'\beta};\theta\right)\right]^{e^{-x'\beta}} = 1-\left[1+\left(\eta te^{x'\beta}\right)^{\alpha}\right]^{\frac{k^{\alpha}e^{-x'\beta}}{\eta^{\alpha}}}. \tag{6.8}$$

The chf for the GLL–AH model is

$$H(t;\theta,\beta,x) = H_0\left(te^{x'\beta};\theta\right)e^{-x'\beta} = \left(\frac{k^{\alpha}}{\eta^{\alpha}}\log\left[1+\left(\eta te^{x'\beta}\right)^{\alpha}\right]\right)e^{-x'\beta}. \tag{6.9}$$

The pdf for the GLL-AH model:

$$f(t;x) = f_0\left(te^{x'\beta}\right)\left[S_0\left(te^{x'\beta}\right)\right]^{e^{-x'\beta}} = \frac{\alpha k\left(kte^{x'\beta}\right)^{\alpha-1}}{\left[1+\left(\eta te^{x'\beta}\right)^{\alpha}\right]^{\frac{k^{\alpha}}{\eta^{\alpha}}+1}}\left[1+\left(\eta te^{x'\beta}\right)^{\alpha}\right]^{\frac{k^{\alpha}e^{-x'\beta}}{\eta^{\alpha}}} \tag{6.10}$$

6.3.1 Submodels

The proposed parametric hazard-based GLL–AH model framework has three submodels that are also closed under the AH framework.

6.3.1.1 Submodel I: $\eta = 1$

If we put $\eta = 1$ in Equation (6.6), we get the hrf of the BXII–AH model, which is expressed mathematically as

$$h(t;x) = \frac{\alpha k \left(kte^{x'\beta}\right)^{\alpha-1}}{1 + (te^{x'\beta})^{\alpha}}. \tag{6.11}$$

6.3.1.2 Submodel II: $\eta = k$

If we put $\eta = k$ in Equation (6.6), we are referred to the hrf of the LL–AH model, which is stated mathematically as

$$h(t;x) = \frac{\alpha k \left(kte^{x'\beta}\right)^{\alpha-1}}{1 + (kte^{x'\beta})^{\alpha}}. \tag{6.12}$$

6.3.1.3 Submodel III: $\eta^{\alpha} \to 0$.

If we put $\eta^{\alpha} \to 0$ in Equation (6.6), we are referred to the hrf of the W–AH model, which is stated mathematically as

$$h(t;x) = \alpha k \left(kte^{x'\beta}\right)^{\alpha-1}. \tag{6.13}$$

6.4 Inferential Procedures

In this section, the unknown parameters of the proposed parametric accelerated hazard model with generalized log-logistic baseline hazard are estimated using classical approach (via maximum likelihood method) and Bayesian approach using non-

informative priors.

6.4.1 Classical Approach

We are concerned in this subsection with a full likelihood function for the proposed parametric AH model. The likelihood function is an important component not only in the Bayesian approach but also in classical inference, in which the standard approach for estimating parameters involves maximizing it. Consider both noninformative and independent (right) censorship.

Suppose there are n individuals with survival times denoted by $T_1, T_2, \ldots, T_n$. Assuming that the data are subject to right censoring, we observe $t_i = \min(T_i, RC_i)$, where $RC_i > 0$ being the censoring time for individual i. Letting $\delta_i = \mathrm{I}(T_i \leq RC_i)$ that equals 1 if $T_i \leq RC_i$ and 0 otherwise, the observed data for individual i consists of $\{t_i, \delta_i, x_i\}, i = 1, 2, \ldots, n$, where t_i is a survival time or censoring time according to whether $\delta_i = 1$ or 0 , respectively, and $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})'$ is a $p \times 1$ column vector of external covariates.

When considering a parametric AH model, the censored likelihood function can be written as follows:

$$\begin{aligned}
L(\theta, \beta; D) &= \prod_{i=1}^{n} \left[f\left(t_i; \theta, \beta, x_i\right)\right]^{\delta_i} \left[S\left(t_i; \theta, \beta, x_i\right)\right]^{1-\delta_i} \\
&= \prod_{i=1}^{n} \left[\frac{h\left(t_i; \theta, \beta, x_i\right)}{S\left(t_i; \theta, \beta, x_i\right)}\right]^{\delta_i} \left[S\left(t_i; \theta, \beta, x_i\right)\right]^{1-\delta_i} \\
&= \prod_{i=1}^{n} \left[h\left(t_i; \theta, \beta, x_i\right)\right]^{\delta_i} S\left(t_i; \theta, \beta, x_i\right) \qquad (6.14) \\
&= \prod_{i=1}^{n} \left[h\left(t_i; \theta, \beta, x_i\right)\right]^{\delta_i} \exp\left[-H\left(t_i; \theta, \beta, x_i\right)\right] \\
&= \prod_{i=1}^{n} \left[h_0\left(t_i e^{x_i'\beta}; \theta\right)\right]^{\delta_i} \exp\left[-H_0\left(t_i e^{x_i'\beta}; \theta\right) e^{-x_i'\beta}\right],
\end{aligned}$$

where $D = (t_i, \delta_i, x_i, i = 1, 2, \ldots, n)$ represents the observed data, which includes

survival times, censoring time, and covariates. In our expression, we recall that θ is the vector of baseline distributional parameters, and β is the regression coefficients. An iterative optimization approach can be used to produce the MLE (e.g., the Newton–Raphson algorithm). Because the MLEs are approaching normalcy, various hypothesis tests and interval constructions of model parameters are conceivable.

The log-likelihood function is expressed as follows:

$$\ell(\theta, \beta; D) = \sum_{i=1}^{n} \delta_i \log \left[h_0 \left(t e^{x_i' \beta}; \theta \right) \right] - \sum_{i=1}^{n} H_0 \left(t_i e^{x_i' \beta}; \theta \right) e^{-x_i' \beta}. \tag{6.15}$$

The GLL–AH model's full log-likelihood function is expressed as follows:

$$\begin{aligned} \ell(\theta, \beta; D) = & \sum_{i=1}^{n} \delta_i \log(\alpha) + \sum_{i=1}^{n} \delta_i \alpha \log(k) + (\alpha - 1) \sum_{i=1}^{n} \delta_i \log \left(t_i e^{x_i' \beta} \right) \\ & - \sum_{i=1}^{n} \delta_i \log \left[1 + \left(\eta t_i e^{x_i' \beta} \right)^{\alpha} \right] - \left(\frac{k}{\eta} \right)^{\alpha} \sum_{i=1}^{n} e^{-x_i' \beta} \log \left[1 + \left(\eta t_i e^{x_i' \beta} \right)^{\alpha} \right]. \end{aligned} \tag{6.16}$$

To obtain the MLE of $\theta' = (k, \alpha, \eta)$, and β, we can directly maximize Equation (6.16) with respect to (k, α, η), and β. Alternatively, we can express the first derivative of the log-likelihood function in order to solve the nonlinear equations below for the log-likelihood function's first derivative.

With this aim, let us set $\varphi = (k, \alpha, \eta, \beta)$. Then the first derivatives of the log-

likelihood functions are as follows:

$$\begin{aligned}\frac{\partial \ell(\varphi)}{\partial \alpha} &= \frac{1}{\alpha}\sum_{i=1}^{n}\delta_i + \sum_{i=1}^{n}\delta_i \log(k) + \sum_{i=1}^{n}\delta_i \log\left(t_i e^{x_i'\beta}\right) \\ &\quad - \sum_{i=1}^{n}\delta_i \frac{\left(\eta t_i e^{x_i'\beta}\right)^{\alpha}\log\left(\eta t_i e^{x_i'\beta}\right)}{1+\left(\eta t_i e^{x_i'\beta}\right)^{\alpha}} \\ &\quad - \left(\frac{k}{\eta}\right)^{\alpha}\log(k)\sum_{i=1}^{n} e^{-x_i'\beta}\log\left[1+\left(\eta t_i e^{x_i'\beta}\right)^{\alpha}\right] \\ &\quad + \left(\frac{k}{\eta}\right)^{\alpha}\log(\eta)\sum_{i=1}^{n} e^{-x_i'\beta}\log\left[1+\left(\eta t_i e^{x_i'\beta}\right)^{\alpha}\right] \\ &\quad - \left(\frac{k}{\eta}\right)^{\alpha}\sum_{i=1}^{n}\frac{e^{-x_i'\beta}\left(\eta t_i e^{x_i'\beta}\right)^{\alpha}\log\left(\eta t_i e^{x_i'\beta}\right)}{1+\left(\eta t_i e^{x_i'\beta}\right)^{\alpha}},\end{aligned} \tag{6.17}$$

$$\begin{aligned}\frac{\partial \ell(\varphi)}{\partial \eta} &= -\left(\frac{\alpha}{\eta}\right)\sum_{i=1}^{n}\delta_i \frac{\left(\eta t_i e^{x_i'\beta}\right)^{\alpha}}{1+\left(\eta t_i e^{x_i'\beta}\right)^{\alpha}} \\ &\quad + \left(\frac{\alpha}{\eta}\right)\left(\frac{k}{\eta}\right)^{\alpha}\sum_{i=1}^{n} e^{-x_i'\beta}\log\left[1+\left(\eta t_i e^{x_i'\beta}\right)^{\alpha}\right] \\ &\quad - \left(\frac{\alpha}{\eta}\right)\left(\frac{k}{\eta}\right)^{\alpha}\sum_{i=1}^{n}\frac{e^{-x_i'\beta}\left(\eta t_i e^{x_i'\beta}\right)^{\alpha}}{1+\left(\eta t_i e^{x_i'\beta}\right)^{\alpha}},\end{aligned} \tag{6.18}$$

$$\frac{\partial \ell(\varphi)}{\partial k} = \left(\frac{\alpha}{k}\right)\sum_{i=1}^{n}\delta_i - \left(\frac{\alpha}{k}\right)\left(\frac{k}{\eta}\right)^{\alpha}\sum_{i=1}^{n} e^{-x_i'\beta}\log\left[1+\left(\eta t_i e^{x_i'\beta}\right)^{\alpha}\right] \tag{6.19}$$

and

$$\begin{aligned}\frac{\partial \ell(\varphi)}{\partial \beta_j} &= (\alpha-1)\sum_{i=1}^{n}\delta_i x_{ij} - \alpha\sum_{i=1}^{n}\delta_i x_{ij}\frac{\left(\eta t_i e^{x_i'\beta}\right)^{\alpha}}{1+\left(\eta t_i e^{x_i'\beta}\right)^{\alpha}} \\ &\quad + \left(\frac{k}{\eta}\right)^{a}\sum_{i=1}^{n} x_{ij}\log\left[1+\left(\eta t_i e^{x_i'\beta}\right)^{\alpha}\right].\end{aligned} \tag{6.20}$$

6.4.2 Bayesian Approach

In this subsection, the prior distributions for the parameters of the proposed model are first established, and these distributions are then multiplied by the likelihood function to create the Bayesian model.

6.4.2.1 Prior Distribution

The formulation of a prior distribution is a crucial step in every Bayesian approach. This is especially true for fully parametric survival regression models. Because we lack prior knowledge from historical data or from prior experiments, we set the prior scenario in this study using a noninformative independent gamma distribution, Gamma $(10, 10)$, as the baseline distribution parameters. Gamma distributions are flexible and include noninformative priors (uniform) and the marginal priors distribution for each regression coefficient is taken as a normal distribution centered at zero with a wide known variance $(0, 100)$. Numerous study articles in the literature, such as (Muse et al., 2022g,a; Khan, 2018; Al-Aziz et al., 2022; Alvares and Rubio, 2021), take these priors into account. Here, we consider

$$\pi(\alpha) \sim G\left(a_1, b_1\right) = \frac{b_1^{a_1}}{\Gamma\left(a_1\right)} \alpha^{a_1-1} e^{-b_1 \alpha}; a_1, b_1, \alpha > 0, \tag{6.21}$$

$$\pi(\eta) \sim G\left(a_2, b_2\right) = \frac{b_2^{a_2}}{\Gamma\left(a_2\right)} \eta^{a_2-1} e^{-b_2 \eta}; a_2, b_2, \eta > 0, \tag{6.22}$$

$$\pi(k) \sim G\left(a_3, b_3\right) = \frac{b_3^{a_3}}{\Gamma\left(a_3\right)} k^{a_3-1} e^{-b_3 k}; a_3, b_3, k > 0. \tag{6.23}$$

From historical data of the baseline distribution, it is simple to determine the hyperparametric values of the prior distributions (Muse et al., 2021a). When the explanatory variables are assumed to have a prior normal distribution, we have the following regression coefficients:

$$\pi\left(\beta'\right) \sim N\left(a_4, b_4\right). \tag{6.24}$$

The joint prior distribution of all unknown parameters has a pdf given by

$$\pi\left(\alpha, k, \eta, \beta'\right) = \pi(\alpha)\pi(\eta)\pi(k)\pi\left(\beta'\right). \tag{6.25}$$

6.4.2.2 Likelihood Function

The likelihood function for the GLL general hazard model is computed as follows:

$$\begin{aligned} L_{GLL-AH}\left(\theta,\beta;D\right) = &= \prod_{i=1}^{n}\left[h_0\left(t_i e^{x_i'\beta};\theta\right)\right]^{\delta_i}\exp\left[-H_0\left(t_i e^{x_i'\beta};\theta\right)e^{-x_i'\beta}\right] \\ &= \prod_{i=1}^{n}\left[\frac{\alpha k\left(k t_i e^{x_i'\beta}\right)^{\alpha-1}}{1+\left(\eta t_i e^{x_i'\beta}\right)^{\alpha}}\right]^{\delta_i} \\ &\exp\left[-\left\{\frac{k^{\alpha}}{\eta^{\alpha}}\log\left[1+\left(\eta t_i e^{x_i'\beta}\right)^{\alpha}\right]\right\}e^{-x_i'\beta}\right]. \end{aligned} \tag{6.26}$$

6.4.2.3 Posterior Distribution

The joint posterior pdf is expressed as the multiplication of the likelihood function in Equation (6.26) and the prior distribution in Equation (6.25):

$$\begin{aligned} p\left(\alpha,k,\eta,\beta;t\right) \propto &\prod_{i=1}^{n}\left[\frac{\alpha k\left(k t_i e^{x_i'\beta}\right)^{\alpha-1}}{1+\left(\eta t_i e^{x_i'\beta}\right)^{\alpha}}\right]^{\delta_i} \\ &\exp\left[-\left\{\frac{k^{\alpha}}{\eta^{\alpha}}\log\left[1+\left(\eta t_i e^{x_i'\beta}\right)^{\alpha}\right]\right\}e^{-x_i'\beta}\right] \\ &\times \pi\left(\alpha,k,\eta,\beta'\right), \end{aligned} \tag{6.27}$$

where the prior specification for the unknown parameters is represented by the first four terms on the right-hand side of the equation.

The joint posterior pdf is analytically intractable because of how challenging it is to integrate. Therefore, the inference can be supported by the Markov chain Monte Carlo (McMC) simulation methods, including the Gibbs sampler and Metropolis–Hastings algorithms, which can be used to generate samples from which features of the relevant marginal distributions can be inferred.

6.5 Simulation Study

In this section, we offer a thorough Monte Carlo (MC) simulation analysis to assess how well the suggested model performs in terms of estimating the parameters of the baseline distribution and the regression coefficients. There are two inferential techniques used in the analysis.

I. Procedure I: An MLE estimate technique.

II. Procedure II: A Bayesian estimation technique with independent gamma priors for the baseline distribution parameters and a normal prior for the regression coefficients, as well as non-informative priors.

Two explanatory variables in an AH regression framework were considered in all simulations: one binary covariate (x_1) generated from Bernoulli (0.5) distribution and one continuous covariate (x_2) generated from the standard normal distribution. Regression parameter values were chosen to be $\beta = (0.75, 0.5)$ corresponding to the covariate vector $x = (x_1, x_2)'$.

The GLL baseline distribution hazard was used to generate the survival data, and the exponential distribution with a rate parameter equal to the censoring proportion of 10% was used to generate the censoring times.

We were particularly interested in the performance and accuracy of the proposed model's estimators in the simulation exercise, specifically the bias, standard error, and mean square error. The simulation's findings were derived from 500 replications with 50, 100, 300, and 500 samples for each parameter value. The results are shown in Table 6.1, which includes the mean estimate (est), standard error (SE), average bias (AB), mean square error (MSE), and coverage probability for the MLE estimates for both inferential techniques. The estimates' averages are extremely close, and generally, the AB and MSE are less as sample size rises. Additionally, as sample sizes are increased, estimates for all evaluated parameters perform better.

Table 6.1: Simulation study for GLL-AH Regression Model. True values (True), Estimates (Est.), standard error (SE), average bias (AB), mean square error (MSE) and coverage probability (CP 95%) are presented for the parameters.

	True	Est.	SE	AB	MSE	CP	Est.	SE	AB	MSE	$\hat{R}$
					Set I $n = 50$						
M_2					MLE approach				Bayesian		
β_1	0.75	0.800	0.100	0.050	0.037	93.85	0.790	0.002	0.040	0.036	1.002
β_2	0.5	0.558	0.042	0.058	0.024	94.50	0.512	0.003	0.012	0.011	1.002
α	1.50	1.590	0.010	0.090	0.008	95.20	1.505	0.001	0.005	0.003	1.000
k	0.75	0.900	0.435	0.150	0.063	92.05	0.850	0.005	0.100	0.045	1.002
η	1.20	1.265	0.011	0.065	0.046	94.25	1.212	0.000	0.012	0.004	1.003
	True	Est.	SE	AB	MSE	CP	Est.	SE	AB	MSE	$\hat{R}$
					Set II $n = 100$						
M_2					MLE approach				Bayesian		
β_1	0.75	0.790	0.100	0.040	0.036	94.10	0.770	0.001	0.020	0.018	1.000
β_2	0.5	0.530	0.030	0.030	0.024	94.80	0.510	0.002	0.010	0.010	1.001
α	1.50	1.610	0.040	0.110	0.087	93.40	1.553	0.001	0.053	0.041	1.003
k	0.75	0.850	0.250	0.100	0.056	93.20	0.800	0.004	0.050	0.037	1.002
η	1.20	1.250	0.008	0.050	0.034	94.80	1.205	0.000	0.005	0.003	1.001
					Set III $n = 300$						
	True	Est.	SE	AB	MSE	CP	Est.	SE	AB	MSE	$\hat{R}$
M_2					MLE approach				Bayesian		
β_1	0.75	0.78	0.092	0.030	0.032	94.40	0.768	0.001	0.018	0.016	1.000
β_2	0.5	0.525	0.013	0.025	0.021	93.90	0.503	0.001	0.003	0.002	1.000
α	1.50	1.592	0.021	0.042	0.030	93.85	1.506	0.001	0.006	0.006	1.001
k	0.75	0.844	0.212	0.094	0.049	93.46	0.798	0.003	0.048	0.036	1.000
η	1.20	1.252	0.008	0.052	0.034	94.60	1.205	0.000	0.005	0.003	1.001
	True	Est.	SE	AB	MSE	CP	Est.	SE	AB	MSE	$\hat{R}$
					Set IV $n = 500$						
M_2					MLE approach				Bayesian		
β_1	0.75	0.775	0.065	0.025	0.017	95.10	0.752	0.000	0.002	0.002	1.000
β_2	0.5	0.526	0.013	0.026	0.021	94.00	0.503	0.001	0.003	0.002	1.000
α	1.50	1.550	0.040	0.050	0.037	94.70	1.503	0.001	0.003	0.001	1.000
k	0.75	0.825	0.110	0.075	0.048	94.07	0.780	0.003	0.030	0.027	1.001
η	1.20	1.205	0.005	0.005	0.003	95.04	1.203	0.000	0.003	0.001	1.001

We also note that, compared to MLE estimates, Bayesian estimates have a lower SE.

Similar results were obtained from a simulation analysis with around 20% censored observations for each dataset (data not shown). In conclusion, our simulation work has shown that the suggested parametric AH model may prove to be a highly helpful parametric hazard-based regression model to accurately represent survival data with or without crossover survival curves.

6.6 Applications

This section examines a right-censored dataset from an oncology clinical trial with crossover survival curves to show how the proposed parametric AH model can be

used to model lifetime data with crossing survival curves. First, the Rstan package's Bayesian analysis of the AH model and its competing models, such as the PH, PO, and AFT models, is provided. After performing a traditional analysis with the MLE technique, add model comparison. Next, by using a frequentist estimation approach, regression analyses were conducted by using the proposed baseline hazard (GLL), power generalized Weibull (PGW), generalized gamma (GG), exponentiated Weibull (EW), log-logistic (LL), and Weibull (W) distributions as a baseline to AH models, and the fits were compared by using information criteria (Akaike information criterion (AIC), Consistent AIC (CAIC), and Hannan–Quinn information criterion (HQIC)). The GLL–AH and its submodels are then used to do a Bayesian analysis.

6.6.1 Gastric Cancer Dataset

We look at the Gastrointestinal Tumor Study Group's gastric cancer data collection (1982). This dataset has frequently been used in studies involving crossing survival curves, particularly in the field related to survival analysis. A few instances include Demarqui and Mayrink (2021) and Diao et al. (2013). The dataset is freely accessible under the label "gastric" by using the R package AmoudSurv (MUSE et al., 2022). This oncology clinical trial includes 90 patients who have been diagnosed with locally advanced gastric cancer. The patients were randomly assigned to the following groups: (i) a control group, which included 45 patients who got chemotherapy; and (ii) a treatment group, which included 45 patients who received radiation therapy along with chemotherapy. In this study, these patients were followed for around 5 years. For each patient, three variables are reported in the datasets: the response time, which indicates failure (time to death) or right censoring (the censoring proportion in this data set is around 12.22%), a binary failure indicator, which identifies patients who experienced the event of interest, and a group binary indicator with 1, indicating the type of treatment.

Figure 6.1 shows the overall survival curve for the gastric cancer dataset as well as the survival curves for the two types of therapies (chemotherapy vs. chemotherapy mixed with radiotherapy) used to treat locally unresectable gastric cancer. Close inspection reveals crossovers and crossings between the curves, which supports the AH model's efficacy and suitability for this data analysis. The fundamental non-parametric plots for the survival time of the gastric cancer dataset are presented in Figure 6.2.

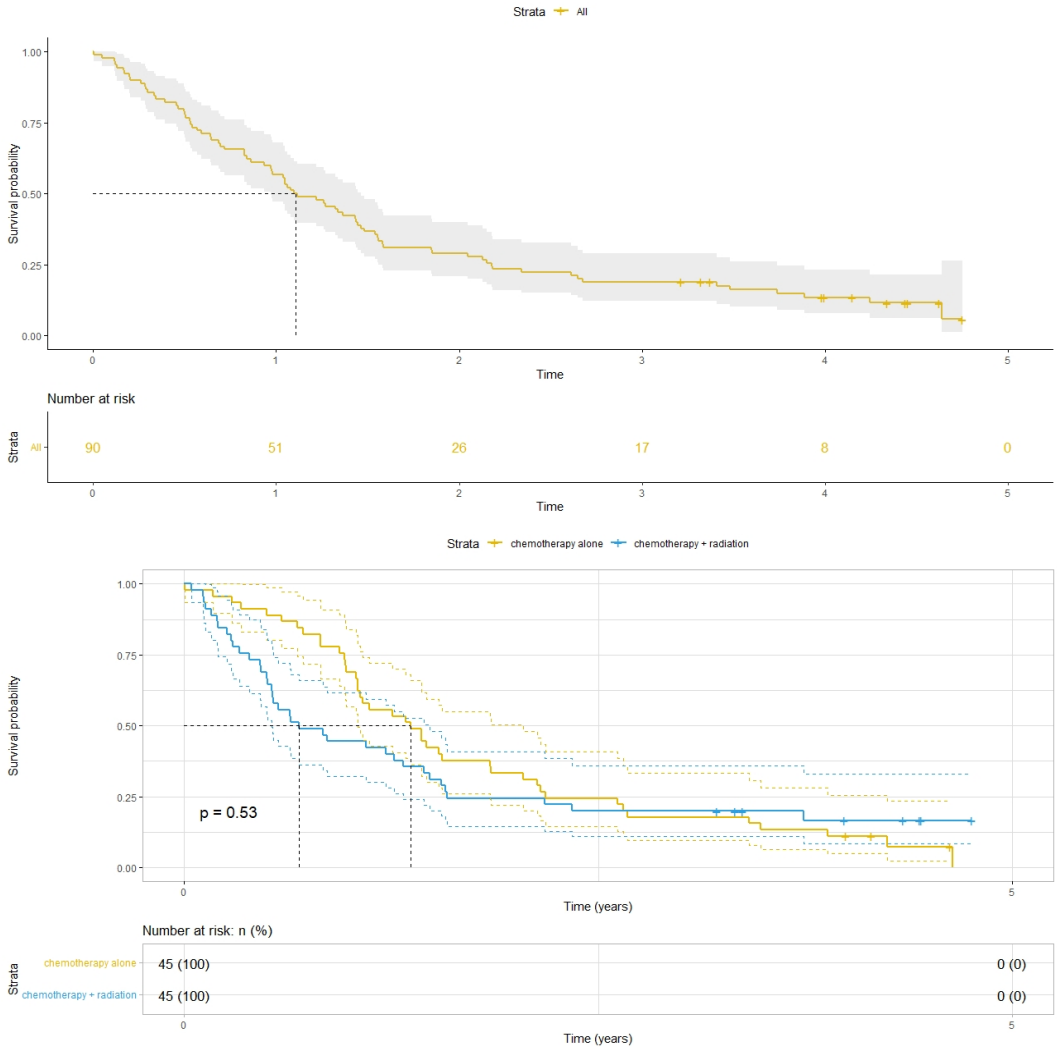

Figure 6.1: Overall survival and the survival curves for the two types of treatments of the gastric cancer data set.

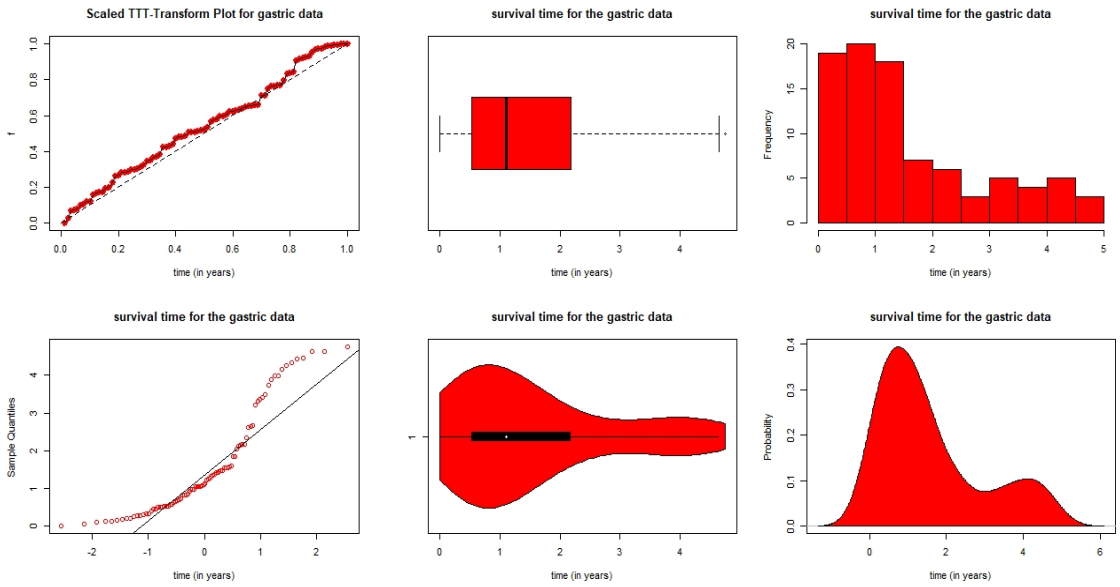

Figure 6.2: Fundamental plots for the survival time of the gastric cancer data set.

6.6.2 Classical Analysis

The MLE estimates for baseline distribution parameters and coefficients of regression for the proposed AH model with different baseline distributions and other survival regression models with the GLL baseline distribution are provided in Tables 6.2 and 6.3.

Table 6.2 summarizes the statistics for the GLL–AH model and other survival regression models, including the PH, PO, and AFT models with all GLL baseline distributions. Based on the information criterion values, we conclude that the GLL–AH model has the lowest AIC, CAIC, and HQIC values compared to the other survival regression models, which indicates that the GLL–AH model outperforms its competing models.

Table 6.2: Results from the fitted proposed fully-parametric accelerated hazard regression model and other survival regression models with all GLL baseline distribution to gastric cancer data set

Models	Parameter(s)	Estimate	SE	AIC	CAIC	HQIC
GLL-AH	β	2.690	0.021	**244.318**	**242.845**	**248.351**
	α	1.505	0.040			
	κ	0.542	0.036			
	η	0.133	0.022			
GLL-PO	β	0.750	0.101	251.816	250.522	255.848
	α	1.382	0.100			
	κ	0.650	0.074			
	η	0.500	0.042			
GLL-PH	β	0.130	0.241	255.565	254.345	259.598
	α	1.302	0.140			
	κ	0.759	0.136			
	η	0.580	0.222			
GLL-AFT	β	0.540	0.135	252.139	250.851	256.171
	α	1.545	0.127			
	κ	0.557	0.106			
	η	0.728	0.231			

The statistics summary under the GLL–AH model, and other AH models with different baseline distributions are presented in Table 6.3. Based on the information criteria values, we deduce that the GLL–AH model beats its rival AH models becaue it has the lowest AIC, CAIC, and HQIC values when compared to the other AH models with various baseline distributions.

Table 6.3: Results from the fitted proposed fully-parametric accelerated hazard regression model with different baseline distributions to gastric cancer data set

Models	Parameter(s)	Estimate	SE	AIC	CAIC	HQIC
GLL-AH	β	2.690	0.021	**244.318**	**242.845**	**248.351**
	α	1.505	0.040			
	κ	0.542	0.036			
	η	0.133	0.022			
PGW-AH	β	1.930	0.082	251.186	249.878	255.218
	α	1.687	0.142			
	κ	0.821	0.066			
	η	2.226	0.102			
GG-AH	β	2.688	0.130	252.645	251.368	256.677
	α	1.821	0.122			
	κ	0.482	0.236			
	η	0.737	0.042			
EW-AH	β	2.066	0.110	252.667	251.390	256.699
	α	0.789	0.212			
	κ	0.911	0.086			
	η	2.283	0.052			
LL-AH	β	1.097	0.020	247.492	246.686	250.517
	α	1.913	0.052			
	κ	1.213	0.019			
LN-AH	β	0.261	0.120	263.830	263.197	266.854
	α	0.065	0.101			
	κ	1.260	0.032			
BXII-AH	β	0.923	0.142	249.144	248.359	252.168
	α	0.880	0.119			
	κ	1.890	0.120			
W-AH	β	2.581	0.214	256.776	256.078	259.800
	α	1.013	0.049			
	κ	1.818	0.112			
G-AH	β	2.367	0.430	255.121	254.406	258.145
	α	1.495	0.039			
	κ	1.252	0.123			

6.6.3 Likelihood Ratio Test

The proposed AH model with the GLL baseline distribution is compared to its submodels, which include the log-logistic AH, Burr-XII AH, and Weibull AH models, by using the likelihood ratio test (LRT). It is required to reduce the number of parameters in a model and evaluate how this affects the model's capacity to match the data in order to draw thorough statistical conclusions about the model. In Table 6.4, statistics and related P-values demonstrate that the GLL–AH model fits the gastric dataset with crossing survival curves better than its submodels.

Table 6.4: LRT test for the GH model and its sub-models

Model	Hypothesis	LRT	P-value
GLL-AH vs. BXII-AH	$H_0 : \eta = 1, H_1 : H_0$ is false,	6.999	0.008
GLL-AH vs. LL-AH	H_0: $\eta = \kappa, H_1 : H_0$ is false,	5.347	0.021
GLL-AH vs. W-AH	H_0: $\eta^{\alpha} \to 0$, $H_1 : H_0$ is false ,	14.533	<0.001

6.6.4 Bayesian Analysis

We used Bayesian analysis to compare the proposed GLL–AH model with its competing models, such as the GLL–PH, GLL–AH, and GLL–AFT models, and some of its submodels, including the LL–AH, BXII–AH, and W–AH regression models. The baseline distribution parameters $\alpha \sim G(a_1, b_1), \eta \sim G(a_2, b_2)$, and $k \sim G(a_3, b_3)$ with hyperparameter values $(a_1 = b_1 = a_2 = b_2 = a_3 = b_3 = 10)$ are assumed to have separate gamma priors that are independent and noninformative normal prior with a value of $N(0, 100)$ for $\beta's$ (regression coefficients). The Rstan package was utilized for our analysis (Carpenter et al., 2017).

6.6.4.1 Numerical Summary

In this section, we used the McMC sample of posterior properties for the proposed fully parametric AH, PO, AFT, and PH models with the GLL baseline distribution in Table 6.5 to examine several posterior characteristics of interest and their numerical values. The submodels of the GLL baseline distribution using the AH model are also examined in Table 6.6 to assess several posterior characteristics of interest and their numerical values.

Table 6.5: Results for the posterior properties of the GLL-AH, GLL-PO, GLL-PH and GLL-AFT models.

Models	Par (s)	Estimate	SE	SD	2.5%	Medium	97.5%	N_{eff}	$\hat{R}$
GLL-AH	β	1.016	0.009	0.476	0.030	1.027	1.909	2684	1.001
	α	0.836	0.002	0.106	0.648	0.829	1.064	3097	1.002
	κ	1.553	0.004	0.196	1.205	1.544	1.969	2714	1.001
	η	0.674	0.003	0.191	0.353	0.653	1.105	3023	1.001
GLL-PO	β	0.565	0.006	0.353	-0.135	0.562	1.268	3617	1.001
	α	1.414	0.003	0.156	1.136	1.405	1.741	3257	1.000
	κ	0.804	0.002	0.115	0.600	0.796	1.054	2951	1.001
	η	0.806	0.004	0.214	0.429	0.792	1.262	2918	1.000
GLL-PH	β	0.106	0.004	0.224	-0.330	0.107	0.540	3216	1.000
	α	1.341	0.002	0.146	1.077	1.332	1.646	3588	1.001
	κ	0.876	0.002	0.122	0.662	0.869	1.134	3068	1.001
	η	0.837	0.004	0.221	0.452	0.820	1.315	3239	1.001
GLL-AFT	β	0.418	0.005	0.269	-0.116	0.415	0.949	3396	1.000
	α	1.435	0.003	0.177	1.124	1.423	1.804	3373	1.000
	κ	0.809	0.002	0.114	0.609	0.801	1.060	2963	1.000
	η	0.850	0.004	0.210	0.479	0.836	1.311	2728	1.000

Table 6.6: Results for the posterior properties of the sub-models of the GLL-AH model including LL-AH, W-AH, and BXII-AH models.

Models	Par (s)	Estimate	SE	SD	2.5%	Medium	97.5%	N_{eff}	$\hat{R}$
LL-AH	β	0.764	0.007	0.385	-0.073	0.800	1.421	3228	1.001
	α	1.636	0.004	0.197	1.261	1.629	2.039	2930	1.000
	κ	0.879	0.002	0.107	0.688	0.873	1.109	3681	1.001
W-AH	β	-0.007	0.014	0.949	-1.850	-0.019	1.860	4377	1.000
	α	0.984	0.001	0.085	0.821	0.982	1.152	3521	1.000
	κ	0.559	0.001	0.068	0.437	0.554	0.702	3875	1.001
BXII-AH	β	0.678	0.007	0.378	-0.135	0.697	1.345	3291	1.000
	α	1.627	0.004	0.209	1.247	1.620	2.062	3099	1.000
	κ	0.949	0.002	0.115	0.740	0.943	1.186	3932	1.000

6.6.4.2 Visual Summary

Figures 6.3–6.9 provide the trace and autocorrelation (AC) plots for the baseline distribution parameters and regression coefficients of the proposed AH model and its submodels, plus other competing survival regression models, including the GLL–PH, GLL–PO, and GLL–AFT models, indicating convergence of the chains.

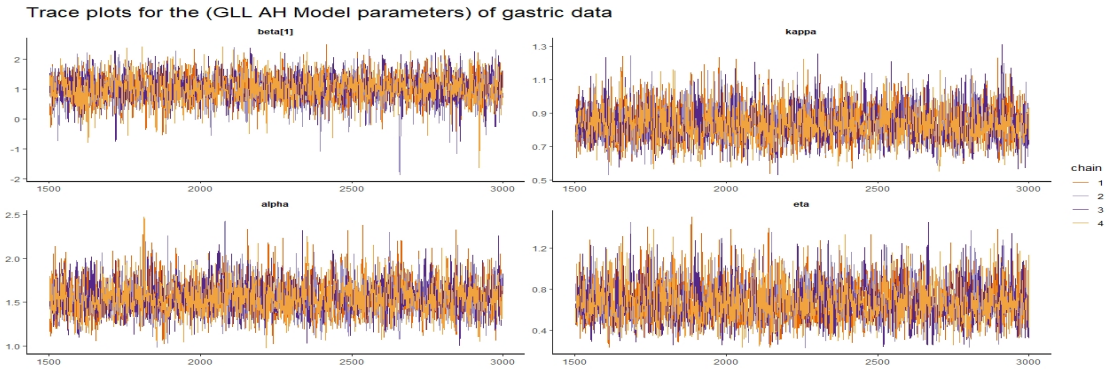

Figure 6.3: Trace plots for the GLL-AH model parameters

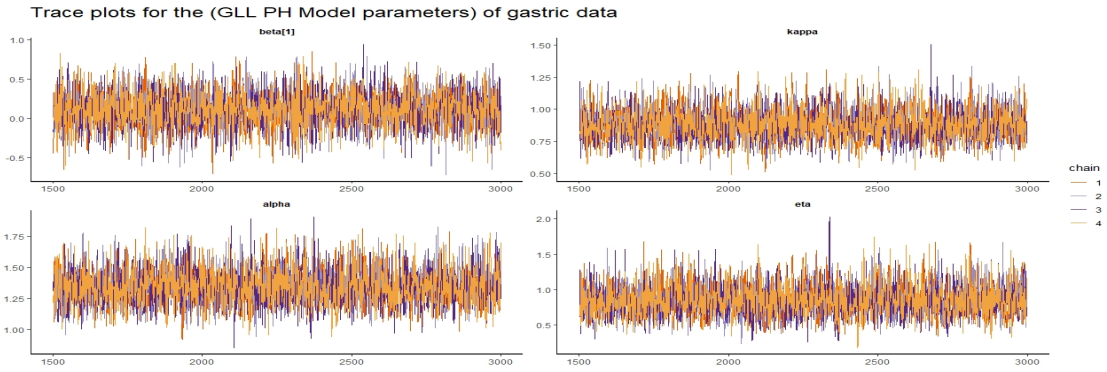

Figure 6.4: Trace plots for the GLL-PH model parameters

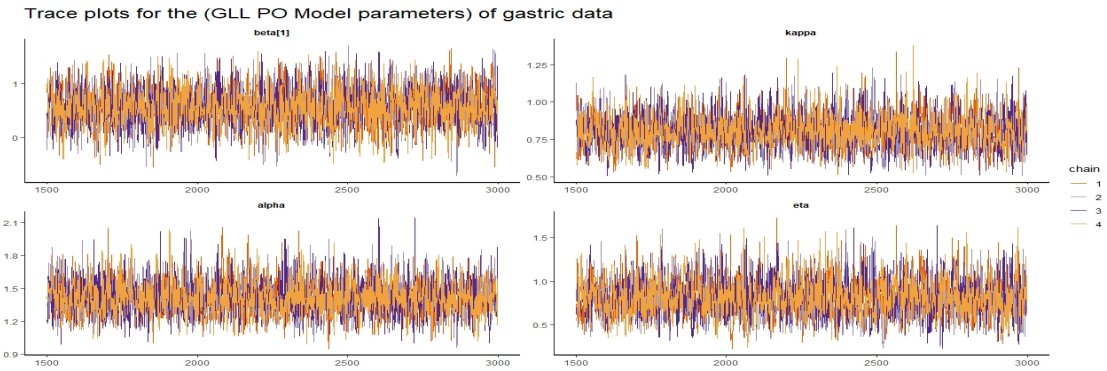

Figure 6.5: Trace plots for the GLL-PO model parameters

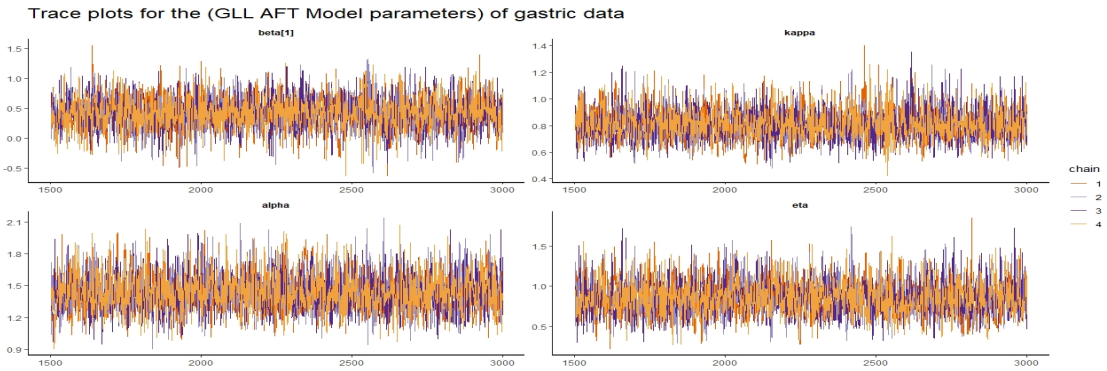

Figure 6.6: Trace plots for the GLL-AFT model parameters

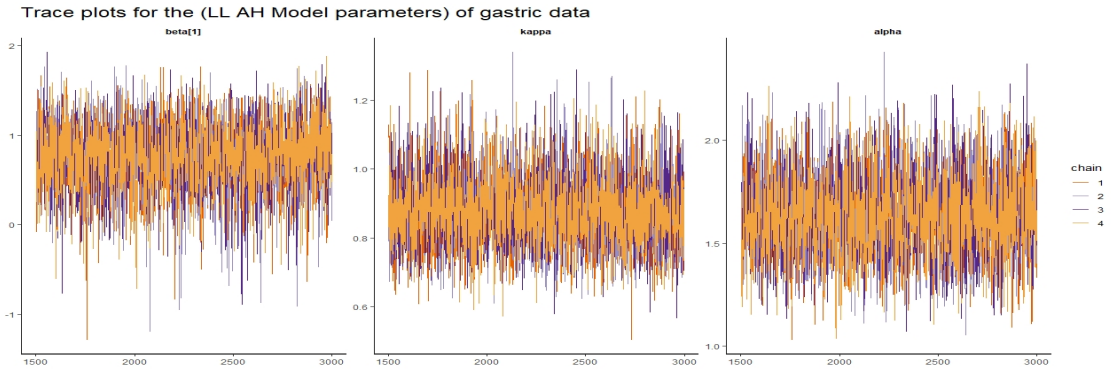

Figure 6.7: Trace plots for the LL-AH model parameters

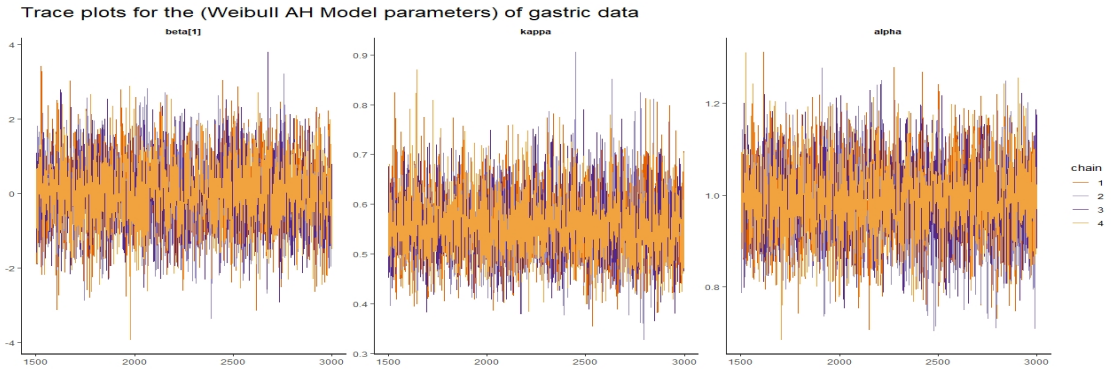

Figure 6.8: Trace plots for the W-AH model parameters

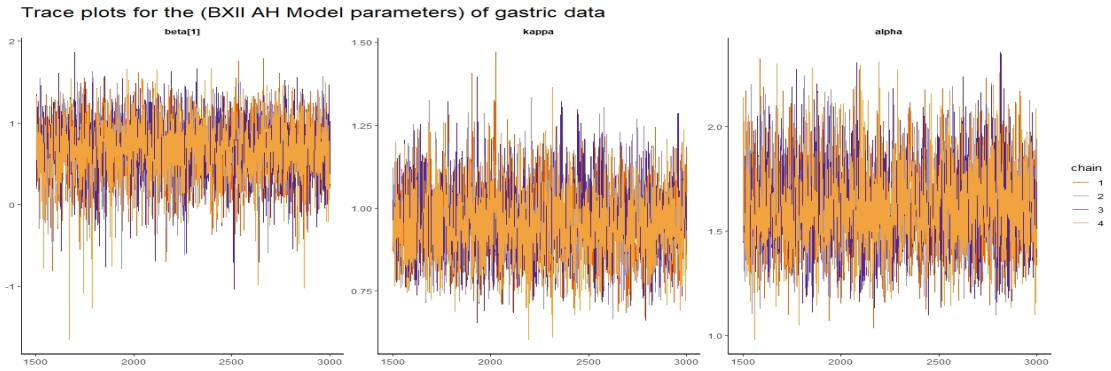

Figure 6.9: Trace plots for the BXII-AH model parameters

6.6.4.3 Posterior Predictive Checks

If a fitted Bayesian parametric hazard-based regression model predicts future observations that are consistent with the current data, it is considered sufficient or performing well. By using the Bayesplot R package, posterior predictive check (PPC)

plots are used to visually evaluate model fit. It can be seen from PPC in Figure 6.10, that the GLL–AH model fits the data quite well.

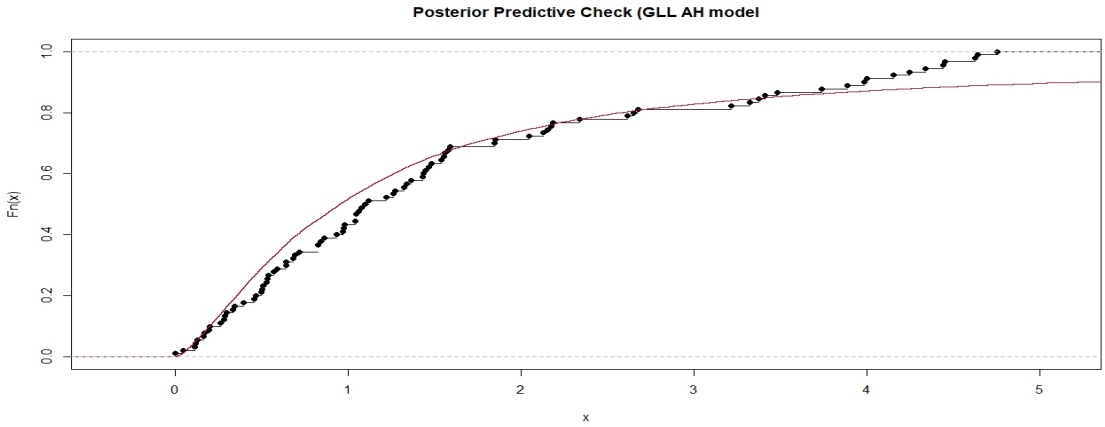

Figure 6.10: The empirical cdf, the dotted line and the cdf of the fitted model, the smooth curve, show that the fitted GLL-AH model predicts the future observations that are consistent with the current data.

6.6.4.4 McMC Convergence Diagnostics

We applied both numerical and visual methods to evaluate the convergence of the McMC algorithm for the proposed models and their special cases. The McMC algorithm HMC-NUTS has converged to the joint posterior distribution, as shown by the summary results in the above table, because the potential scale reduction factor $hatR$ is 1, the effective sample size (n_{eff}) is greater than 400, and the MC error (SE) is less than 0.05 of the posterior standard deviations for all parameters. Visually assessing convergence is often done by using AC and trace graphs (Ashraf-Ul-Alam and Khan, 2021). Figures 6.3–6.9 show a stationary pattern fluctuating within a band, demonstrating the convergence of the McMC algorithm. Figure 6.11, showing the AC plot, demonstrates how the AC rapidly decreases to zero as the period of lag increases, indicating good mixing and the convergence of the algorithm to the desired posterior distribution. Finally, Figure 6.12 indicates the pdf plots for the GLL-AH model posterior parameters.

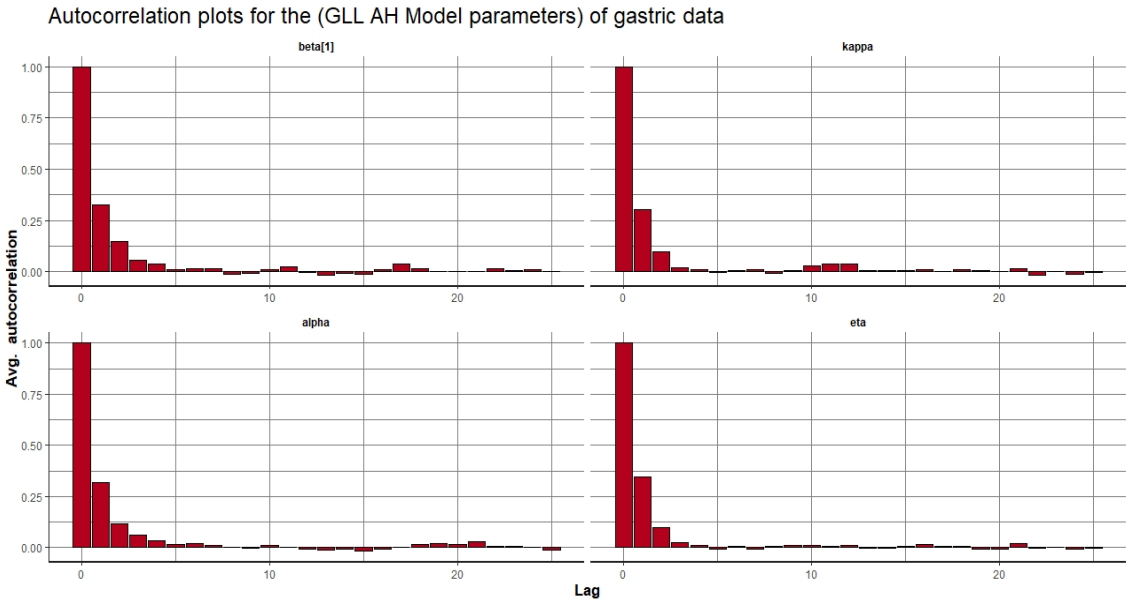

Figure 6.11: Auto-correlation plots for the GLL-AH model parameters

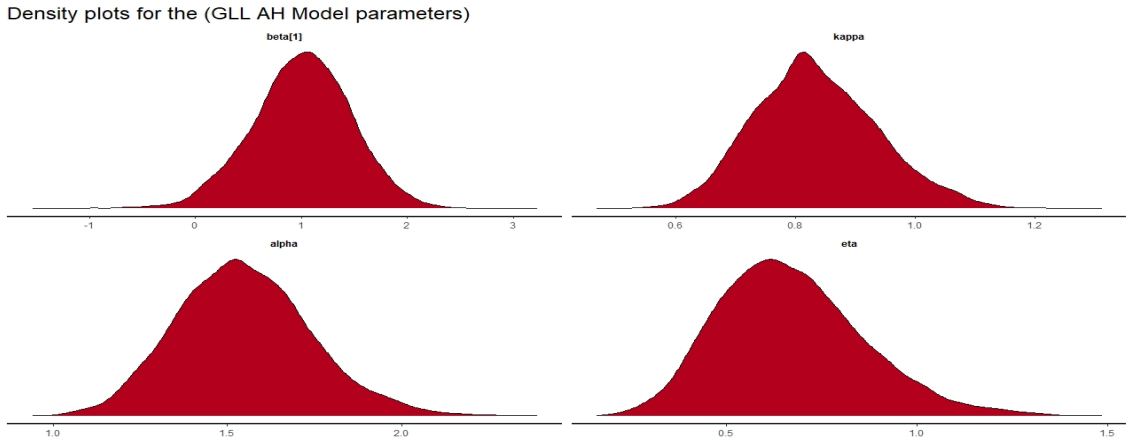

Figure 6.12: Density plots for the GLL-AH model parameters

6.6.4.5 Bayesian Model Selection

We implemented two information criteria, the Watanabe–Akaike information criterion (WAIC), proposed by Watanabe (2013), for the Bayesian model comparison, and the leave-one-out information criterion (LOOIC) proposed by Vehtari et al. (2017). A model may be said to be best suited if it has the lowest WAIC and LOOIC values for both information criteria. In addition to Stan fitting, posterior predictive check (PPC) and determining WAIC and LOOIC are performed by using

Table 6.7: Bayesian model comparison for the GLL-AH, GLL-PO, GLL-AFT, and GLL-PH models

Model	**WAIC**	**LOOIC**
GLL-AH	243.20	243.20
GLL-PO	251.40	251.42
GLL-AFT	251.80	251.90
GLL-PH	254.80	254.82

Table 6.8: Bayesian model comparison for the GLL-AH and its special cases including LL-AH, W-AH, and BXII-AH models

Model	**WAIC**	**LOOIC**
GLL-AH	243.20	243.20
LL-AH	249.30	249.40
W-AH	255.01	255.00
BXII-AH	247.05	247.08

the R package loo (Vehtari et al., 2017). Table 6.7 below shows that, when compared to its rival models, the GLL–AH model is the most effective. In addition, Table 6.8 demonstrates that, when compared to its sub-models, again the GLL–AH model is the superior one. Figure 6.13 indicates the Kaplan-Meier estimate and the survival estimate curve for the proposed GLL-AH model parameters.

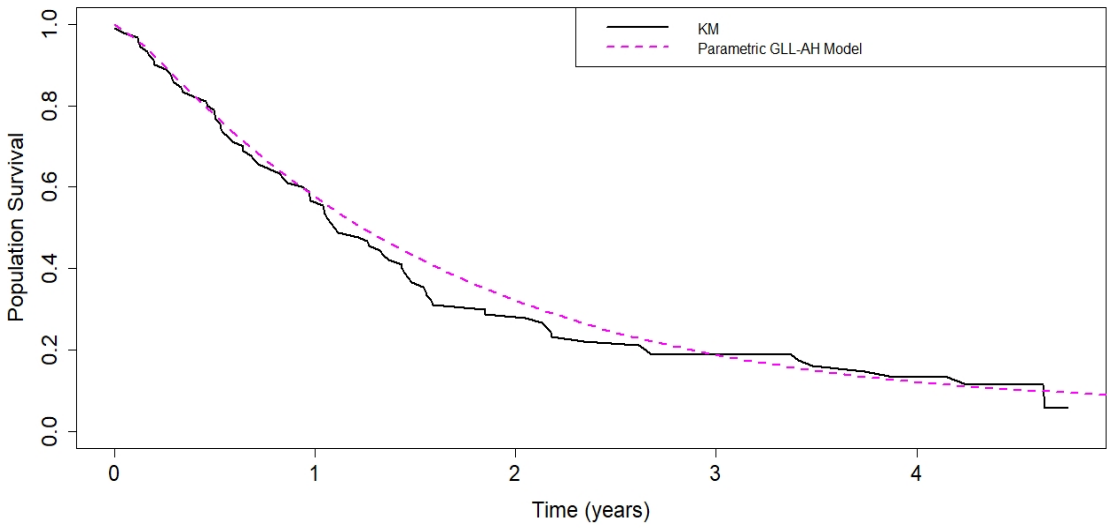

Figure 6.13: Kaplan-Meier and fitted survival curve for the GLL-AH model of the gastric cancer data set.

Figure 6.13 indicates the Kaplan–Meier estimate and the sf estimate for the proposed GLL–AH model parameters.

Figures 6.14 and 6.15 demonstrate the Kaplan–Meier estimate and the survival estimate curves for the proposed regression models with GLL baseline distribution and the AH model with various baseline hazards. In Figure 6.14, the GLL–AH model survival curve is closer to the KM survival curve compared to all other survival regression models. The same thing occurred in Figure 6.15.

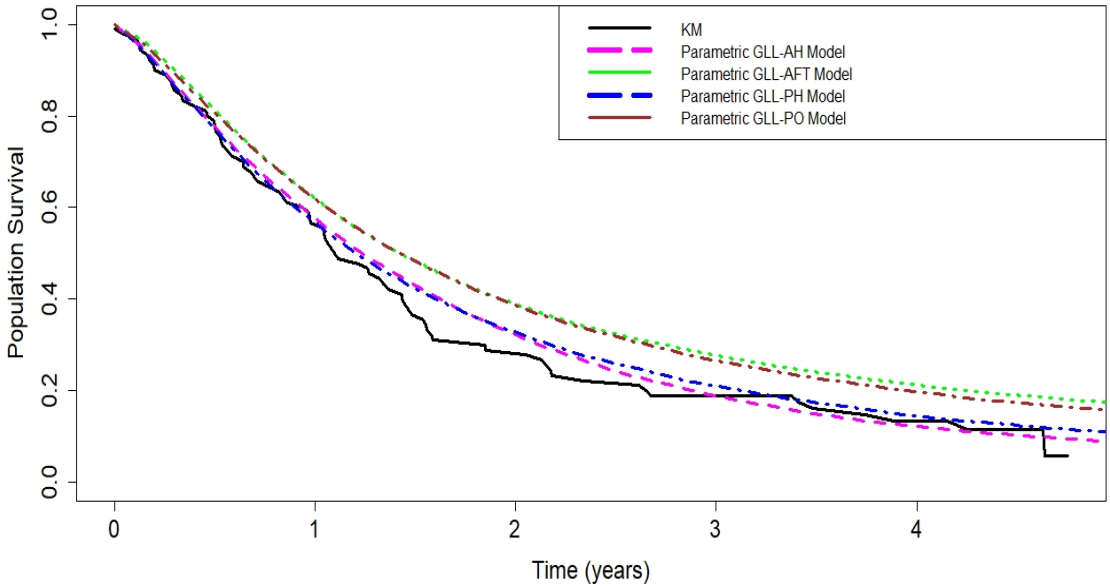

Figure 6.14: Kaplan-Meier and Estimated Survival Plots for the competitive Regression models with GLL baseline of the gastric cancer data set.

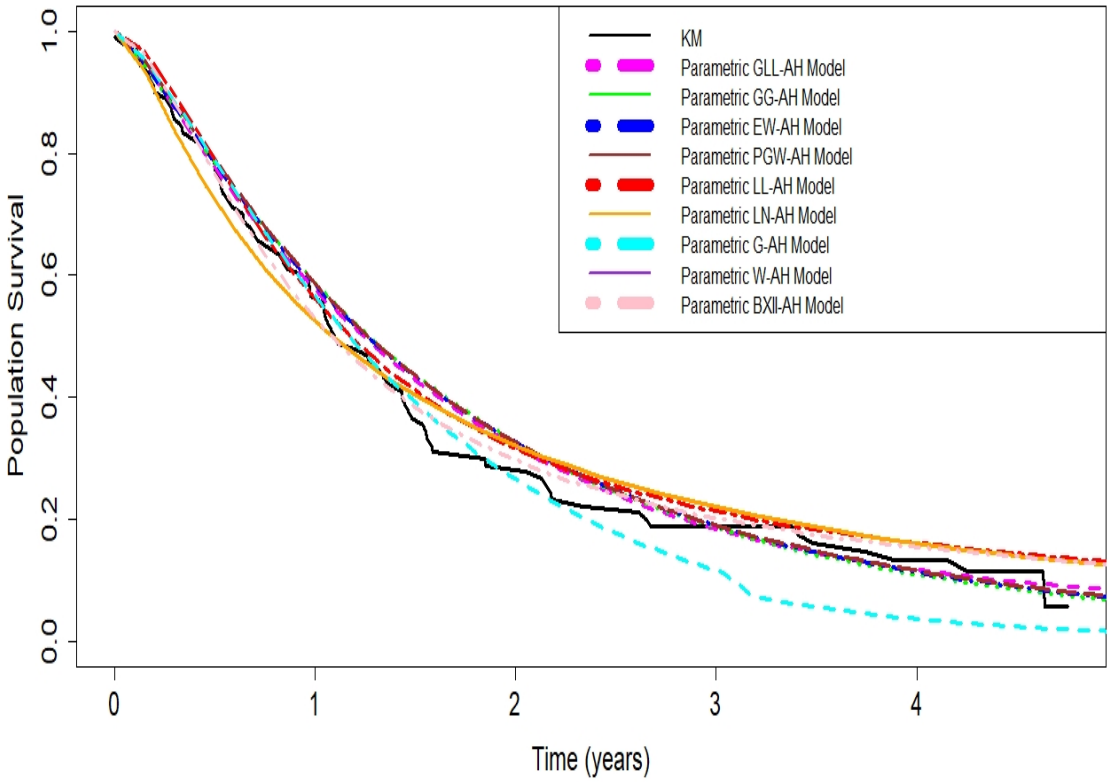

Figure 6.15: Kaplan-Meier and Estimated Survival Plots for the competitive AH models of the gastric cancer data set.

6.7 Conclusions

This chapter proposes a fully parametric AH model for dealing with censored lifetime data with crossover survival curves as an extension of the semi-parametric AH model (Chen and Wang, 2000). The primary distinction between this modification and others is that we used a modified baseline distribution that can capture different hazard rate shapes to provide a more flexible depiction of the baseline hazard. By adopting a flexible parametric baseline distribution like the GLL distribution, we showed that it is possible to carry out both Bayesian and classical likelihood inference using the rstan package of the R programming language.

This also defines the chapter's key contribution, as no other study combining these

two characteristics (AH model and a modified baseline distribution) can be found in the time-to-event analysis field. Furthermore, employing both Bayesian and classical inference via MLE will address the semi-parametric AH model's limited use due to a lack of efficient and trustworthy estimation methods. Additionally, using the GLL distribution as a baseline hazard offers several benefits as compared to other parametric baseline distributions that may accept different hazard rate shapes, such as the gamma, GG, Weibull, EW, PGW, LL, Bur-XII, and LN distributions.

Following the simulation study, the chapter gave a real-world demonstration involving a well-known dataset with crossover survival curves and was concerned with a clinical study for patients with gastric cancer. In summary, the GLL–AH model outperforms the other competing parametric AH models with various baseline hazards and other survival regression models with the same baseline hazard. Finally, we developed an R package, "AHSurv", to fit the proposed model in this study as an addendum to this paper; the source code is accessible at (Muse et al., 2022e).

CHAPTER 7

Bayesian and Frequentist Approaches for a Tractable Parametric General Class of Hazard-Based Regression Models: An Application to Oncology Data

In this chapter, we present our sixth manuscript [1] about a tractable parametric general hazard model. Note that the materials of this chapter have been reproduced from our article which is under review.

7.1 Introduction

One of the main goals of censored time-to-event data analysis with covariates is to find and quantify the relationship between the baseline hazard rate function and the covariates so that the covariates can be employed in disease prevention and management (Zhou and Hanson, 2015; Alvares et al., 2021; Rubio et al., 2019; Demarqui and Mayrink, 2021; Rubio et al., 2021). The assumption leads to hazard-based regression models, and the study' goal is to estimate a vector of regression coefficients for the components of covariates.

Cox Cox (1972) developed a hazard-based regression model in which the covariate has a multiplicative relationship with the hazard rate function, and is called the proportional hazard (PH) model. The PH is without a doubt the most widely used in practice. Let $h(t;x)$ be the hazard rate function for a subject with $x =$

[1] Muse, A. H., Mwalili, S., Ngesa, O., Chesneau, C., Al-Bossly, A., & El-Morshedy, M. (2022). Bayesian and Frequentist Approaches for a Tractable Parametric General Class of Hazard-Based Regression Models: An Application to Oncology Data. Mathematics, 10(20), 3813.

$(x_1, x_2, \ldots, x_p)^T$ variables, and $h_0(t)$ be the baseline hazard rate function for those with $x = 0$. The following is a formula for the Cox PH model:

$$h(t; x) = h_0(t) \cdot \psi(\boldsymbol{\beta} \cdot x) = h_0(t) e^{x'\beta}. \tag{7.1}$$

where ψ is a positive link function with $\psi(0) = 1$, most often the exponential function is used to represent the link function of the covariates; $\beta = (\beta_1, \beta_2, \ldots, \beta_p)^T$ is the vector of the regression coefficients, e^{β_j} denotes the hazard ratio resulting from an increase in the jth covariate by one unit. The Cox PH model leads to the estimation of β by means of a "partial likelihood" approach Lawless (2011).

The Cox PH model is usually used to model censored lifetime data. However, there may be some benefits to using parametric PH models for such data. According to Hjort Hjort (1992), the success of the Cox PH regression model may have had the unintended consequence of practitioners paying too little attention to the baseline hazard. If proved to be acceptable, a parametric version of the Cox model would allow for more exact estimation of survival probability while also contributing to a better understanding of the phenomenon under investigation. Parametric PH models, for example, which might be a challenge with the Cox PH model, can sometimes be handled simply, and visualizations of the hazard rate function are considerably easier. To accommodate variable hazard rate forms, modifications to the log-logistic and Weibull models are presented. For example, exponentiated-Weibull Mudholkar and Hutson (1996), sine Kumaraswamy-Weibull Chesneau and Jamal (2020), arctan-Weibull Alkhairy et al. (2021), exponentiated generalized cosine-Weibull Mahmood et al. (2022), secant kumaraswamy-Weibull Souza et al. (2022), tan log-logistic Muse et al. (2021c) and the generalized log-logistic Muse et al. (2021a) models.

When the Cox PH model proportionality assumption is not satisfied, flexible parametric non-proportional hazards can be relaxed. For example, the accelerated failure time (AFT) model can be considered Kalbfleisch and Prentice (2011). The AFT

model can be written as:

$$h(t;x) = h_0\left(te^{x'\beta}\right)e^{x'\beta}. \tag{7.2}$$

The AFT assumption in Equation 7.2 postulates that the covariates have time-dependent and non-proportional effects on the hazard rate, while PH assumption in Equation 7.1 postulates that the covariates have time-independent and proportional effects.

The PH and AFT models have been widely employed in a variety of time-to-event analysis applications. These models, despite their popularity, are unsuitable for handling time-to-event data with crossed survival and hazard curves Demarqui et al. (2019). Chen and Weng Chen and Wang (2000) presented a new class of hazard-based regression models, known as the accelerated hazard (AH) model, that may be used to analyze crossing survival curves. The AH model can be written as:

$$h(t;x) = h_0\left(te^{x'\beta}\right). \tag{7.3}$$

To describe the shift in hazard progression across time, assumption 7.3 assumes that the covariates have a time-scale change to the hazard rate function. The AH model has the advantage of being non-proportional, it can accept the phenomenon of identical hazards at time $t = 0$, which is common in randomized clinical trials.

As a result, the relationship between covariates and baseline hazard can be described as follows: covariates with a "proportional" effect on the hazard rate function must be time-independent and able to scale up the baseline hazard function, whereas covariates with a "non-proportional" effect on the hazard rate function must be time-dependent or interact with time in the baseline hazard function. As a result, whereas the influence of a time-independent covariate on the hazard rate varies over time, the effect of a time-independent covariate on the hazard rate function remains constant.

Hence, when creating the link between the covariates and the hazard rate function, we have four options: time-dependent or time-independent, proportional or non-proportional. As a result, neither the AFT, AH, or PH models can enable some factors to have proportional and time-dependent impacts on the hazard while others have non-proportional and time-dependent effects in one model Wang et al. (2018a). To solve this issue, the goal of this chapter is to introduce a tractable parametric general class of hazard-based regression models for overall survival data that includes the AFT, AH, and PH as special cases, which will subsequently be used to model right-censored cancer data sets with or without crossover survival curves.

The motivating ideas behind our work on Bayesian and frequentist approaches for the general class of hazard-based regression models with GLL baseline distribution are as follows: (i) The baseline continuous probability distributions closed under the PH and AFT frameworks have drawbacks in that most of them are not adaptable enough to take into account both monotone and non-monotone hazard rates; (ii) For statistical inference, the Bayesian approach does not depend on asymptotic approximations; and to the author's knowledge, there are no previous studies for the Bayesian inference of the general class of parametric hazard-based regression models; (iii) due to the accessibility of software, Bayesian application for hazard-based complex models is considerably easier and simpler than the frequentist approach; (iv) if the baseline hazard distribution is valid and correct, parametric hazard-based regression models may yield more accurate estimates than semi-parametric hazard-based regression models; and, last but just not least, (v) what sets our work apart and appeals to healthcare professionals, epidemiologists, bio-statisticians, and other applied researchers in numerous fields is the use of modified distributions that may accommodate different hazard rate shapes data.

Based on the above-mentioned motivations and discussions, the main purpose of this paper is to introduce a general parametric hazard-based regression model with a generalized log-logistic baseline distribution. So, presenting the parametric general

hazard (GH) class of hazard-based regression models and their special cases using GLL baseline hazard, deriving the probabilistic functions for the GH model, formulating and interpreting all the special cases, applying the Bayesian and frequentist inference procedures, developing computational algorithms to fit the proposed GH model, estimating the covariate effect on the hazard rate, and applying it to a right-censored cancer data set is the novelty of the study.

The rest of this chapter is folded as follows: Section 7.2 describes the proposed general hazard (GH) model formulation, assumptions, its nested structure, and its probabilistic functions. Section 7.3 lists the special cases of the general hazard model and their probabilistic functions. The parameter interpretations of the sub-models are discussed in Section 7.4. Section 7.5 presents the proposed general hazard model with generalized log-logistic baseline distribution. Section 7.6 presents the inferential approaches of the proposed GH model. Two extensive simulation studies are presented in Section 7.7. Section 7.8 displays two right-censored cancer data sets, one of which contains crossover survival curves. The final Section 7.9 contains the major conclusion, final remarks, and discussion of future work.

7.2 Model Formulation

7.2.1 Review of Current Literature and Recent Research

Prior to the model formulation, we discuss the state of scientific progress in the context of current survival models. Specifically, we look at the work that has been done in relation to the closely related extended hazard (EH) and generalized hazard (GH) models. The EH model is actually very similar to the GH model; the only distinction is that the EH was developed before the development of the AH model. In the study of censored lifetime data with covariates, Ciampi and Etezadi-Amoli (1985) constructed a universal model for evaluating the PH and the AFT hypothesis.

Then, using spline approximation, Etezadi-Amoli and Ciampi (1987) proposed an EH model for censored lifetime data with variables.

Following the research done by Etazadi-Amoli and Ciampi, Louzada-Neto (1997) proposed an EH regression model that permits the spread parameter to rely on covariates. For EH models, Louzada-Neto (2001) presented a simple Bayesian analysis. Then, after the development of the AH model, Chen and Jewell (2001) developed the GH model, which combines the EH model with the AH model, another hazard-based model. An extended linear EH model with applicability to time-dependent covariates was proposed by (Elsayed et al., 2006). Tong et al. (2013) addressed a few inferential research questions in the semi-parametric GH model. The modification of the GH model was discussed by (Wang et al., 2018a) in the context of time-independent and time-dependent factors.

The majority of the work developed in relation to GH models dealt with semi-parametric models. The parametric GH regression models for the relative survival data were recently examined by (Rubio et al., 2019). Other efforts, for the GH models were discussed by Li et al. (2015) by extending the GH to a spatial model. Finally, a mixed-effect GH model was created by Rubio Alvarez and Drikvandi (2022) to model for clustered survival data.

There is a research gap following the current literature that needs to be filled. Both the application of the GH model to the overall survival data and the field of statistical inference for the Bayesian approach are unresolved issues that need to be resolved. Here, we suggest a parametric GH model to close that gap, and we estimate the model's parameters using both maximum likelihood estimation (MLE) and Bayesian methods.

7.2.2 Model Formulation

The hazard rate function (hrf) and the cumulative hazard function (chf), rather than the probability density function (pdf) and the cumulative distribution function (cdf) are typically used to interpret the most common parametric hazard-based regression models.

Assume that x is a vector of covariates, T is a non-negative random variable that represents the amount of time till the occurrence of an event of concern, and $\psi(x)$ is the link function for the covariates, which is most often employed as an exponential or (log-linear function) $e^{x^{'}\beta}$, where β is a vector of regression coefficients. Ciampi and Etezadi-Amoli (1985) developed a generalized version of the PH, and AFT models to incorporate more versatile interaction terms in terms of the covariates and time. The hrf and chf of the general class of hazard-based regression models are expressed as follows:

$$h_{GH}(t;x) = h_0\left(t\psi\left(\boldsymbol{x}'\boldsymbol{\beta_1}\right)\right)\psi\left(\boldsymbol{x}'\boldsymbol{\beta}_2\right) = h_0\left(te^{x'\beta_1}\right)e^{x'\beta_2}. \tag{7.4}$$

$$H_{GH}(t;x) = H_0\left(t\psi\left(\boldsymbol{x}'\boldsymbol{\beta_1}\right)\right)\psi\left(-\boldsymbol{x}'\boldsymbol{\beta_1} + \boldsymbol{x}'\boldsymbol{\beta}_2\right) = H_0\left(te^{x'\beta_1}\right)e^{x'\beta_2 - x'\beta_1}. \tag{7.5}$$

where $h_0(t)$ and $H_0(t)$ are called the baseline hazard and the baseline cumulative hazard rates respectively; $h_{GH}(t;x)$ is the hazard rate function at time t, and $H_{GH}(t;x)$ is the cumulative hazard rate function at time t.

7.2.3 Nested Structure of the GH Model

The importance of the general class of parametric hazard-based regression model is that it represents a GH structure that contains, as special cases, the proportional hazard (PH), accelerated hazard (AH), and the accelerated failure time (AFT) models. To be more precise,

i. If $\beta_1 = 0$, then GH = PH;

ii. If $\beta_2 = 0$, then GH = AH; and

iii. If $\beta_1 = \beta_2$, then GH = AFT.

Hence, the GH model can be used as a numerical tool to determine which of them is more appropriate for a given censored survival data. The nested structure of the model illustrates the richness of the GH class and motivates its investigation Etezadi-Amoli and Ciampi (1987).

7.2.4 Model Assumption

The basic assumption of the general class of parametric hazard-based regression models is that the effect of covariates on the hazard rate function is identified as having two separate components, namely:

i. A time-scale change in the hazard rate function, and

ii. A relative hazards ratio

$$h_{GH}(t;x) = h_0\left(te^{x'\beta_1}\right)e^{x'\beta_2}$$

In other words, the model has both

i. time-dependent and time-independent (time-fixed) covariates

ii. proportional and non-proportional hazards separately,

for evaluating the hazard function and hazard ratio over time Chen and Jewell (2001).

The assumption of the special cases is different in nature. For instance, the PH framework postulates that the covariates multiply the hazard rate function, causing the hazard function to fluctuate in level (Rezaei et al., 2014). The AH framework

postulates that that each covariate has a time-dependent effect since it states that the effect of a unit change in a covariate affects the time scale of the baseline hazard rate (Co, 2010). In the AFT framework, it is postulated that the covariates have an effect both on the hazard and the time scale (Khan, 2018). Note that, the AH, AFT, and PH models coincide for the case when the baseline hazard is the Weibull distribution (Rubio et al., 2019).

7.2.5 Probabilistic Functions for the GH Model

In this section, we derive the most common probabilistic functions for the GH model. The other probabilistic functions for the model with equations 7.4 and 7.5 are computed as follows:

The survival function (sf) of the GH model is computed as follows:

$$S_{GH}(t;x) = S_0\left(te^{x'\beta_1}\right)^{e^{x'\beta_2 - x'\beta_1}}. \tag{7.6}$$

where $S_0(.)$ is the baseline survival function.

The cdf of the GH model is expressed as follows:

$$F_{GH}(t;x) = 1 - S_{GH}(t;x) = 1 - S_0\left(te^{x'\beta_1}\right)^{e^{x'\beta_2 - x'\beta_1}}. \tag{7.7}$$

The pdf of the GH model can be obtained by using:

$$f_{GH}(t;x) = h_0\left(te^{x'\beta_1}\right)e^{x'\beta_2}S_0\left(te^{x'\beta_1}\right)^{e^{x'\beta_2 - x'\beta_1}}. \tag{7.8}$$

7.3 Special Cases of the GH Model

In this section, the three common sub-models for the general class of the hazard-based regression models are discussed.

7.3.1 Proportional Hazard Model

In the GH model framework, if $\beta_1 = 0$, then GH = PH framework. Hence, the hrf is expressed as follows:

$$h_{PH}(t;x) = h_0(t)e^{x'\beta}. \tag{7.9}$$

The chf is expressed as:

$$H_{PH}(t;x) = H_0(t)e^{x'\beta}. \tag{7.10}$$

The sf of the PH model is computes as follows:

$$S_{PH}(t;x) = S_0(t)^{e^{x'\beta}}. \tag{7.11}$$

The cdf of the PH model is expressed as follows:

$$F_{PH}(t;x) = 1 - S_{PH}(t;x) = 1 - S_0(t)^{e^{x'\beta}}. \tag{7.12}$$

The pdf of the PH model can be obtained by using:

$$f_{PH}(t;x) = f_0(t)e^{x'\beta}S_0(t)^{e^{x'\beta}-1} \tag{7.13}$$

7.3.2 Accelerated Hazard Model

In the GH model framework, if $\beta_2 = 0$, then GH = AH framework. Hence, the hrf is expressed as follows:

$$h_{AH}(t;x) = h_0\left(te^{x'\beta}\right). \tag{7.14}$$

The chf is expressed as:

$$H_{AH}(t;x) = H_0\left(te^{x'\beta}\right)e^{-x'\beta}. \tag{7.15}$$

The sf of the AH model is computed as follows:

$$S_{AH}(t;x) = S_0\left(te^{x'\beta}\right)^{e^{-x'\beta}}. \tag{7.16}$$

The cdf of the AH model is expressed as follows:

$$F_{AH}(t;x) = 1 - S_{AH}(t;x) = 1 - S_0\left(te^{x'\beta}\right)^{e^{-x'\beta}}. \tag{7.17}$$

The pdf of the AH model can be obtained by using:

$$f_{AH}(t;x) = f_0\left(te^{x'\beta}\right)S_0\left(te^{x'\beta}\right)^{e^{-x'\beta}}. \tag{7.18}$$

7.3.3 Accelerated Failure Time Model

In the GH model framework, if $\beta_1 = \beta_2$, then GH = AFT framework. Hence, the hrf is expressed as follows:

$$h_{AFT}(t;x) = h_0\left(te^{x'\beta}\right)e^{x'\beta}. \tag{7.19}$$

The chf is expressed as:

$$H_{AFT}(t;x) = H_0\left(te^{x'\beta}\right). \tag{7.20}$$

The sf of the AFT model is computed as follows:

$$S_{AFT}(t;x) = S_0\left(te^{x'\beta}\right) \tag{7.21}$$

The cdf of the AFT model is expressed as follows:

$$F_{AFT}(t;x) = 1 - S_{AFT}(t;x) = 1 - S_0\left(te^{x'\beta}\right). \tag{7.22}$$

The pdf of the AFT model can be obtained by using:

$$f_{AFT}(t;x) = f_0\left(te^{x'\beta}\right)e^{x'\beta}. \tag{7.23}$$

The GH model is a general class that includes the PH, AFT, and AH models as special examples. More specifically, if $\beta_1 = 0$, then GH = PH; if $\beta_2 = 0$, then GH = AH; and if $\beta_2 = \beta_1$, then GH=AFT (Rubio et al., 2019; Etezadi-Amoli and Ciampi, 1987; Ciampi and Etezadi-Amoli, 1985).

7.4 Parameter Interpretation for the Parametric Hazard-based Regression Models

In this section, it is crucial to determine if a positive coefficient of a covariate can monotonically raise the time scale or lower the hazard in order to ensure that the broad class of hazard-based regression models has a plausible interpretation.

7.4.1 Proportional Hazard Model

In this sub-section, we start the interpretation of the parameters of the PH model, as this will facilitate the interpretation of the general class model. if $\beta_1 = 0$, then the GH model is the same as that in a PH model:

$$h(t;x) = h_0(t)e^{x'\beta} \tag{7.24}$$

$$H(t;x) = H_0(t)e^{x'\beta} \tag{7.25}$$

$$\frac{dH(t;x)}{dx} = \beta e^{x'\beta}H_0(t) \tag{7.26}$$

where $H_0(t)$ is an increasing function that takes a value greater than 0 . When $\beta > 0$, $\frac{dH(t;x)}{dx} > 0$, showing that when the covariate x increases, the hazard increases and

the survival time decreases.

7.4.2 Accelerated Failure Time Model

In the GH model, if the covariate $\beta_2 = \beta_1$, the GH model is the same as the AFT model. Hence,

$$h(t;x) = h_0\left(te^{x'\beta}\right)e^{x'\beta} \tag{7.27}$$

$$H(t;x) = H_0\left(te^{x'\beta}\right) \tag{7.28}$$

$$\frac{dH(t;x)}{dx} = \beta te^{x'\beta}H_0'\left(te^{x'\beta}\right) \tag{7.29}$$

Since $H_0(.)$ is an increasing function, $H_0'(.)$ Is greater than 0 . When $\beta > 0$, $\frac{dH(t;x)}{dx} > 0$, showing that when the covariate x increases, the hazard increases and the survival time decreases.

7.4.3 Accelerated Hazard Model

In the GH model, if the covariate $\beta_2 = 0$, the GH model is the same as the AH model. Hence,

$$h(t;x) = h_0\left(te^{x'\beta}\right) \tag{7.30}$$

$$H(t;x) = e^{-x'\beta}H_0\left(te^{x'\beta}\right) \tag{7.31}$$

$$\frac{dH(t;x)}{dx} = -\beta e^{-x'\beta}\left[H_0\left(te^{x'\beta}\right) - te^{x'\beta}H_0'\left(te^{x'\beta}\right)\right] \tag{7.32}$$

No matter whether the β is less than or greater than zero, when $t = \frac{H_0\left(te^{x'\beta}\right)}{e^{x'\beta}H_0'\left(te^{x'\beta}\right)'}$, $\frac{dH(t;x)}{dx} = 0$. Thus, with the increase of t, the sign of $\frac{dH(t;x)}{dx}$ may change. The time when the sign changes depend on the form $H_0(.)$ Function. It means that x with a positive coefficient does not always increase the duration when it increases, which results in a challenge to interpret the sign of a variable coefficient.

Hence, the AH model also has the merit of being applicable to the crossing of hazards

Table 7.1: Summary of Parameter interpretation and comparison of PH, AH, and AFT Models

	AH Model	PH Model	AFT Model
effect	$\beta_{AH} > 0$ treatment accelerates the hazard by a factor of $e^{\beta_{AH}}$	$\beta_{PH} > 0$ treatment proportionally increases hazard by a factor of $e^{\beta_{PH}}$	$\beta_{AFT} > 0$ treatment decelerates failure time of the survivor function by a factor of $e^{\beta_{AFT}}$
	$\beta_{AH} < 0$ treatment decelerates the hazard by a factor of $e^{\beta_{AH}}$	$\beta_{PH} < 0$ treatment proportionally decreases hazard by a factor of $e^{\beta_{PH}}$	$\beta_{AFT} < 0$ treatment accelerates failure time of the survivor function by a factor of $e^{\beta_{AFT}}$
β's interpretation	Hazard progression time ratio	hazard ratio	Survival time ratio
limited to crossover in hazards	No	Yes	Yes
limited to crossover in survival	No	Yes	No

and survivor functions. In other circumstances, this benefit may make it difficult to comprehend the parameters accurately (Co, 2010; Qing Chen, 2001). As a result, rather than providing merely the survival functions, showing the hazards according to distinct covariate patterns is recommended to assist in illustrating the survival time process (Chen and Jewell, 2001). In fact, GH and AH are suitable for the analysis of crossover survival and hazard functions (Zhang and Peng, 2009).

In general, the parameter interpretation depends on the shape of the baseline hazard, which we classify here as monotone (decreasing or increasing) or non-monotone (bathtub or unimodal). A summary of parameter interpretation for the AFT, AH, and PH models is presented in Table 7.1 below. **N.B.** $\beta_{AH} = \beta_{PH} = \beta_{AFT} = 0$, treatment does not have an effect

7.5 Generalized Log-logistic General Hazard Model

For the GH model, the generalized log-logistic baseline hazard is

$$h(t) = \frac{\alpha k (kt)^{\alpha-1}}{\lceil 1 + (\eta t)^{\alpha} \rceil},$$

so, according to Equation (7.4) the hazard rate for an individual with covariate vector x and link function $\psi(x)$ is

$$h_{GLL-GH}\left(t; \boldsymbol{\theta}, \beta_1, \beta_2\right) = h_0\left(t\psi\left(\boldsymbol{x}'\boldsymbol{\beta_1}\right)\right)\psi\left(\boldsymbol{x}'\boldsymbol{\beta}_2\right) = \frac{\alpha k \left(kt\psi\left(\boldsymbol{x}'\boldsymbol{\beta_1}\right)\right)^{\alpha-1}}{\lceil 1 + \left(\eta t\psi\left(\boldsymbol{x}'\boldsymbol{\beta_1}\right)\right)^{\alpha} \rceil}\psi\left(\boldsymbol{x}'\boldsymbol{\beta}_2\right) \tag{7.33}$$

applying the log-linear function $\psi\left(\boldsymbol{x}'\boldsymbol{\beta}\right) = \exp\left(\boldsymbol{x}'\boldsymbol{\beta}\right)$, we can simplify into

$$h_{GLL-GH}\left(t; \boldsymbol{\theta}, \beta_1, \beta_2\right) = h_0\left(t\psi\left(\boldsymbol{x}'\boldsymbol{\beta}_1\right)\right)\psi\left(\boldsymbol{x}'\boldsymbol{\beta}_2\right) = \frac{\alpha k \left(kte^{x'\beta_1}\right)^{\alpha-1}}{\left[1 + \left(\eta te^{x'\beta_1}\right)^{\alpha}\right]}e^{x'\beta_2}, \tag{7.34}$$

The chf of the GLL-GH model using Equation (7.5) is obtained as follows:

$$H_{GLL-GH}\left(t; \boldsymbol{\theta}, \beta_1, \beta_2\right) = e^{x'\beta_2 - x'\beta_1}\frac{k^{\alpha}}{\lambda^{\alpha}}\log\left[1 + \left(\lambda te^{x'\beta_1}\right)^{\alpha}\right], \tag{7.35}$$

7.6 Model Inference

In this section, we discuss the classical approach (via maximum likelihood estimation technique) and Bayesian inference (assuming non-informative priors for both baseline hazard parameters and regression coefficients) for the general class of hazard-based regression models with generalized log-logistic baseline.

7.6.1 Classical Inference

The general class of hazard-based regression models is considered in this sub-section with a fully parametric specification. To obtain the frequentist inference about the vector of model parameters, we assume that the time-to-event data is right-censored and that the censoring mechanism is non-informative. The censored likelihood function can be defined as follows when a parametric general class of hazard-based regression model is considered:

$$\begin{aligned} L\left(\boldsymbol{\theta}, \beta_1, \beta_2 \mid D\right) &= \prod_{i=1}^{n} \left[f\left(t_i; \boldsymbol{\theta}, \beta_1, \beta_2, x\right)\right]^{\delta_i} \left[S\left(t_i; \boldsymbol{\theta}, \beta_1, \beta_2, x\right)\right]^{1-\delta_i} \\ &= \prod_{i=1}^{n} \left[\frac{h\left(t_i; \boldsymbol{\theta}, \beta_1, \beta_2, x\right)}{S\left(t_i; \boldsymbol{\theta}, \beta_1, \beta_2, x\right)}\right]^{\delta_i} \left[S\left(t_i; \boldsymbol{\theta}, \beta_1, \beta_2, x\right)\right]^{1-\delta_i} \\ &= \prod_{i=1}^{n} \left[h\left(t_i; \boldsymbol{\theta}, \beta_1, \beta_2, x\right)\right]^{\delta_i} S\left(t_i; \boldsymbol{\theta}, \beta_1, \beta_2, x\right) \\ &= \prod_{i=1}^{n} \left[h\left(t_i; \boldsymbol{\theta}, \beta_1, \beta_2, x\right)\right]^{\delta_i} \exp\left[-H\left(t_i; \boldsymbol{\theta}, \beta_1, \beta_2, x\right)\right], \\ &= \prod_{i=1}^{n} \left[h_0\left(t_i e^{{x_i}'\beta_1}, \boldsymbol{\theta}\right) e^{x_i\beta_2}\right]^{\delta_i} \exp\left[-H_0\left(t_i e^{x_i'\beta_1}, \boldsymbol{\theta}\right) e^{x_i'\beta_2 - x_i'\beta_1}\right], \end{aligned} \tag{7.36}$$

where θ is the vector of distributional parameters with the baseline hazard, and $D = (t_i, \delta_i, x_i, i = 1, 2, \ldots, n)$ denotes the observed data including t_i = survival time, δ_i = censoring time, and x_i = covariates respectively. The maximum likelihood estimation can be obtained via an iterative optimization process (e.g., the Newton-Raphson algorithm). Hypothesis testing and interval estimations of model parameters are possible due to the MLEs' approaching normalcy. The likelihood function's natural logarithm, often known as the log-likelihood function, is expressed as follows:

$$\begin{aligned} \ell\left(\theta, \beta_1, \beta_2 \mid D\right) = &\sum_{i=1}^{n} \delta_i \log\left[h_0\left(te^{{x_i}'\beta_1} \mid \theta\right) + {x_i}'\beta_2\right] \\ &- \sum_{i=1}^{n} H_0\left(t_i e^{{x_i}'\beta_1} \mid \theta\right) e^{-x_i\beta_1 + {x_i}'\beta_2}, \end{aligned} \tag{7.37}$$

Using Equation (**??**) for the $h_0(.)$ and Equation (**??**) for the $H_0(.)$ as the baseline hazard and cumulative hazard functions respectively for the GLL-GH model. The full log-likelihood function of the GLL-GH model can be expressed as follows:

$$\begin{aligned}\ell\left(\theta,\beta_1,\beta_2 \mid D\right) =& \sum_{i=1}^{n}\delta_i\log(\alpha)+\sum_{i=1}^{n}\delta_i\alpha\log(k)+(\alpha-1)\sum_{i=1}^{n}\delta_i\log\left(t_i e^{x_i'\beta_1}\right)\\ &-\sum_{i=1}^{n}\delta_i\log\left[1+\left(\eta t_i e^{x_i'\beta_1}\right)^{\alpha}\right]+\sum_{i=1}^{n}\delta_i\log\left[e^{x_i'\beta_2}\right]\\ &-\left(\frac{k}{\eta}\right)^{\alpha}\sum_{i=1}^{n}e^{-x_i'\beta_1+x_i\beta_2}\log\left[1+\left(\eta t_i e^{x_i'\beta_1}\right)^{\alpha}\right]\end{aligned} \tag{7.38}$$

To obtain the MLE's of $\boldsymbol{\theta'}=(k,\alpha,\eta),\beta_1$, and β_2, we can maximize (47) directly with respect to $(k,\alpha,\eta),\beta_1$, and β_2. or we can solve the first derivative of the log-likelihood function (non-linear equations below). Let us $\omega=(k,\alpha,\eta,\beta_1$, and $\beta_2)$, then the first derivatives of the log-likelihood functions are as follows:

$$\begin{aligned}\frac{\partial\ell(\boldsymbol{\omega})}{\partial\alpha} =& \frac{1}{\alpha}\sum_{i=1}^{n}\delta_i+\sum_{i=1}^{n}\delta_i\log(k)+\sum_{i=1}^{n}\delta_i\log\left(t_i e^{x_i'\beta_1}\right)\\ &+\sum_{i=1}^{n}\delta_i\frac{\left(\eta t_i e^{x_i'\beta_1}\right)^{\alpha}\log\left(\eta t_i e^{x_i'\beta_1}\right)}{\left[1+\left(\eta t_i e^{x_i\beta_1}\right)^{\alpha}\right]}\\ &-\left(\frac{k}{\eta}\right)^{\alpha}\log(k)\sum_{i=1}^{n}e^{-x_i'\beta_1+x_i'\beta_2}\log\left[1+\left(\eta t_i e^{x_i'\beta_1}\right)^{\alpha}\right]\\ &+\left(\frac{k}{\eta}\right)^{\alpha}\log(\eta)\sum_{i=1}^{n}e^{-x_i'\beta_1+x_i'\beta_2}\log\left[1+\left(\eta t_i e^{x_i'\beta_1}\right)^{\alpha}\right]\\ &-\left(\frac{k}{\eta}\right)^{\alpha}\sum_{i=1}^{n}\frac{e^{-x_i'\beta_1+x_i'\beta_2}\left(\eta t_i e^{x_i'\beta_1}\right)^{\alpha}\log\left(\eta t_i e^{x_i\beta_1}\right)}{\left[1+\left(\eta t_i e^{x_i'\beta_1}\right)^{\alpha}\right]}\end{aligned} \tag{7.39}$$

$$\begin{aligned}\frac{\partial\ell(\omega)}{\partial\eta} =& -\left(\frac{\alpha}{\eta}\right)\sum_{i=1}^{n}\delta_i\frac{\left(\eta t_i e^{x_i'\beta_1}\right)^{\alpha}}{\left[1+\left(\eta t_i e^{x_i'\beta_1}\right)^{\alpha}\right]}\\ &+\left(\frac{\alpha}{\eta}\right)\left(\frac{k}{\eta}\right)^{\alpha}\sum_{i=1}^{n}e^{-x_i'\beta_1+x_i'\beta_2}\log\left[1+\left(\eta t_i e^{x_i'\beta_1}\right)^{\alpha}\right]\\ &-\left(\frac{\alpha}{\eta}\right)\left(\frac{k}{\eta}\right)^{\alpha}\sum_{i=1}^{n}\frac{e^{-x_i'\beta_1+x_i'\beta_2}\left(\eta t_i e^{x_i'\beta_1}\right)^{\alpha}}{\left[1+\left(\eta t_i e^{x_i\beta_1}\right)^{\alpha}\right]}\end{aligned} \tag{7.40}$$

$$\frac{\partial \ell(\omega)}{\partial \beta_{1,j}} = (\alpha-1)\sum_{i=1}^{n}\delta_i x_{ij} - \alpha\sum_{i=1}^{n}\delta_i x_{ij}\frac{\left(\eta t_i e^{x_i{}'\beta_1}\right)^{\alpha}}{[1+(\eta t_i e^{x_i{}'\beta_1})^{\alpha}]}$$
$$+\left(\frac{k}{\eta}\right)^{\alpha}\sum_{i=1}^{n}x_{ij}\log\left[1+\left(\eta t_i e^{x_i'\beta_1}\right)^{\alpha}\right] \tag{7.41}$$
$$-\left(\frac{k}{\eta}\right)^{\alpha}\alpha\sum_{i=1}^{n}x_{ij}\frac{e^{-x_i{}'\beta_1+x_i'\beta_2}\left(\eta t_i e^{x_i{}'\beta_1}\right)^{\alpha}}{[1+(\eta t_i e^{x_i{}'\beta_1})^{\alpha}]} + \text{ for } j=1,2,\ldots,p,$$

$$\frac{\partial \ell(\omega)}{\partial \beta_{1,j}} = (\alpha-1)\sum_{i=1}^{n}\delta_i x_{ij} - \alpha\sum_{i=1}^{n}\delta_i x_{ij}\frac{\left(\eta t_i e^{x_i{}'\beta_1}\right)^{\alpha}}{[1+(\eta t_i e^{x_i{}'\beta_1})^{\alpha}]}$$
$$+\left(\frac{k}{\eta}\right)^{\alpha}\sum_{i=1}^{n}x_{ij}\log\left[1+\left(\eta t_i e^{x_i'\beta_1}\right)^{\alpha}\right] \tag{7.42}$$
$$-\left(\frac{k}{\eta}\right)^{\alpha}\alpha\sum_{i=1}^{n}x_{ij}\frac{e^{-x_i{}'\beta_1+x_i'\beta_2}\left(\eta t_i e^{x_i{}'\beta_1}\right)^{\alpha}}{[1+(\eta t_i e^{x_i{}'\beta_1})^{\alpha}]} + \text{ for } j=1,2,\ldots,p,$$

$$\frac{\partial \ell(\boldsymbol{\omega})}{\partial \beta_{2,j}} = \sum_{i=1}^{n}\delta_i x_{ij}$$
$$-\left(\frac{k}{\eta}\right)^{\alpha}\sum_{i=1}^{n}x_{ij}\left[1+\left(\eta t_i e^{x_i'\beta_1}\right)^{\alpha}\right]e^{-x_i'\beta_1+x_i'\beta_2} \quad \text{for } j=1,2,\ldots,p, \tag{7.43}$$

By adjusting the initial partial derivatives, the MLEs for the unknown distributional parameters are obtained for $\boldsymbol{\theta}' = (k, \alpha, \eta)$, and the regression coefficients β_1, and β_2, by solving the non-linear equations $\frac{\partial \ell(\omega^*)}{\partial \alpha^*} = 0$, $\frac{\partial \ell(\omega^*)}{\partial \eta^*} = 0$, $\frac{\partial \ell(\omega^*)}{\partial k^*} = 0$, $\frac{\partial \ell(\omega^*)}{\partial \beta_{1,j^*}} = 0$, and $\frac{\partial \ell(\omega^*)}{\partial \beta_{2,j^*}} = 0$ iteratively. To maximize log-likelihood functions, many software packages include proven optimization algorithms. We utilized the function nlminb to optimize our computer code, which was written in R software.

The approximate normality of the maximum likelihood estimators is used in the tests and interval estimates for the model distributional parameters and regression coefficients. The asymptotic distribution of $\widehat{\omega}^*$ is roughly a $(p+3)$ - variate normal distribution having mean ω^* and covariance matrix $I\left(\widehat{\omega}^*\right)^{-1}$ where is the observed information matrix $(p+3)\times(p+3)$. The observed information matrix is utilized to build confidence intervals for the model parameters because the expected information matrix is difficult. The following is a representation of the observed information

matrix:

$$J\left(\hat{\boldsymbol{\omega}}^{*}\right)=-\left[\begin{array}{ccccc}
\frac{\partial \ell^{2}\left(\boldsymbol{\omega}^{*}\right)}{\partial \alpha^{*2}} & \frac{\partial \ell^{2}\left(\boldsymbol{\omega}^{*}\right)}{\partial \alpha^{*} \partial k^{*}} & \frac{\partial \ell^{2}\left(\boldsymbol{\omega}^{*}\right)}{\partial \alpha^{*} \partial \eta^{*}} & \frac{\partial \ell^{2}\left(\boldsymbol{\omega}^{*}\right)}{\partial \alpha^{*} \partial \beta_{1, j}{ }^{*}} & \frac{\partial \ell^{2}\left(\boldsymbol{\omega}^{*}\right)}{\partial \alpha^{*} \partial \beta_{2, j}{ }^{*}} \\
& \frac{\partial \ell^{2}\left(\omega^{*}\right)}{\partial k^{*2}} & \frac{\partial \ell^{2}\left(\boldsymbol{\omega}^{*}\right)}{\partial k^{*} \partial \eta^{*}} & \frac{\partial \ell^{2}\left(\boldsymbol{\omega}^{*}\right)}{\partial k^{*} \partial \beta_{1, j}{ }^{*}} & \frac{\partial \ell^{2}\left(\boldsymbol{\omega}^{*}\right)}{\partial k^{*} \partial \beta_{2, j}{ }^{*}} \\
& & \frac{\partial \ell^{2}\left(\boldsymbol{\omega}^{*}\right)}{\partial \eta^{* 2}} & \frac{\partial \ell^{2}\left(\boldsymbol{\omega}^{*}\right)}{\partial \eta^{*} \partial \beta_{1, j}{ }^{*}} & \frac{\partial \ell^{2}\left(\boldsymbol{\omega}^{*}\right)}{\partial \eta^{*} \partial \beta_{2, j}{ }^{*}} \\
& & & \frac{\partial \ell^{2}\left(\boldsymbol{\omega}^{*}\right)}{\partial \beta_{1, j}{ }^{* 2}} & \frac{\partial \ell^{2}\left(\omega^{*}\right)}{\partial \beta_{1, j}{ }^{*} \partial \beta_{2, j}{ }^{*}} \\
& & & & \frac{\partial \ell^{2}\left(\omega^{*}\right)}{\partial \beta_{2, j}{ }^{*}}
\end{array}\right]$$

The asymptotic distribution is likewise nearly normal using the multivariate delta technique, with mean ω^{*} and covariance matrix $D \sum D^{\prime}$, where D is the $(p+3) \times (p+3)$ diagonal matrix $\operatorname{diag}(\hat{\theta}, 1, 1, \ldots, 1), I\left(\widehat{\omega}^{*}\right)^{-1}$. As a result, for the model parameters, the asymptotic multivariate normal distribution $N_{5}\left(0, I\left(\widehat{\omega}^{*}\right)^{-1}\right)$ can be utilized to create $100(1-\varphi)\%$ two-sided confidence intervals. The significant level is denoted by the letter φ.

7.6.2 Bayesian Inference

As an alternative, we apply the Bayesian approach, which enables the incorporation of prior understanding of the model parameters using informative prior density functions. In the absence of this knowledge, a noninformative prior may be taken into account. The information pertaining to the model parameters is retrieved using a posterior marginal distribution in the Bayesian technique. Two problems typically result from this. The first speaks of obtaining a marginal posterior distribution, and the second, of computing the important moments. In both situations, numerical integration frequently does not offer an analytical answer. Here, we utilize the Gibbs sampler and Metropolis-Hastings's algorithm as part of the Markov chain Monte Carlo (McMC) simulation approach.

By defining the prior distributions for model unknown parameters, followed by multiplying by the likelihood function, the Bayesian model is created.

7.6.2.1 Priors for the Parameters

An essential component of every Bayesian inference is the specification of a prior distribution. This is particularly true for parametric hazard-based regression models. Due to the flexibility of gamma distributions, which include non-informative priors (uniform) and the marginal prior distribution for each regression coefficient $(m,)m = 1,\ldots,,5$, the prior scenario is set in this study using a non-informative independent gamma distribution, Gamma $(10, 10)$, as the baseline distribution parameters because we have no prior information from historical data or from previous experiments and a normal distribution with zero mean and a wide known variance $(0, 100)$ for the regression coefficients. These priors are taken into consideration in numerous study publications in the literature, including (Alvares et al., 2021; Khan and Basharat, 2022; Muse et al., 2022g,a). Here,

$$\pi(\alpha) \sim G\left(a_1, b_1\right) = \frac{b_1^{a_1}}{\Gamma\left(a_1\right)} \alpha^{a_1-1} e^{-b_1\alpha}; a_1, b_1, \alpha > 0 \tag{7.44}$$

$$\pi(\eta) \sim G\left(a_2, b_2\right) = \frac{b_2^{a_2}}{\Gamma\left(a_2\right)} \eta^{a_2-1} e^{-b_2\eta}; a_2, b_2, \eta > 0 \tag{7.45}$$

$$\pi(k) \sim G\left(a_3, b_3\right) = \frac{b_3^{a_3}}{\Gamma\left(a_2\right)} k^{a_3-1} e^{-b_3 k}; a_3, b_3, k > 0 \tag{7.46}$$

The hyper-parametric values of the prior distributions may be simply determined from historical data of the baseline distribution (Muse et al., 2021a). For the regression coefficients prior (taken as a normal distribution), we have

$$\pi\left(\boldsymbol{\beta}_1'\right) \sim N\left(a_4, b_4\right) \tag{7.47}$$

$$\pi\left(\boldsymbol{\beta}_2'\right) \sim N\left(a_5, b_5\right) \tag{7.48}$$

The joint prior distribution of all unknown parameters has a density function given

by

$$\pi\left(\alpha, k, \eta, \boldsymbol{\beta}_1', \boldsymbol{\beta}_2'\right) = \pi(\alpha)\pi(\eta)\pi(k)\pi\left(\boldsymbol{\beta}_1'\right)\pi\left(\boldsymbol{\beta}_2'\right) \tag{7.49}$$

7.6.2.2 Likelihood Function

The likelihood function for the generalized-log-logistic general hazard model is computed as follows:

$$\begin{aligned} L_{GLL-GH}\left(\theta, \beta_1, \beta_2 \mid D\right) &= \prod_{i=1}^{n}\left[h_0\left(t_i e^{x_i'\beta_1} \mid \theta\right) e^{x_i'\beta_2}\right]^{\delta_i} \exp\left[-H_0\left(t_i e^{x_i'\beta_1} \mid \theta\right) e^{x_i'\beta_2 - x_i'\beta_1}\right] \\ &= \prod_{i=1}^{n}\left[\frac{\alpha k\left(k t_i e^{x_i'\beta_1}\right)^{\alpha-1}}{\left[1+\left(\eta t_i e^{x_i'\beta_1}\right)^{\alpha}\right]} e^{x_i'\beta_2}\right]^{\delta_i} \\ &\quad \exp\left[-\left\{\frac{k^{\alpha}}{\eta^{\alpha}}\log\left[1+\left(\eta t_i e^{x_i{}'\beta_1}\right)^{\alpha}\right]\right\} e^{x_i'\beta_2 - x_i'\beta_1}\right] \end{aligned} \tag{7.50}$$

7.6.2.3 Posterior Distribution

The joint posterior density function is expressed as the multiplication of the likelihood function in Equation (7.50) and the prior distribution in Equation (7.49):

$$\begin{aligned} & p\left(\alpha, k, \eta, \boldsymbol{\beta_1}', \boldsymbol{\beta_2}' \mid t\right) \\ &\propto \prod_{i=1}^{n}\left[\frac{\alpha k\left(k t_i e^{x_i'\beta_1}\right)^{\alpha-1}}{\left[1+\left(\eta t_i e^{x_i'\beta_1}\right)^{\alpha}\right]} e^{x_i'\beta_2}\right]^{\delta_i} \\ &\quad \exp\left[-\left\{\frac{k^{\alpha}}{\eta^{\alpha}}\log\left[1+\left(\eta t_i e^{x_i{}'\beta_1}\right)^{\alpha}\right]\right\} e^{x_i'\beta_2 - x_i'\beta_1}\right] \\ &\quad \times \pi\left(\alpha, k, \eta, \boldsymbol{\beta_1}', \boldsymbol{\beta_2}'\right) \end{aligned} \tag{7.51}$$

where the prior specification for the unknown parameters is represented by the first four terms on the right-hand side of the equation.

Due to the difficulty of integrating the joint posterior density, the joint posterior density is analytically intractable. The inference can therefore be based on McMC simulation techniques, including the Gibbs sampler and Metropolis-Hastings algo-

rithms, which can be used to provide samples from which properties of the marginal distributions of interest can be inferred.

7.7 Simulation Study

We provide two simulation experiments in this section that illustrate the inferential properties, model suitability, nested structure, and estimator performance of the suggested model.

7.7.1 Simulation Study I: Comparative Study

Simulation studies I discuss the proposed model's classical approach as well as their special cases, which include the AH, AFT, and PH models. The goal of this study is to show how the proposed model's nested structure compares to the most commonly used parametric approaches for survival data analysis. We use information criterion such as AIC and CAIC to choose models that accurately reflect the underlying model structure as well as the effects of censoring percentage and sample size on parameter estimation.

7.7.1.1 Data Generation from the Hazard-Based Regression Models

We employed the inversion method to generate survival data from the general class of hazard-based regression models and their special cases such as AH, AFT, and PH. This technique is based on the relationship between the cumulative hazard function (chf) of a survival random variable and a standard uniform random variable. Whenever the chf has a closed form solution, it may be applied, inverted, and easily implemented in R (Team et al., 2013).

Since the Cox PH model is the most widely used in survival analysis, we took into consideration the method of Bender et al. (Bender et al., 2005) that they used to simulate data from the Cox regression model. We also thought about the Leemis et

al. (Leemis et al., 1990) methods for simulating survival data from an AFT model, and we used the same method for the rest of the AH and our proposed GH model (Austin, 2012).

The cdf is deduced from the survival function from the following formula:

$$F(t;x) = 1 - S(t;x) \tag{7.52}$$

As a result, for lifetime data generation, if Y is a random variable that follows a cdf F, then $U = F(Y)$ follows a uniform distribution on the interval $[0,1]$, and $(1-U)$ also follows a uniform distribution $\mathcal{U}[0,1]$. We eventually get that

$$1 - U = S(t) = \exp\{-H_0(t;x)\} \tag{7.53}$$

Then,

$$\exp\{-H_0(t;x)\} = 1 - U \tag{7.54}$$

Taking into account the cumulative hazard function for the GH model in Equation (7.5) it follows as:

$$\exp\left\{-H_0\left(te^{x'\beta_1}\right)e^{x'\beta_2 - x'\beta_1}\right\} = 1 - U \tag{7.55}$$

The generation of survival times for the proposed model and its special cases can be described in the following general structural form:

$$T = \frac{1}{e^{x'\beta_1}} H_0^{-1}\left(\frac{-\log(1-U)}{e^{x'\beta_2 - x'\beta_1}}\right) \tag{7.56}$$

With

$$(e^{x'\beta_1}, e^{x'\beta_2}) = \begin{cases} \beta_1, \beta_2 = 0 & \text{for the AH model} \\ \beta 1 = \beta_2 & \text{for the AFT model} \\ \beta_1 = 0, \beta_2 & \text{for the PH model} \\ \beta_1, \beta_2 & \text{for the GH model} \end{cases}$$

If the baseline hazard rate is strictly positive for all t,then baseline cumulative hazard rate can be inverted, and we can express the lifetime data of each of the hazard-based regression models considered (PH, AFT, PH and GH) model from $H_0^{-1}(u)$.
In our case, the cumulative hazard rate for the GLL distribution is of the form:

$$H(t; \alpha, \eta, k) = \frac{k^\alpha}{\eta^\alpha} \log\left[1 + (\eta t)^\alpha\right], \quad t \geq 0, \quad k, \alpha, \eta > 0$$

Consequently, the inverse of the cumulative hazard function is expressed as follows:

$$H_0^{-1}(U; \alpha, \eta, k) = \frac{\left(e^{\eta^\alpha k^{-\alpha} U} - 1\right)^{\frac{1}{\alpha}}}{\eta}$$

In this study, we used the baseline cumulative hazard rate function and its reverse $(H_0(t)$ and $H_0^{-1}(t))$ to generate the survival data.

Case I: GH Model

In case 1 , the lifetimes of the GH model is expressed as follows:

$$T = \frac{1}{e^{x'\beta_1}} H_0^{-1}\left(\frac{-\log(1-U)}{e^{x'\beta_2 - x'\beta_1}}\right).$$

For this simulation, we consider that the survival times follow a GLL baseline, therefore the survival times can therefore be simulated from:

$$T = \frac{\left(e^{\eta^\alpha k^{-\alpha}\left(-\log(1-u)e^{-x'\beta_2}\right).} - 1\right)^{\frac{1}{\alpha}}}{\eta e^{x'\beta_2 - x'\beta_1}}.$$

Case II: AFT Model

In case 2, we generate the survival data from an AFT model as follows:

$$T = \frac{H_0^{-1}\{-\log(1-U)\}}{e^{x'\beta}}.$$

Using GLL baseline:

$$T = \frac{\left(e^{\eta^\alpha k^{-\alpha}(-\log(1-u)}\right) - 1)^{\frac{1}{\alpha}}}{\eta e^{x'\beta}}.$$

Case III: AH Model

In case 3 , we generate the survival data from an AH model as follows:

$$T = \frac{1}{e^{x'\beta}} H_0^{-1}\left(-\frac{\log(1-U)}{e^{-x'\beta})}\right).$$

Using GLL baseline:

$$T = \frac{\left(e^{\eta^\alpha k^{-\alpha}\left(-\log(1-u)e^{x'\beta}\right)\cdot} - 1\right)^{\frac{1}{\alpha}}}{\eta e^{-x'\beta}}.$$

Case IV: PH Model

In case 4 , we generate the survival data from a PH model as follows:

$$T = H_0^{-1}\left(\frac{-\log(1-U)}{e^{x'\beta}}\right).$$

Using GLL baseline:

$$T = \frac{\left(e^{\eta^{\alpha^\alpha} k^{-\alpha}\left(-\log(1-u)e^{-x'\beta}\right)} - 1\right)^{\frac{1}{\alpha}}}{\eta}.$$

7.7.1.2 Simulation Design

Using a severe cancer (with a reduced five-year survival rate), such as lung cancer, the initial values of the parameters are set to create scenarios that imitate cancer population studies (Rubio et al., 2019, 2021).

i. The administrative censorship at Tc = 5 years, which produced an average of 20percent censoring in all data sets.

ii. an extra random samples censorship (drop out) utilizing exponential distribution with the rate parameter r were employed to estimate the censoring rates.

In the second scenario, we select r values that will cause censoring of roughly 40 percent. Based on the GH model in Equation (7.4), a series of simulations with N = 10000 data sets of various sample size (n = 5000, and 10000) set and censoring percentages (Tc = 20 and 40 percentages) were conducted.
The covariates' values were simulated as follows: (1) the continuous covariate "age" was simulated using a collection of uniform distributions with 0.25 probability on $(30, 65)$, 0.35 probability on $(65, 75)$, and 0.40 probability on $(75, 85)$ years old; as well as (2) the binary covariates "treatment" and "gender" were simulated using a 0.5 binomial distribution. For more information, we advise the reader to visit (Rubio et al., 2019, 2021; Muse et al., 2022g; Rubio et al., 2022).

7.7.1.3 Simulation Scenarios

To compare the nested structure of the proposed GH model with generalized log-logistic baseline hazard to its special cases, the AH, PH, and AFT regression models. We conducted four simulation scenarios based on the types of hazard-based regression model frameworks (AH, AFT, PH, and GH).

Scenario 1: GH Framework

In Scenario 1, the survival times data were obtained from a general hazard (GH) framework with a GLL baseline hazard using the distributional parameter values for ($k = 0.625, \alpha = 1.50$, and $\eta = 1.0$) and the covariates values for ($\beta = 0.75, 0.85, 0.95$, $\beta_H = 0.35, 0.45, 0.55$).The censored times data were produced from assuming administrative censoring (a) Tc = 5, which generated about 20% censoring, (b) an extra independent random censoring (i.e., dropout) using an exponential distribution with rate parameter r and we select values for r to generate about 40% censoring.

Scenario 2: AFT Framework

In Scenario 2, the survival times data were obtained from an AFT framework with a GLL baseline hazard using the using the distributional parameter values for ($k = 0.675$, $\alpha = 1.50$, and $\eta = 1.0$) and the covariates values for ($\beta = 0.75, 0.85, 0.95, \beta_H = 0.75, 0.85, 0.95$).

The censored times data were produced from assuming administrative censoring (a) Tc = 5, which generated about 20% censoring, (b) an extra independent random censoring (i.e., dropout) using an exponential distribution with rate parameter r and we select values for r to generate about 40% censoring.

Scenario 3: PH Framework

In Scenario 3, the survival times data were obtained from a PH framework with a GLL baseline hazard using the using the distributional parameter values for ($k = 0.65$, $\alpha = 1.50$, and $\eta = 1.0$) and the covariates values for ($\beta = 0, 0, 0, \beta_H = 0.15, 0.25, 0.35$). The censored times data were produced from assuming administrative censoring (a) $Tc = 5$, which generated about 20% censoring, (b) an extra independent random censoring (i.e., dropout) using an exponential distribution with rate parameter r and we select values for r to generate about 40% censoring.

Scenario 4: AH Framework

In Scenario 4, the survival times data were obtained from an AH framework with a GLL baseline hazard using the distributional parameter values for ($k = 0.80, \alpha =$

1.50, and $\eta = 1.0$) and the covariates values for $(\beta = 0.15, 0.25, 0.35, \beta_H = 0, 0, 0)$. The censored times data were produced from assuming administrative censoring (a) $Tc = 5$, which generated about 20% censoring, (b) an extra independent random censoring (i.e., dropout) using an exponential distribution with rate parameter r and we select values for r to generate about 40% censoring.

7.7.1.4 Results for Scenario 1

For Scenario 1, the degree of censoring seems to affect how well a model fits the data. The GH model performs better than the AFT, PH, and AH models overall. The AFT outperforms the other hazard-based models, including PH and AH, in terms of information criteria. Generally speaking, it appears that the AH has the most information criteria. As can be seen in Figures 7.1 through 7.4, the AFT and PH models suit the data the best, while the AH model fits the data the least well. Although it appears that the AFT is overestimated and the PH is underestimated, both of them suit the data better than the AH model. As seen in Table 7.2, every competing model demonstrates how the censoring proportions' increase has an impact on the model's performance in terms of information criteria. The identical thing takes place in Table 7.3. The PH outperforms better than the AFT and AH in Table 7.3 for the lighter censoring, while when the censoring becomes heavier, the AFT is the one that outperforms better compared to the PH and AH models. In general, AFT is the one that is superior after the GH model, since the covariates of the model effect are for both hazard and time scale.

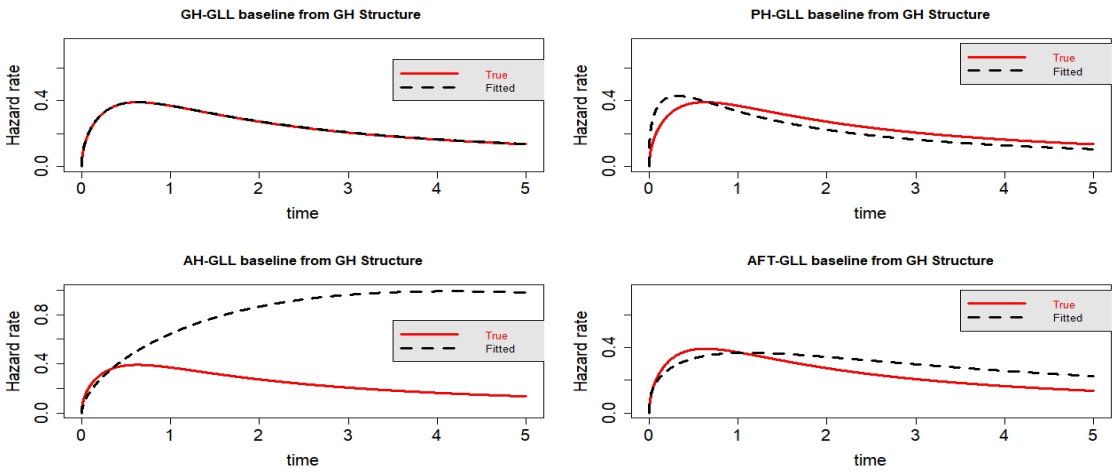

Figure 7.1: Estimated baseline hazard functions with censoring proportion of 20% and a sample size of $n = 5000$. The dashed and solid curves indicate the estimated and true hazard rate functions, accordingly. The data generated from a GH structure

Table 7.2: Simulation results from GH model with $(k = 0.625, \alpha = 1.50,$ and $\eta = 1.0)$, covariates values for $(\beta_1 = 0.75, 0.85, 0.95,\ \beta_2 = 0.35, 0.45, 0.55\ )$ and $n = 5,000$ with about 20% censoring. AIC, CAIC, and HQIC values for the competitive models.

Model	**AIC**	**CAIC**	**HQIC**
		20% Censoring	
GLL-GH Model	9068.327	9068.705	9088.884
GLL-AFT Model	9202.449	9202.656	9216.154
GLL-AH Model	12264.223	12264.149	122277.928
GLL-PH Model	9222.095	9222.307	9235.800
		40% censoring	
GLL-GH Model	5463.519	5463.598	5484.077
GLL-AFT Model	5505.834	5505.872	5519.540
GLL-AH Model	7424.734	7424.799	7438.440
GLL-PH Model	5548.739	5548.777	5562.444

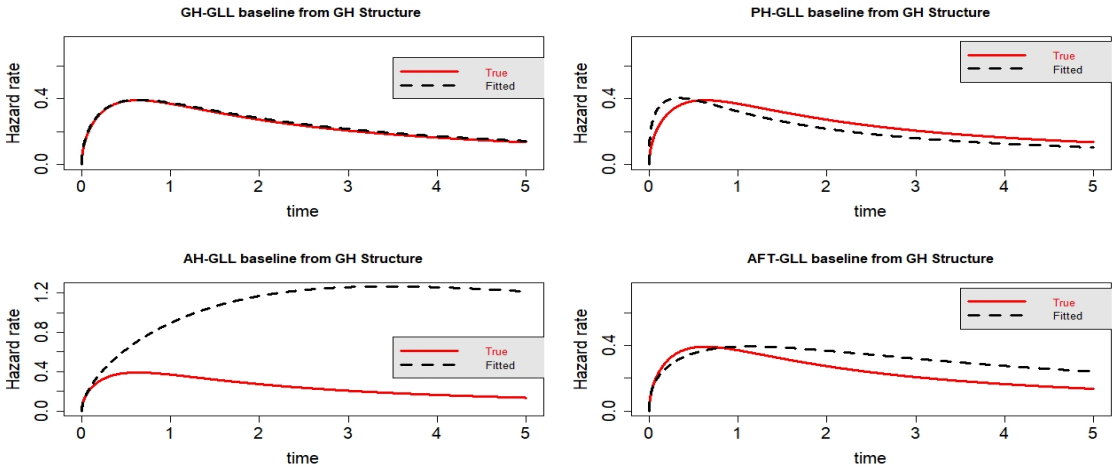

Figure 7.2: Estimated baseline hazard functions with censoring proportion of 40% and a sample size of $n = 5000$. The dashed and solid curves indicate the estimated and true hazard rate functions, accordingly. The data generated from a GH structure

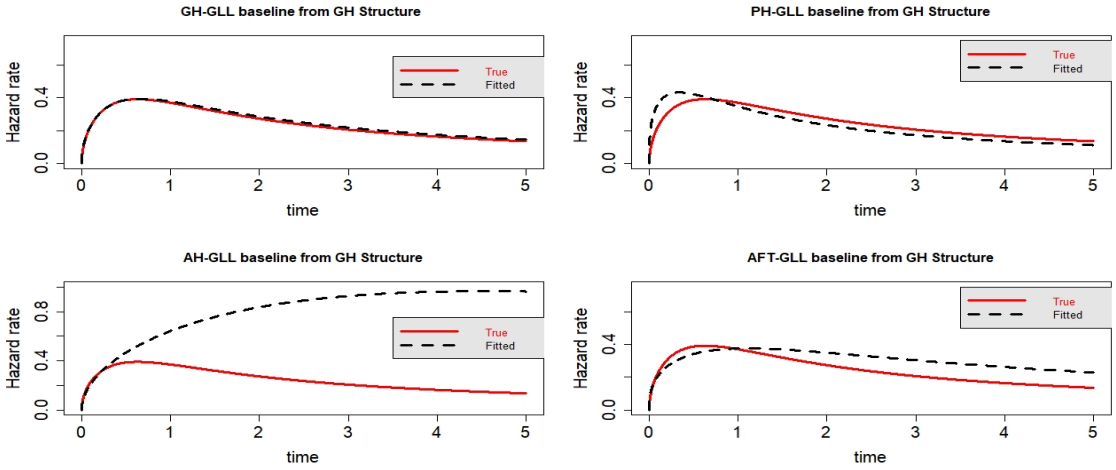

Figure 7.3: Estimated baseline hazard functions with censoring proportion of 40% and a sample size of $n = 10000$. The dashed and solid curves indicate the estimated and true hazard rate functions, accordingly. The data generated from a GH structure

Table 7.3: Simulation results from GH model with $(k = 0.625, \alpha = 1.50,$ **and** $\eta = 1.0)$**, covariates values for** $(\beta_1 = 0.75, 0.85, 0.95,\ \beta_2 = 0.35, 0.45, 0.55\ )$ **and** $n = 10,000$ **with about** 20% **censoring. AIC, CAIC, and HQIC values for the competitive models.**

Model	**AIC**	**CAIC**	**HQIC**
		20% Censoring	
GLL-GH Model	17567.525	175672	17589.491
GLL-AFT Model	17843.352	17843.429	17857.996
GLL-AH Model	24035.240	24035.198	24049.884
GLL-PH Model	17818.738	17818.184	17833.382
		40% censoring	
GLL-GH Model	10719.862	10719.901	10741.828
GLL-AFT Model	10809.107	10809.125	10823.751
GLL-AH Model	14662.298	14662.330	14676.942
GLL-PH Model	10913.655	10913.673	10928.299

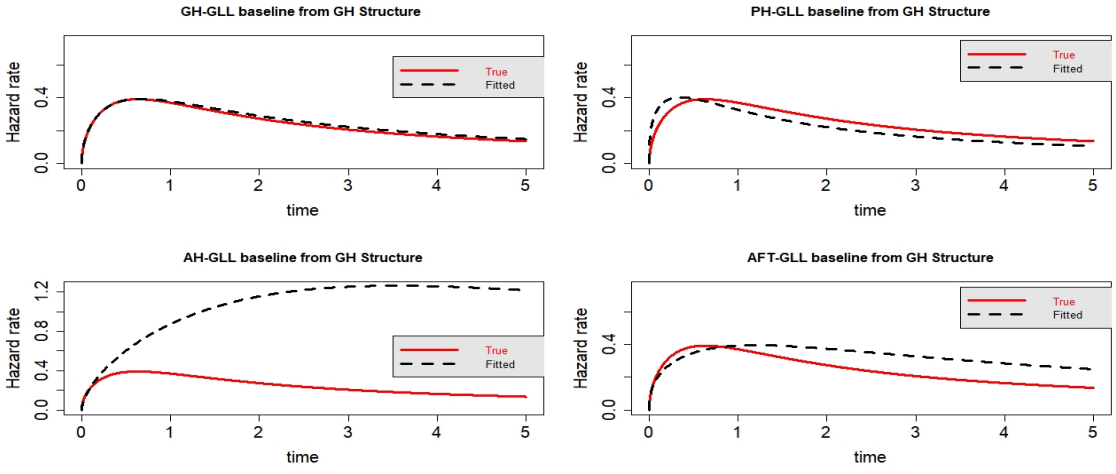

Figure 7.4: Estimated baseline hazard functions with censoring proportion of 40% **and a sample size of** $n = 10000$**. The dashed and solid curves indicate the estimated and true hazard rate functions, accordingly. The data generated from a GH structure**

7.7.1.5 Results for Scenario 2

In Scenario 2, a simulation is generated using an AFT framework. The GLL-GH model and the true generated GLL-AFT model were all superior to the GLL-PH and GLL-AH models for the information criteria values, such as the AIC and CAIC values for Scenario 2. This demonstrates how the AFT model is a specific case of the GH framework. The generated model has the lowest AIC, CAIC, and HQIC values as would be predicted given that it is an AFT framework.

The GLL-GH model has the lowest AIC and CAIC values when the sample size and censoring fraction increase to $n = 10000$ and 40% censoring, respectively, as shown in Tables 7.4 and 7.5. This shows that the GH structure performs better than its special cases when there is heavy censoring of the data. However, from the visual representations, it is clear that the GLL-GH and GLL-AFT models exhibit certain similarities and provide the best fit when compared to the other two rival models. We can therefore deduce from Scenario 2 that the AFT model is a sub-model of the GH model, as illustrated in Figures 7.5 to 7.8.

Table 7.4: Simulation results from AFT model with $(k = 0.675, \alpha = 1.50,$ **and** $\eta = 1.0)$**, covariates values for** $(\beta_1 = 0.75, 0.85, 0.95,\ \beta_2 = 0.75, 0.85, 0.95)$ **and** $n = 5,000$ **with about** 20% **and** 40% **censoring. AIC, CAIC, and HQIC values for the competitive models.**

	AIC	**CAIC**	**HQIC**
		20% Censoring	
GLL-AFT Model	9710.030	9710.583	9723.735
GLL-GH Model	9713.632	9714.807	9734.190
GLL-AH Model	12766.930	12766.869	12780.635
GLL-PH Model	10241.712	10240.974	10255.417
		40% Censoring	
GLL-AFT Model	6119.325	6119.367	6133.030
GLL-GH Model	6122.154	6122.246	6142.711
GLL-PH Model	6500.689	6500.739	6514.394
GLL-AH Model	8259.059	8249.154	8262.764

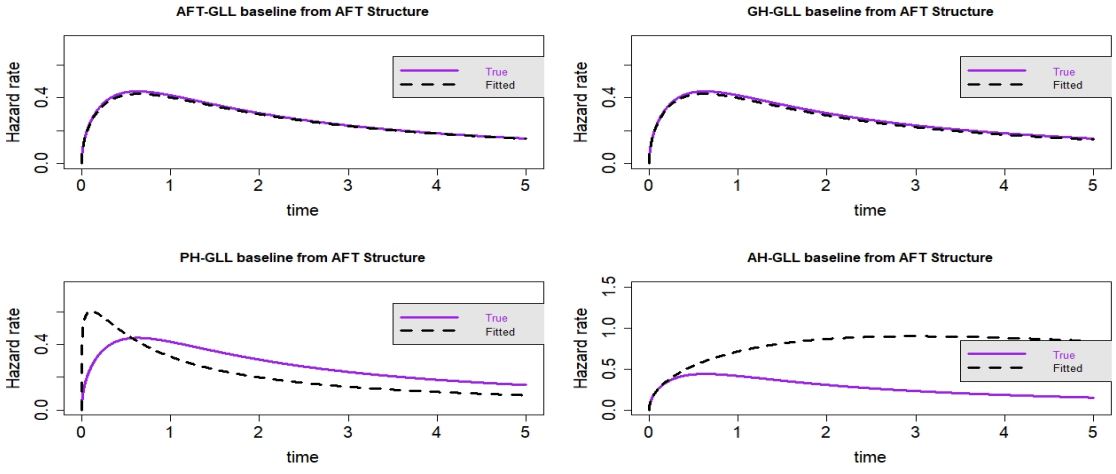

Figure 7.5: Estimated baseline hazard functions with censoring proportion of 20% **and a sample size of** $n = 5000$**. The dashed and solid curves indicate the estimated and true hazard rate functions, accordingly. The data generated from an AFT structure**

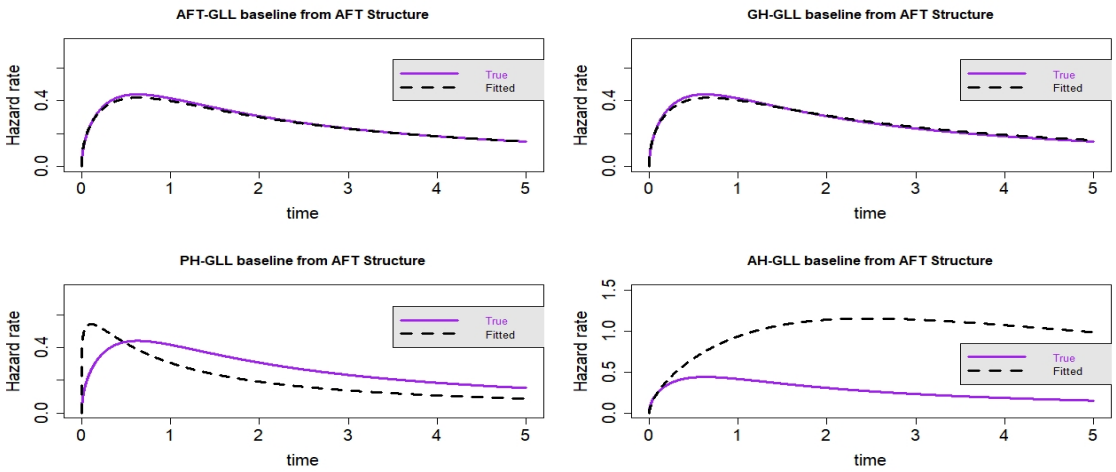

Figure 7.6: Estimated baseline hazard functions with censoring proportion of 40% and a sample size of $n = 5000$. The dashed and solid curves indicate the estimated and true hazard rate functions, accordingly. The data generated from an AFT structure

Table 7.5: Simulation results from AFT model with ($k = 0.675, \alpha = 1.50$, and $\eta = 1.0$), covariates values for ($\beta_1 = 0.75, 0.85, 0.95$, $\beta_2 = 0.75, 0.85, 0.95$) and $n = 10,000$ with about 20% and 40% censoring. AIC, CAIC, and HQIC values for the competitive models.

Model	**AIC**	**CAIC**	**HQIC**
		20% Censoring	
GLL-AFT Model	19192.366	19192.571	19207.011
GLL-GH Model	19195.876	19196.313	19217.842
GLL-AH Model	25043.290	12766.257	25057.934
GLL-PH Model	20201.167	20200.269	20215.811
		40% censoring	
GLL-AFT Model	12214.303	12214.324	12228.947
GLL-GH Model	12211.367	12211.413	12233.332
GLL-PH Model	13036.386	13036.410	13051.030
GLL-AH Model	16233.077	16233.122	16247.721

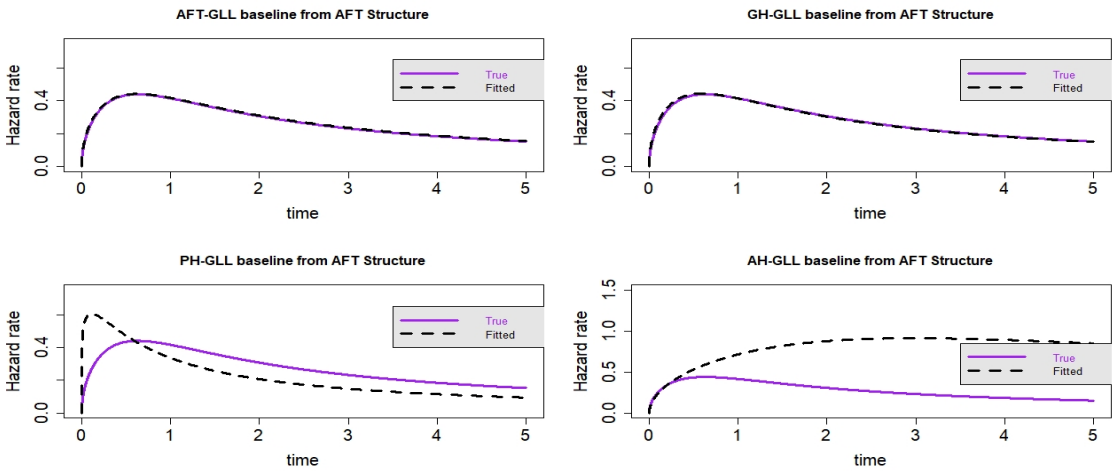

Figure 7.7: Estimated baseline hazard functions with censoring proportion of 40% and a sample size of $n = 10000$. The dashed and solid curves indicate the estimated and true hazard rate functions, accordingly. The data generated from an AFT structure

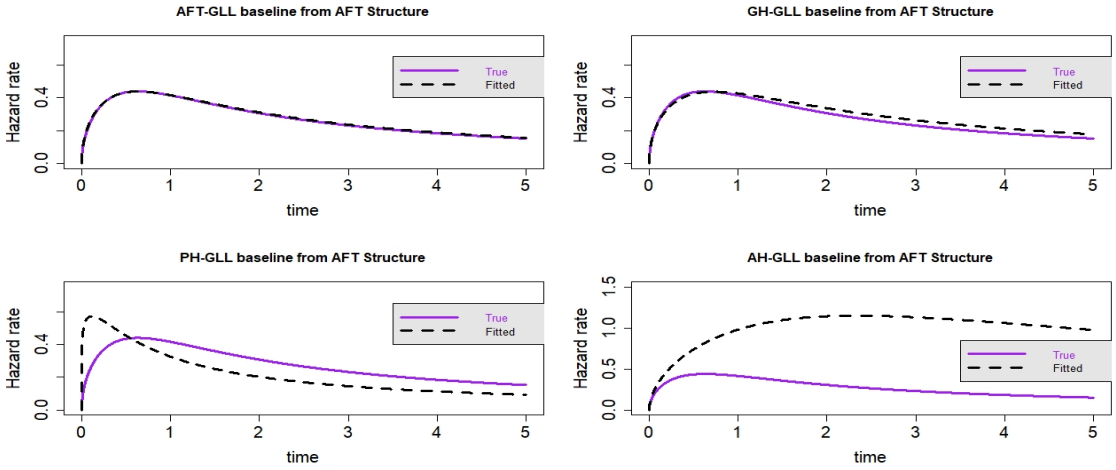

Figure 7.8: Estimated baseline hazard functions with censoring proportion of 40% and a sample size of $n = 10000$. The dashed and solid curves indicate the estimated and true hazard rate functions, accordingly. The data generated from an AFT structure

7.7.1.6 Scenario 3 Results

In Scenario 3, a PH framework is used to generate simulation data. For information criteria values like the AIC and CAIC values for Scenario 3, the GLL-GH model and the true generated GLL-PH model were all superior to the GLL-AH and GLL-AFT

models. This explains how the GH framework is used specifically in the PH model. Since the model created from the data was closed using the PH framework, the resultant GLL-PH model has the lowest AIC, CAIC, and HQIC values as would be expected.

As demonstrated in Table 7.6 and 7.7, the GLL-GH model's AIC and CAIC values are comparable when the censoring fraction is increased to 40%. This demonstrates that when there is severe data censoring, the GH structure outperforms its special cases. But it is obvious from the visual representations in Figures 7.9 to 7.12 that the GLL-GH and GLL-PH models share several characteristics and offer the best fit when compared to the other two competing models. As we may infer from Scenario 3, the PH model is a sub-model of the GH model.

Table 7.6: Simulation results from PH model with $(k = 0.65, \alpha = 1.50$**, and** $\eta = 1.0)$**, covariates values for** $(\beta_2 = 0.15, 0.25, 0.35,\ \beta_1 = 0.0, 0.0, 0.0)$ **and** $n = 5,000$ **with about** 20% **and** 40% **censoring. AIC, CAIC, and HQIC values for the competitive models.**

Model	**AIC**	**CAIC**	**HQIC**
		20 % Censoring	
GLL-PH Model	14941.383	14941.349	14955.088
GLL-GH Model	14947.070	14946.997	14967.627
GLL-AFT Model	15037.394	15037.361	15051.100
GLL-AH Model	15224.134	15224.102	15237.840
		40%	
GLL-PH Model	10966.406	10966.229	10980.111
GLL-GH Model	10966.497	10966.116	10987.054
GLL-AFT Model	11038.579	11038.415	11052.284
GLL-AH Model	11206.019	11205.877	11219.724

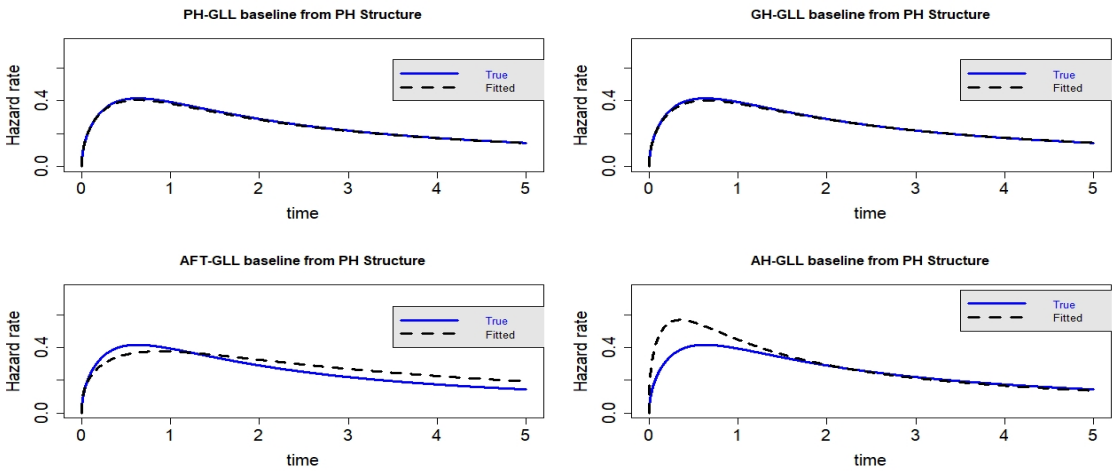

Figure 7.9: Estimated baseline hazard functions with censoring proportion of 20% and a sample size of $n = 5000$. The dashed and solid curves indicate the estimated and true hazard rate functions, accordingly. The data generated from a PH structure

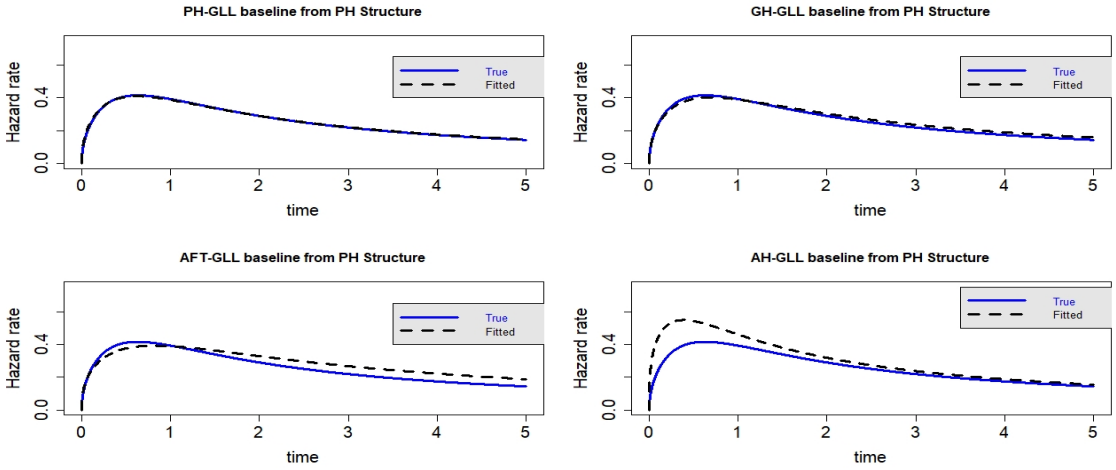

Figure 7.10: Estimated baseline hazard functions with censoring proportion of 40% and a sample size of $n = 5000$. The dashed and solid curves indicate the estimated and true hazard rate functions, accordingly. The data generated from a PH structure

Table 7.7: Simulation results from PH model with ($k = 0.65, \alpha = 1.50$, **and** $\eta = 1.0$)**, covariates values for** ($\beta_2 = 0.15, 0.25, 0.35$, $\beta_1 = 0.0, 0.0, 0.0$) **and** $n = 10,000$ **with about** 20% **and** 40% **censoring. AIC, CAIC, and HQIC values for the competitive models.**

Model	**AIC**	**CAIC**	**BIC**
		20% Censoring	
GLL-PH Model	29768.943	29768.927	29783.588
GLL-GH Model	29772.045	29772.008	29794.011
GLL-AFT Model	29969.703	29969.686	29984.345
GLL-AH Model	30321.478	30321.462	30336.123
		40% Censoring	
GLL-PH Model	22085.485	22085.404	22100.129
GLL-GH Model	22088.317	22088.143	22110.283
GLL-AFT Model	22219.687	22219.611	22234.331
GLL-AH Model	22560.893	22560.828	22575.538

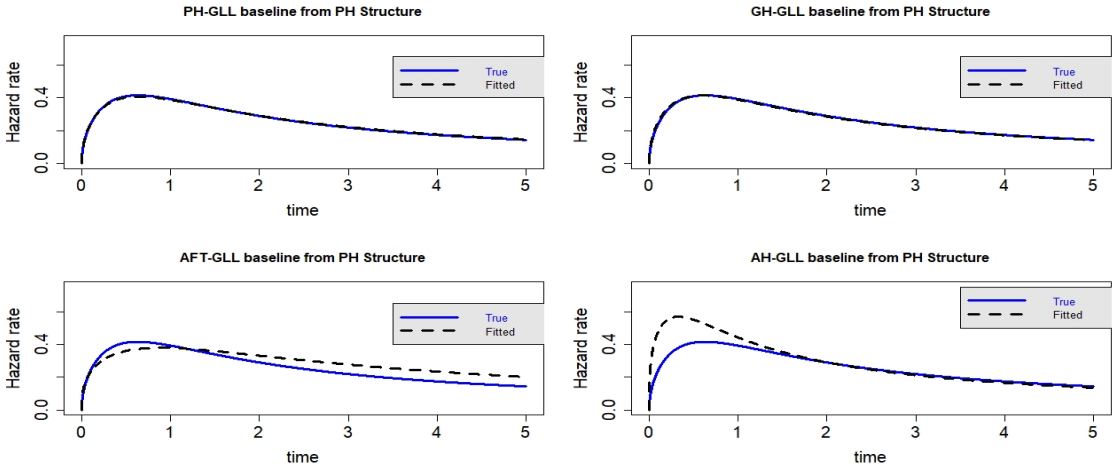

Figure 7.11: Estimated baseline hazard functions with censoring proportion of 40% **and a sample size of** $n = 10000$**. The dashed and solid curves indicate the estimated and true hazard rate functions, accordingly. The data generated from a PH structure**

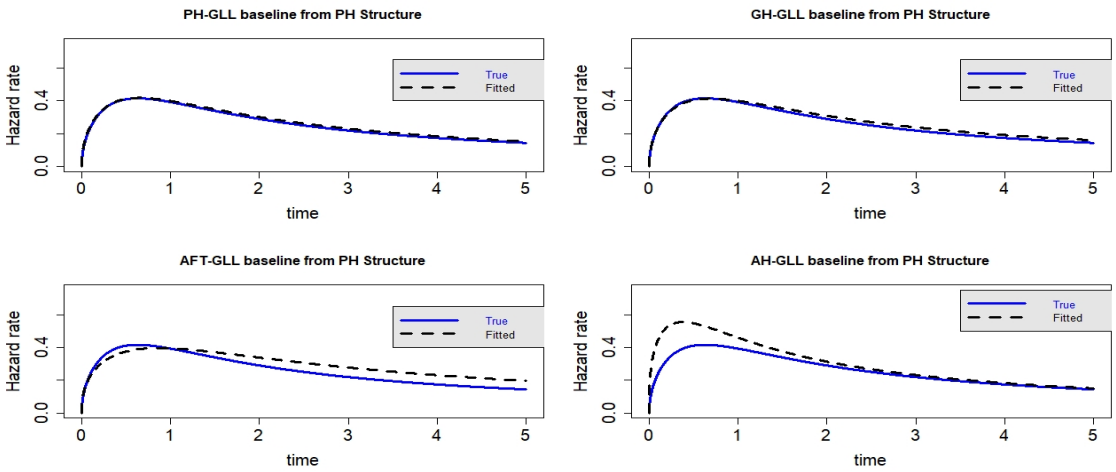

Figure 7.12: Estimated baseline hazard functions with censoring proportion of 40% and a sample size of $n = 10000$. The dashed and solid curves indicate the estimated and true hazard rate functions, accordingly. The data generated from a PH structure

7.7.1.7 Scenario 4 Results

In the context of Scenario 4, some proximity can be accounted for by the visual representation in Figures 7.13 to 7.16 of all the fitted models. Nevertheless, as anticipated, the GLL-GH and GLL-AH models outperform the two other rival models, the GLL-AFT and GLL-PH models. When compared to the other models, the GLL-GH model is the most similar to the actual created model, demonstrating that the GH framework is a general instance of the AH framework. As demonstrated in Tables 7.8 and 7.9, GLL-AH has the lowest value for information criteria because it is the generated model from simulation data.

Table 7.8: Simulation results from AH model with $(k = 0.80, \alpha = 1.50,$ **and** $\eta = 1.0)$**, covariates values for** $(\beta_1 = 0.15, 0.25, 0.35,\ \beta_2 = 0.0, 0.0, 0.0)$ **and** $n = 5,000$ **with about** 20% **and** 40% **censoring. AIC, CAIC, and HQIC values for the competitive models.**

Model	**AIC**	**CAIC**	**HQIC**
		20% Censoring	
GLL-AH Model	15084.818	15084.785	15098.523
GLL-GH Model	15086.243	15086.172	15106.801
GLL-PH Model	15161.428	15161.396	15175.133
GLL-AFT Model	15212.691	15212.658	15226.396
		40% Censoring	
GLL-AH Model	11154.225	11154.078	11167.930
GLL-GH Model	11156.478	11156.161	11177.035
GLL-PH Model	11229.268	11229.130	11242.974
GLL-AFT Model	11252.678	11252.542	11266.382

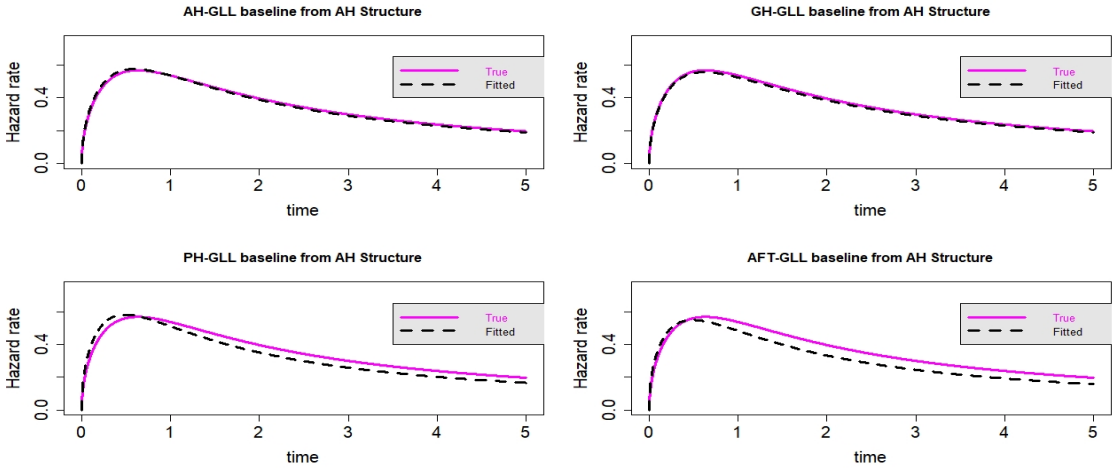

Figure 7.13: Estimated baseline hazard functions with censoring proportion of 20% **and a sample size of** $n = 5000$**. The dashed and solid curves indicate the estimated and true hazard rate functions, accordingly. The data generated from an AH structure**

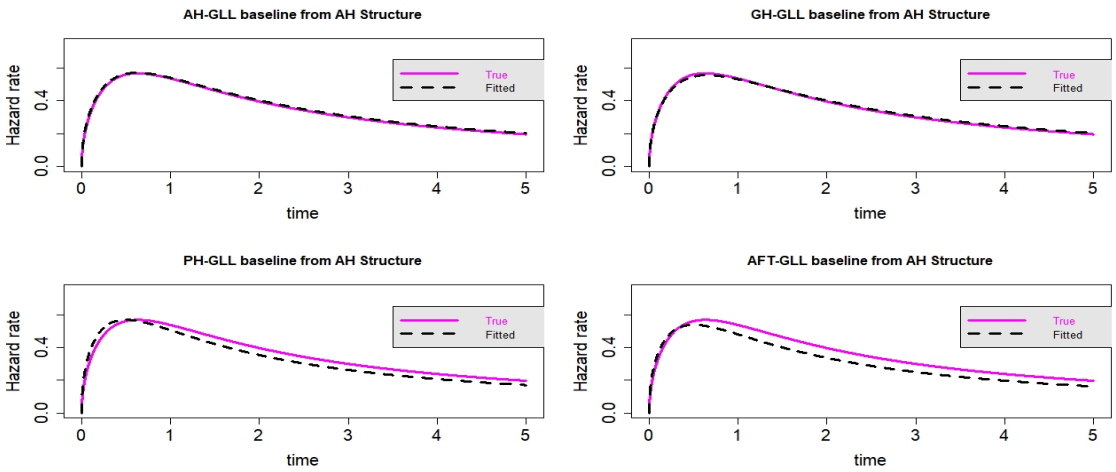

Figure 7.14: Estimated baseline hazard functions with censoring proportion of 40% and a sample size of $n = 5000$. The dashed and solid curves indicate the estimated and true hazard rate functions, accordingly. The data generated from an AH structure

Table 7.9: Simulation results from AH model with $(k = 0.80, \alpha = 1.50,$ and $\eta = 1.0)$, covariates values for $(\beta_1 = 0.15, 0.25, 0.35,\ \beta_2 = 0.0, 0.0, 0.0)$ and $n = 10,000$ with about 20% and 40% censoring. AIC, CAIC, and HQIC values for the competitive models.

Model	**AIC**	**CAIC**	**HQIC**
		20 % Censoring	
GLL-AFT Model	30458.953	30458.937	30473.597
GLL-GH Model	30463.451	30463.416	30485.416
GLL-AH Model	30664.667	30664.651	30679.310
GLL-PH Model	30741.659	30741.644	30756.303
		40 % Censoring	
GLL-AFT Model	22592.371	22592.306	22607.015
GLL-AFT Model	22598.029	22597.890	22619.996
GLL-GH Model	22776.538	22776.477	22791.182
GLL-AH Model	22822.921	22822.861	22837.565

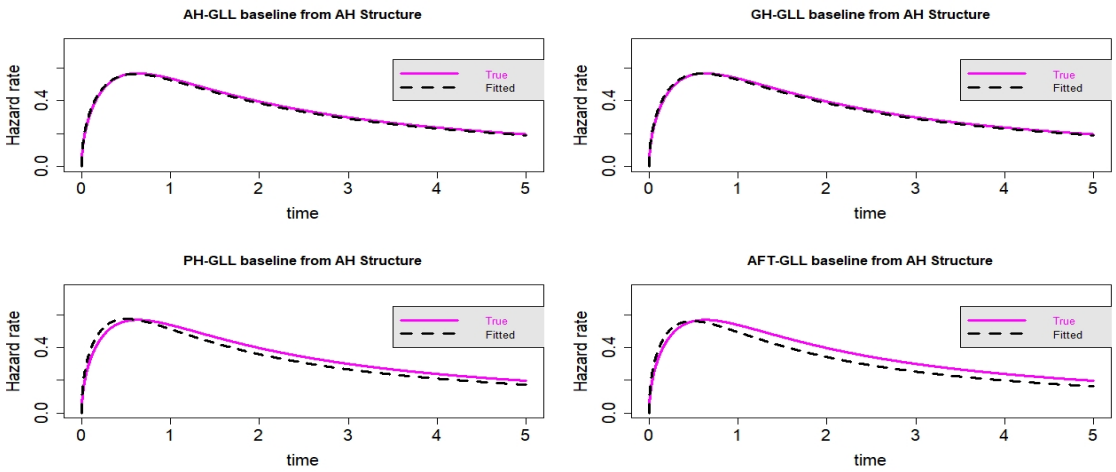

Figure 7.15: Estimated baseline hazard functions with censoring proportion of 40% and a sample size of $n = 10000$. The dashed and solid curves indicate the estimated and true hazard rate functions, accordingly. The data generated from an AH structure

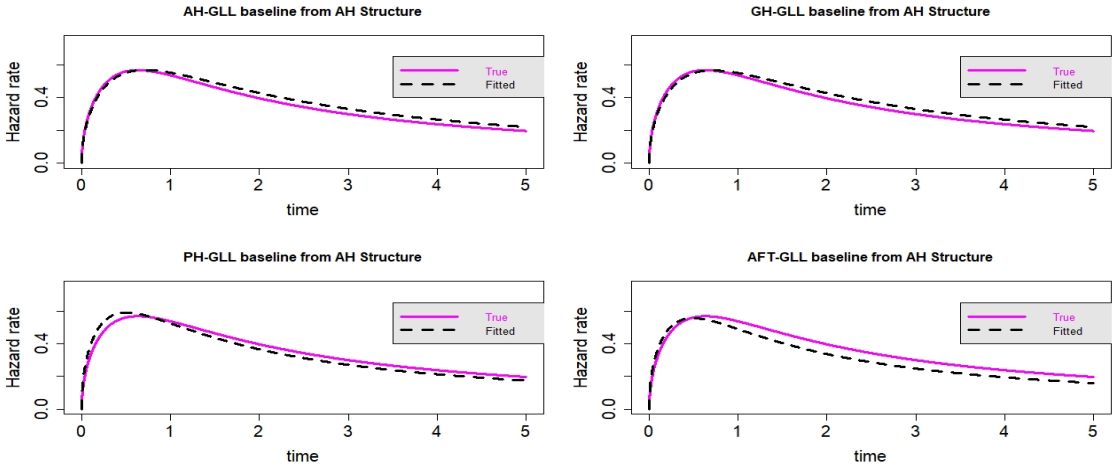

Figure 7.16: Estimated baseline hazard functions with censoring proportion of 40% and a sample size of $n = 10000$. The dashed and solid curves indicate the estimated and true hazard rate functions, accordingly. The data generated from an AH structure

From the data in Tables 7.2 - 7.7, the GH model may have a better performance, but it is hard to say that improvements for the GH model is significant. The GH model, as expected, has a nested structure with a versatile closed-form expression, making it more suitable for censored lifetime data analysis. Finally, the simulation results noted that the GH model has the ability to be a very valuable tractable parametric

hazard-based model for sufficiently describing various types of lifetime data from various failure rate shapes and censoring percent-ages, as well as a numerical tool for making comparisons between the three different approaches for hazard-based models, namely, the AH, PH, and AFT structures.

7.7.2 Simulation Study II: Performance Study

The Bayesian methodology of the proposed model is addressed in Simulation Study II. The aim of this analysis is to illustrate the Bayesian inferential properties of the estimators in the proposed model. In particular, we show how sample size and censoring percentages affect the proposed model's Bayesian inferential properties.

7.7.2.1 Measures of Performance

The posterior mean, absolute bias (AB), mean square error (MSE), effective number of different simulations draws (n_{eff}), coverage probability (CP), and potential scale reduction factor ($\widehat{R}$) were used to evaluate the Bayesian inferential features of the proposed GH model.

The estimators' bias is determined as follows:

$$\text{Bias}(\hat{\theta}) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i - \theta) \tag{7.57}$$

The MSE is a useful indication of overall correctness and is computed as:

$$\text{MSE}(\hat{\theta}) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i - \theta)^2 \tag{7.58}$$

where $\theta = \alpha, \kappa, \eta, \beta^{'}$.

The following is a description of CP :

$$\text{CP} = \hat{\theta} \mp 1.96 \times \text{SE}(\hat{\theta}) \tag{7.59}$$

According to Gelman et al. Gelman et al. (1995), the effective number of sample size simulation draws should be more than or equal to 400 in order to verify the convergence diagnostics of McMC simulations. The maximum permitted limit of $(\widehat{R})$ should also be close to $1(\widehat{R} < 1.10)$.

7.7.2.2 Posterior Analysis of Simulation Study II

In the simulation sets, we incorporated the proposed parametric GH model with the GLL baseline distribution to examine its Bayesian inferential characteristics. Each simulation set was utilized to estimate the suggested GH model with various censoring rates and sample sizes. Three parallel chains with 60,000 iterations each plus additional 6,000 for the burn-in time were utilized to approximate posterior distributions using JAGS software Denwood (2016). The chains were shortened further by storing every 10^{th} draw to reduce auto-correlation in the sequences.

7.7.2.3 Simulation Results of Simulation Study II

Based on these findings reported in Tables 7.10 and 7.11, we can infer that the estimators' biases and MSE decrease with sample size, and that the estimators' bias and MSE are also influenced by the censoring %, with greater censoring proportions increasing MSE and absolute bias (AB). The Gelman-Rubin diagnostic (potential scale reduction factor) and the number of efficiency sample size draws, on the other hand, illustrate that convergence has been reached. The estimators' coverage probability was close to 95%.

Table 7.10: Results of the McMC Simulation for Study II (Bayesian inference). GLL-GH framework with baseline hazard parameter values of $(\alpha = 1.40, k = 0.80,$ **and** $\eta = 1.20)$**; covariate values of** $\beta_1 = (0.25, 0.35, 0.45), \beta_2 = (0.55, 0.65, 0.75)$**; sample size of 100 ; and two censoring proportions for rates of 20 and** 40%

True Value ($\hat{\theta}$)	Posterior Mean	AB	MSE	CP	n-eff	$\hat{R}$
			20% Censoring			
$\beta_{11} = 0.25$	0.260	0.010	0.005	94.13	5232	1.002
$\beta_{12} = 0.35$	0.378	0.028	0.014	93.96	4313	1.002
$\beta_{13} = 0.45$	0.475	0.025	0.010	94.60	4717	1.000
$\beta_{21} = 0.55$	0.595	0.045	0.015	93.88	3987	1.000
$\beta_{22} = 0.65$	0.714	0.064	0.026	92.37	3243	1.000
$\beta_{23} = 0.75$	0.822	0.072	0.032	95.03	3456	1.000
$\alpha = 1.40$	1.425	0.025	0.011	94.50	5404	1.001
$\kappa = 0.80$	0.833	0.033	0.025	94.23	5039	1.002
$\eta = 1.20$	1.224	0.024	0.021	95.67	4788	1.002
			30% Censoring			
$\beta_{11} = 0.25$	0.292	0.042	0.007	92.47	5032	1.003
$\beta_{12} = 0.35$	0.386	0.036	0.015	92.55	4519	1.004
$\beta_{13} = 0.45$	0.483	0.033	0.012	93.41	4788	1.001
$\beta_{21} = 0.55$	0.606	0.056	0.018	94.69	3987	1.001
$\beta_{22} = 0.65$	0.720	0.070	0.023	96.82	3832	1.001
$\beta_{23} = 0.75$	0.831	0.081	0.033	94.45	4156	1.000
$\alpha = 1.40$	1.436	0.036	0.013	96.32	5122	1.002
$\kappa = 0.80$	0.847	0.047	0.028	91.60	5039	1.001
$\eta = 1.20$	1.232	0.032	0.024	93.09	5188	1.001

Table 7.11: Results of the McMC Simulation for Study II (Bayesian inference). GLL-GH framework with baseline hazard parameter values of $(\alpha = 1.40, k = 0.80,$ **and** $\eta = 1.20)$**; covariate values of** $\beta_1 = (0.25, 0.35, 0.45), \beta_2 = (0.55, 0.65, 0.75)$**; sample size of 300 ; and two censoring proportions for rates of 20 and** 40%

True Value ($\hat{\theta}$)	Posterior Mean	AB	MSE	CP	n-eff	$\hat{R}$
			20% Censoring			
$\beta_{11} = 0.25$	0.258	0.008	0.004	95.14	4898	1.000
$\beta_{12} = 0.35$	0.368	0.018	0.012	95.03	4020	1.000
$\beta_{13} = 0.45$	0.470	0.020	0.008	94.93	5676	1.000
$\beta_{21} = 0.55$	0.590	0.040	0.013	95.80	5413	1.000
$\beta_{22} = 0.65$	0.704	0.054	0.024	94.80	5213	1.000
$\beta_{23} = 0.75$	0.802	0.052	0.030	95.04	5093	1.000
$\alpha = 1.40$	1.425	0.025	0.010	95.12	5003	1.000
$\kappa = 0.80$	0.803	0.003	0.002	95.07	4937	1.000
$\eta = 1.20$	1.204	0.004	0.003	94.92	4953	1.000
			30% Censoring			
$\beta_{11} = 0.25$	0.292	0.042	0.006	94.00	5099	1.000
$\beta_{12} = 0.35$	0.386	0.036	0.014	95.25	5676	1.001
$\beta_{13} = 0.45$	0.483	0.033	0.011	94.763	55122	1.000
$\beta_{21} = 0.55$	0.606	0.321	0.013	95.43	3056	1.002
$\beta_{22} = 0.65$	0.720	0.70	0.020	95.55	3898	1.001
$\beta_{23} = 0.75$	0.831	0.081	0.031	94.34	4454	1.000
$\alpha = 1.40$	1.436	0.036	0.012	96.32	4989	1.001
$\kappa = 0.80$	0.847	0.047	0.026	95.00	5006	1.000
$\eta = 1.20$	1.232	0.032	0.020	96.05	5012	1.001

7.8 Application

The most commonly used type of censored data in oncology studies is right-censored survival data. In these analyses, the time-to-event is commonly the time between survival and death. This section focuses on the use of parametric hazard-based regression models to reanalyze two real-world right-censored oncology data sets that have previously been addressed in the literature. The purpose of this study is to

compare the parametric general hazard (GH) regression model to its special cases, which include the PH, AFT, and AH frameworks, with the generalized log-logistic baseline. In the first of the two data sets, there are crossing survival curves, but there are no crossover survival curves in the second.

7.8.1 Colon Cancer Data Set

7.8.1.1 Data Description

In this section, we take a look at a genuine survival time data set for people with colon cancer that is openly accessible using the R package survival under the label of colon (Therneau and Lumley, 2013). Initially, Laurie et al. (1989) described the study. Moertel et al. (1990) contains the main report. The final Moertel report's data set and this one are most similar (Moertel et al., 1995). Lin's paper (Lin, 1994) made use of a version of the data with fewer follow-up times. This colon cancer data set has gained widespread use in the literature on survival analysis, and it is particularly simple to locate in research involving parametric hazard-based regression models.

This clinical trial's experiment involved 1858 patients. These findings come from one of the earliest trials of adjuvant treatment for colon cancer that was successful. Levamisole is a low-toxicity drug formerly utilized to treat worm infestations in animals, while 5-FU is a chemotherapy drug that is moderately toxic (as these things go). Each individual has two records: one for recurrence and one for death. The following variables were taken into account for each patient ($i = 1, \ldots, 1858$) :

i. t_i : time until event or censoring

ii. status: censoring status (1 = observed lifetime, 0 = censored)

iii. age: age of the patient in years

iv. surg: time from surgery to registration (1=long, 0 = short)

v. etype: type of event (1 = recurrence, 2 = death)

The total number of patients whose surgery takes a long time are 494 patients (26.59%), among whom 247(50%) died. The Kaplan-Meier plot for the surgery variable is reported in Figure 7.18 .

7.8.1.2 Visual Representation of the Data

Analysis: The non-parametric plots for the survival time of colon cancer patients are reported in Figure 7.17. TTT plot for the survival time indicates an increasing hazard rate pattern.

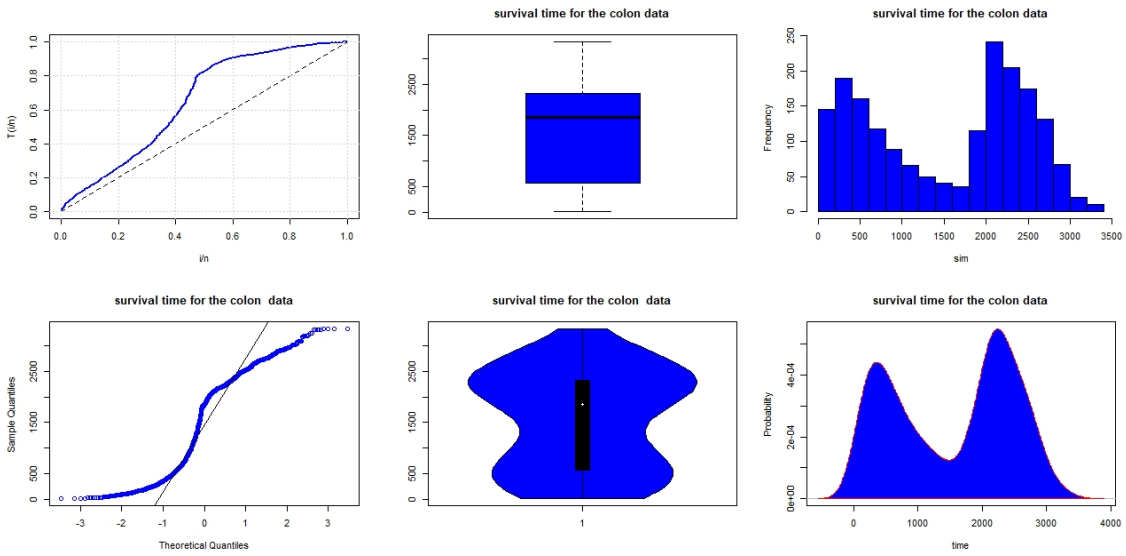

Figure 7.17: Nonparametric plots for the survival time data of colon cancer patients

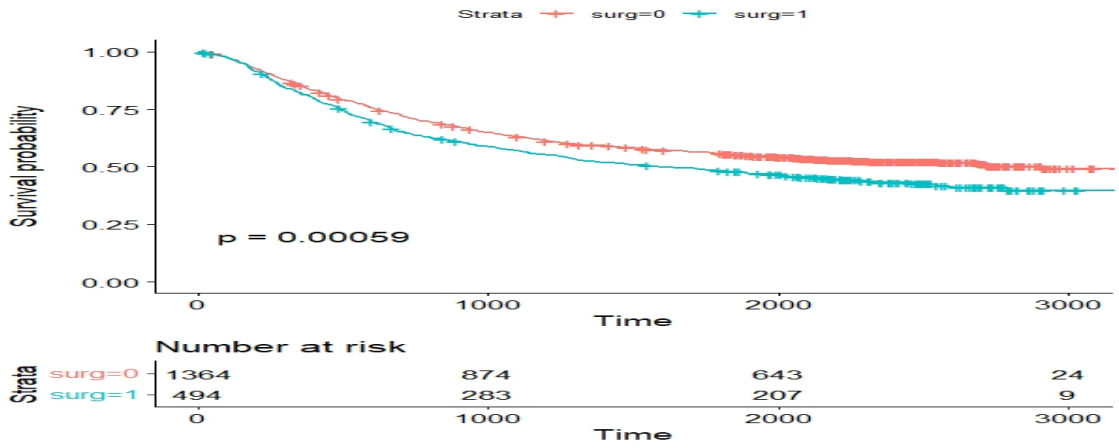

Figure 7.18: Kaplan-Meier survival plot for the variable surgery status(1=long, 0=short).

7.8.2 Classical Analysis

For the proposed GH model and its sub-models, including PH, AH, and AFT with GLL baseline distribution, the MLE estimates for baseline distribution parameters and regression coefficients are provided in Table 7.12

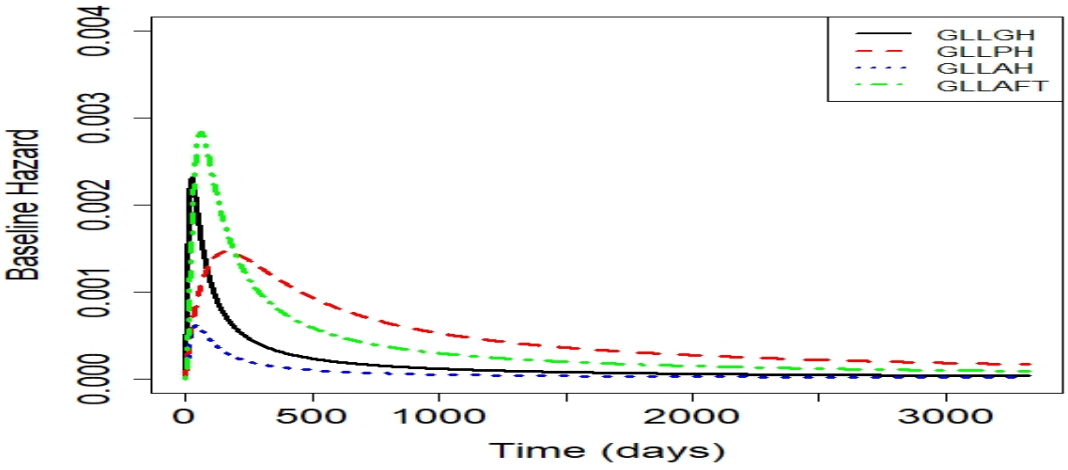

Figure 7.19: Estimated Hazards for the competitive models of the colon cancer data set.

Table 7.12: Results from the fitted parametric hazard-based regression models to Colon cancer data se

Models	Parameter(s)	Estimate	SE	Z-value	P-value	L-95%	U-95%
GLL-GH	β_{11}	0.065	0.038	0.593	0.553	-0.053	0.171
	β_{12}	-0.171	0.129	-1.331	0.183	-0.423	0.081
	β_{13}	-1.549	0.118	-13.185	<0.001	-1.780	-1.139
	β_{21}	0.012	0.039	0.316	0.752	-0.065	0.090
	β_{22}	0.161	0.087	1.850	0.064	-0.010	0.331
	β_{23}	-0.890	0.072	-12.339	<0.001	-1.032	-0.749
	α	1.938	0.127	8.932	<0.001	1.690	2.187
	κ	0.009	0.001	15.290	<0.001	0.007	0.011
	η	0.039	0.006	6.697	<0.001	0.028	0.187
GLL-AFT	β_1	0.023	0.038	0.593	0.553	-0.053	0.098
	β_2	0.100	0.089	1.131	0.258	-0.074	0.274
	β_3	-0.970	0.058	-16.864	<0.001	-1.083	-0.857
	α	2.381	0.139	17.175	<0.001	2.161	2.653
	κ	0.008	0.001	10.986	<0.001	0.011	0.009
	η	0.019	0.002	11.213	¡0.001	0.042	0.022
GLL-PH	β_1	-0.064	0.033	-1.936	0.043	-0.129	0.001
	β_2	0.167	0.072	2.323	0.020	0.026	0.309
	β_3	-0.518	0.052	-9.978	<0.001	-0.620	-0.417
	α	1.938	0.127	8.932	<0.001	1.690	2.187
	κ	0.009	0.001	15.290	<0.001	0.007	0.011
	η	0.039	0.006	6.697	<0.001	0.028	0.187
GLL-AH	β_1	0.029	0.047	0.620	0.535	-0.062	0.120
	β_2	0.-464	0.108	-4.288	<0.001	-0.676	-0.252
	β_3	-1.139	0.090	-12.648	<0.001	-1.315	-0.962
	α	2.113	0.084	25.226	<0.001	2.161	2.277
	κ	0.004	0.000	17.158	<0.001	0.011	0.005
	η	0.024	0.003	9.495	<0.001	0.042	0.029

7.8.3 Frequentist Model Comparison

We take into account the Akaike information criteria (AIC), the Hannan-Quin information criterion (HQIC), and the Corrected Akaike information criterion (CAIC)

when comparing frequentist models. The proposed GH model is the most effective model when compared to its rival models, according to the estimates of the AIC, CAIC, and HQIC in Table 7.13.

Table 7.13: Results for some frequentist inofrmation criterions for the hazard-based regression models

Model	**AIC**	**CAIC**	**HQIC**
GH	16276.09	16276.07	16294.43
PH	16415.50	16415.50	16427.74
AH	16384.96	16384.94	16397.94
AFT	16294.36	16294.35	16306.58

7.8.3.1 Likelihood Ratio Test

To form a comprehensive statistical inference about a model, it is necessary to lower the number of parameters and assess how this impacts the model's ability to match the data. The likelihood ratio test (LRT) is used to compare the GH model to its sub-models, which include the PH, AFT, and AH hazard-based regression models. The LRT statistic and its accompanying P-values in Table 7.14 show that the GH model fits better than its sub-models for the colon cancer lifetime data set.

Table 7.14: LRT test for the GH model and its sub-models

Model	Hypothesis	LRT	P-value
GH vs. PH	H_0: β_2=0, $H_1 : H_0$ is false,	145.424	< 0.001
GH vs. AH	H_0: β_1=0, $H_1 : H_0$ is false,	114.863	< 0.001
GH vs. AFT	H_0: β_1=β_2,$H_1 : H_0$ is false ,	24.268	< 0.001

7.8.4 Bayesian Analysis

We used Bayesian analysis to compare the proposed GLL-GH model with its competing models, such as the GLL-PH, GLL-AH, and GLL-AFT models. The baseline

Table 7.15: Results for the posterior properties of the competitive models.

Models	Par (s)	Estimate	SE	SD	2.5%	Medium	97.5%	N_{eff}	$\hat{R}$
GLL-GH	β_{11}	0.086	0.001	0.	-0.016	0.087	0.187	4013	1.001
	β_{12}	-0.172	0.002	0.122	-0.412	-0.174	0.067	4049	1.001
	β_{13}	-1.500	0.003	0.129	-1.752	-1.497	-1.264	2174	1.004
	β_{21}	0.009	0.001	0.040	-0.070	0.009	0.086	4772	1.000
	β_{22}	0.161	0.001	0.087	-0.004	0.158	0.333	4151	1.001
	β_{23}	-0.925	0.002	0.088	-1.097	-0.925	-0.756	2410	1.003
	α	2.166	0.002	0.138	1.909	2.161	2.449	4951	1.001
	κ	0.011	0.000	0.002	0.008	0.011	0.016	2174	1.004
	η	0.044	0.000	0.009	0.029	0.042	0.064	2016	1.004
GLL-AFT	β_1	0.021	0.001	0.038	-0.057	0.021	0.095	5059	1.001
	β_2	0.101	0.001	0.089	-0.073	0.100	0.271	5069	1.000
	β_3	-1.012	0.002	0.081	-1.168	-1.012	-0.855	2685	1.001
	α	2.282	0.002	0.134	2.033	2.277	2.554	4304	1.000
	κ	0.008	0.000	0.001	0.006	0.008	0.011	2426	1.001
	η	0.020	0.000	0.003	0.015	0.020	0.027	2470	1.001
GLL-PH	β_1	-0.029	0.000	0.034	-0.096	-0.029	0.039	5944	1.000
	β_2	0.238	0.001	0.071	0.099	0.238	0.374	5490	1.000
	β_3	-0.245	0.001	0.066	-0.374	-0.245	-0.116	4336	1.000
	α	1.997	0.002	0.000	-1.767	1.993	2.242	3623	1.001
	κ	0.002	0.000	0.122	0.002	0.002	0.002	2912	1.002
	η	0.004	0.000	0.000	0.004	0.004	0.005	3235	1.001
GLL-AH	β_1	0.071	0.001	0.043	-0.013	0.072	0.157	5046	1.000
	β_2	-0.283	0.001	0.098	-0.479	-0.282	-0.094	4320	1.000
	β_3	-0.744	0.002	0.096	-0.932	-0.744	-0.554	3425	1.001
	α	2.218	0.002	0.132	1.970	2.216	2.487	3258	1.000
	κ	0.003	0.000	0.000	0.003	0.003	0.004	2728	1.001
	η	0.014	0.000	0.002	0.010	0.014	0.018	2904	1.001

distribution parameters $\alpha \sim G(a_1, b_1), \eta \sim G(a_2, b_2)$, and $k \sim G(a_3, b_3)$ with hyper-parameter values $(a_1 = b_1 = a_2 = b_2 = a_3 = b_3 = 10)$ are assumed to have separate gamma priors that are independent and non-informative normal prior with a value of $N(0, 100)$ for $\beta's$ (regression coefficients). Rstan package was utilized for our analysis (Carpenter et al., 2017).

7.8.4.1 Numerical Summary

In this section, we used the McMC sample of posterior properties for the generalized log-logistic general hazard (GLL-GH) model and its special cases including GLL-PH, GLL-AH, and GLL-AFT Table 7.15 to examine several posterior properties of interest and their numerical values.

7.8.4.2 Visual Summary

Figures 7.20 to 7.24 provide the trace and autocorrelation (AC) plots for the baseline distribution parameters and regression coefficients of the proposed GH model and its sub-models, indicating convergence of the chains.

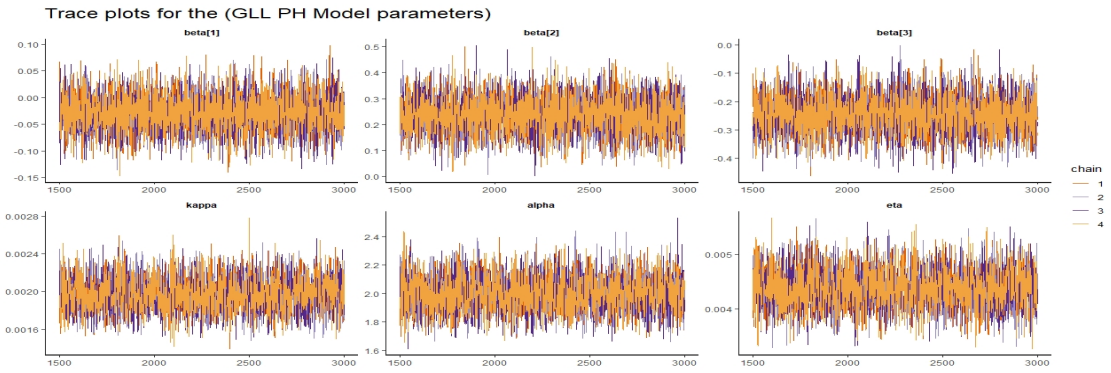

Figure 7.20: Trace plots for the GLL-PH model parameters

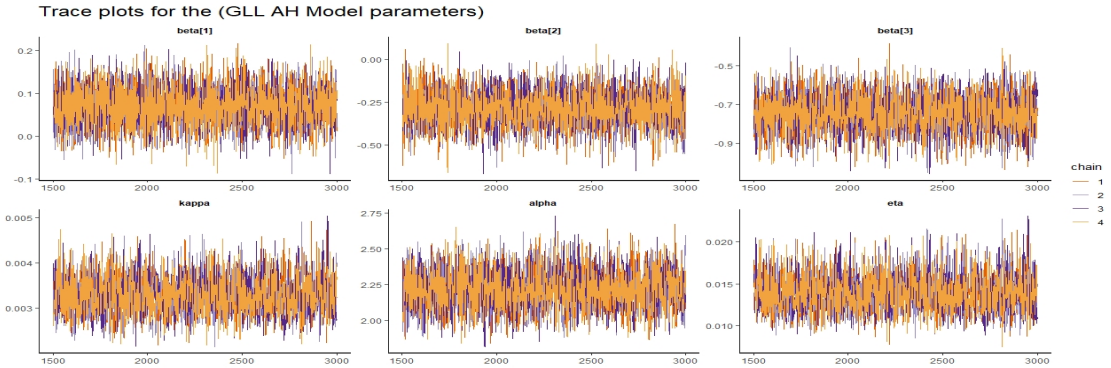

Figure 7.21: Trace plots for the GLL-AH model parameters

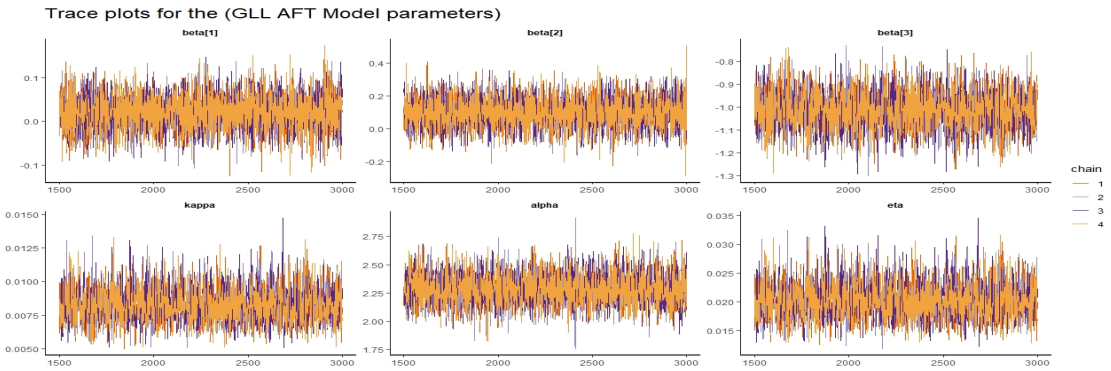

Figure 7.22: Auto-correlation plots for the GLL-AFT model parameters

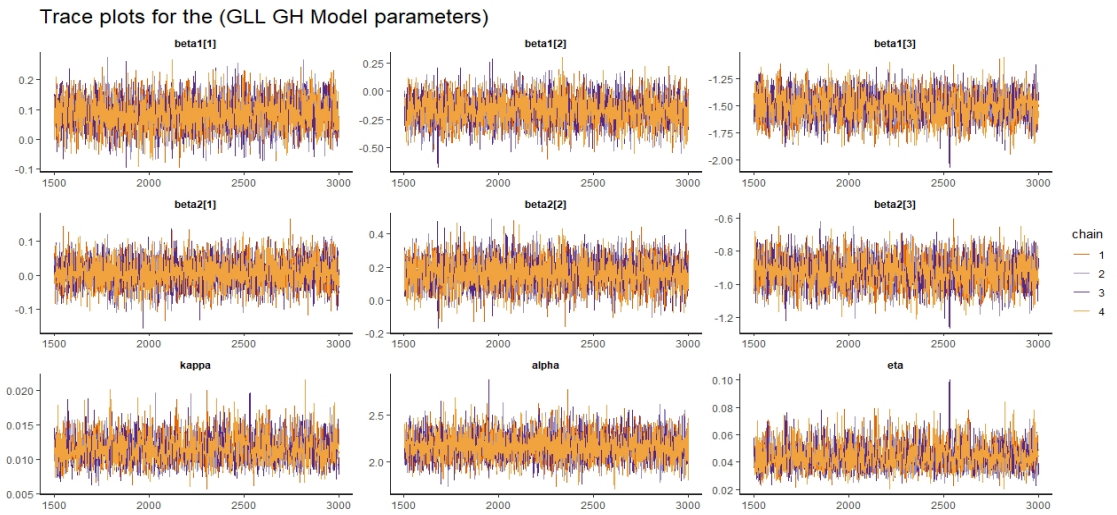

Figure 7.23: Auto-correlation plots for the GLL-GH model parameters

7.8.4.3 McMC Convergence Diagnostics

We applied both numerical and visual methods to evaluate the convergence of the McMC algorithm for the proposed models and their special cases. As can be seen from the summary results in the above table, the McMC algorithm HMC-NUTS has converged to the joint posterior distribution because the potential scale reduction factor $\hat{R}$ is 1 , the effective sample size (n_{eff}) is greater than 400, and the Monte Carlo error (SE) is less than 5% of the posterior standard deviations for all of the parameters.

Visually assessing convergence is often done using auto-correlation and trace graphs (Ashraf-Ul-Alam and Khan, 2021). Figures 7.20 to 7.23 trace plot displays a stationary pattern fluctuating within a band, demonstrating the convergence of the McMC algorithm. Figure **??** auto-correlation plot demonstrates how auto-correlation rapidly decreases to zero as the period of lag increases, indicating good mixing and the convergence of the algorithm to the desired posterior distribution.

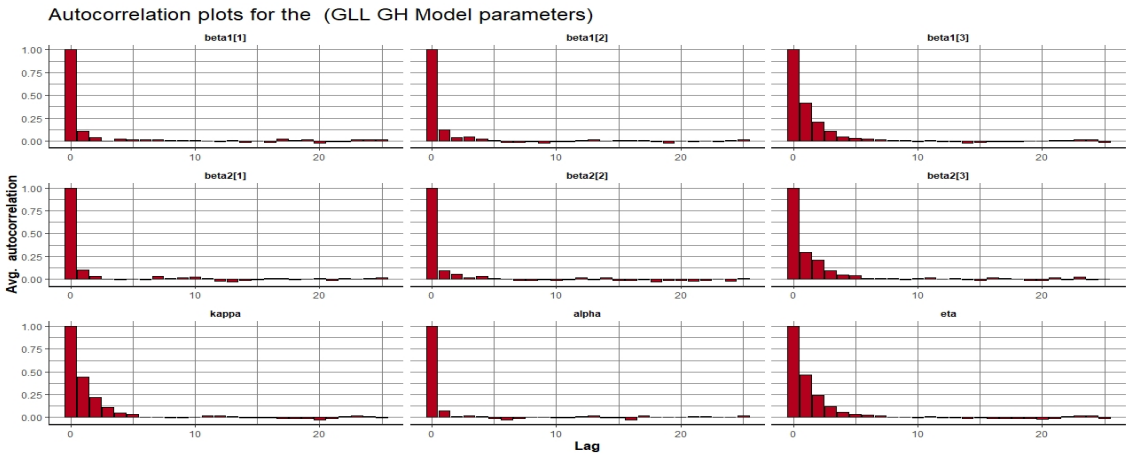

Figure 7.24: Auto-correlation plots for the GLL-GH model parameters

7.8.4.4 Bayesian Model Selection

We implemented two information criteria, Watanabe Akaike information criterion, proposed by (Watanabe, 2013), for the Bayesian model comparison, and the Vehtari et al. (2017) proposed Leave-one-out information criteria. A model may be said to be best suited if it has the lowest WAIC and LOOIC values for both information criteria. In addition to Stan fitting, posterior predictive check (PPC) and determining WAIC and LOOIC are performed using the R package loo (Vehtari et al., 2021). Table 7.16 below shows that when compared to its rival models, the GLL-GH model is the most effective.

Table 7.16: Bayesian model comparison for the GLL-GH and its special cases

Model	**WAIC**	**LOOIC**
GH	16274.00	16274.01
PH	16360.20	16360.21
AH	16345.80	16345.90
AFT	16295.75	16295.80

7.9 Conclusions

The PH, AFT, and AH models are three ways to develop hazard-based regression models for survival data. Because of the relative risk interpretation of the regression coefficients and the existence of a semi-parametric PH model that is robust against the distributional assumption of the survival time, PH models are particularly popular in clinical trials and oncology investigations. The goal of this study was to generalize the hazard-based regression models stated above to include both time-independent and time-dependent covariates in a single model, dubbed the GH model. The main goal of this model is to distinguish scenarios in which covariates have a time-independent or time-dependent effect on the hazard rate when modelling survival data, with a focus on parametric models.

In general, if the underlying distributional assumption is relatively true, a parametric model is chosen in statistical data analysis. In survival data analysis, parametric hazard-based regression models can provide more accurate estimations of the regression coefficients than semi-parametric hazard-based regression models (Collet, 2015). Other essential values, such as quantiles, the hazard function, and survival probabilities, can also be easily estimated using parametric models. It's worth noting that the hazard function is a key part of the time course of a disease process, therefore it's a focus of many clinical investigations.

The Cox PH model does practically all of the modelling of censored survival data. The non-proportional hazard models, such as the AFT and AH models, are chosen as an alternative once the proportionality assumption is discarded. On the other hand, AFT and AH models can only include covariates with time-dependent and non-proportional effects on the hazard overtime, whereas a PH model can only have time-independent and proportional effects. The GH model established in this work can accept a variety of covariates, some of which may have time-independent and

proportional impacts on the hazard value, while others may have time-dependent and non-proportional effects. Another advantage of the model is that it may be used to predict when survival and hazard rates will cross.

A detailed simulation study was conducted to evaluate the performance of the suggested GH model. The findings show that the GH model produces better outcomes, with fewer biases detected for the majority of parameters. The layered structure of the GH model in comparison to the PH, AFT, and AH models for a broad regression setting containing various covariates prevalent in cancer epidemiology studies was further explored using simulated data sets. The results demonstrate the GH model's nested structure and tractability once more. Following the simulation study, this paper shows a real-world data application with right-censored cancer data sets from a patient clinical trial. When the information criterion used in this study was evaluated, the GLL-GH model outperformed the GLL-PH, GLL-AH, and GLL-AFT models.

CHAPTER 8

Amoud Class for Hazard-based and Odds-based Regression Models: An Application to Oncology Studies

In this chapter, we present our seventh manuscript [1] about Amoud Class for hazard-based and odds-based regression models. Note that the materials of this chapter have been reproduced from our article which is under review.

8.1 Introduction

During the last few decades, the semi-parametric Cox proportional hazard (PH) model has dominated survival data analysis. While Cox's original research paper discussed extensions to remove the assumption of PH (Cox, 1972), much work has been carried out to improve the flexibility of survival regression frameworks by using tractable functions for both the baseline and the inclusion of covariates, primarily using probability distributions, splines, or fractional polynomials (Rubio et al., 2019). As a matter of fact, the hazard rate and odds functions are two probabilistic functions with significant practical value in survival analysis. They both take into account the hazard rate or odds for a reference level associated with a link function of the covariates, which is often represented by a log-linear or a multiplicative term $\exp(x_i'\beta)$. Each covariate's associated parameters are represented by the vector β. Given a design matrix X and a subject $i, i \in \{1, \ldots, n\}$, the vector x_i, represents covariate values. The subject i with all of its covariate values equal to zero ($x_i = 0$)

[1] Muse, A. H., Ngesa, O., & Mwalili, S., (2022). Amoud Class for Hazard-based and Odds-based Regression Models: An Application to Oncology Studies. *Axioms*, MDPI, 2022

represents the reference level.

So, based upon the type of probabilistic function utilized as a baseline function, the survival regression model classes can be divided into two primary groups: hazard-based regression models and odds-based regression models. However, the general design of those models did not change; the hazard or odds function was expressed as a baseline function multiplied by the link function of the covariates to either the baseline function, the time scale, or both of them.

Because of the most well-known Cox PH model, hazard-based regression models are the most prevalent survival regression model classes in the field of survival analysis. Consequently, there are four widely used hazard-based regression models: proportional hazard (PH) (Cox, 1972; Kalbfleisch, 1978), accelerated hazard (AH) (Chen and Wang, 2000), accelerated failure time (AFT) (Buckley and James, 1979; Komárek and Lesaffre, 2008), and general hazard (GH) (Ciampi and Etezadi-Amoli, 1985; Etezadi-Amoli and Ciampi, 1987; Louzada-Neto, 1997; Chen and Jewell, 2001).

The odds-based regression models, which are created using a probabilistic function that has recently received more attention and is known as the odds function, are another family of survival regression model classes. Although the odds function is used in epidemiological case-control research, the proportional odds (PO) model class that was presented by Bennett (1983a) is the first to apply it in survival models. AFT model is another odds-based regression model (Kalbfleisch, 1978). As a result, just like hazard-based regression models, odds-based models are divided into four primary categories: PO (Bennett, 1983a), accelerated odds (AO), AFT (Buckley and James, 1979), and general odds (GO) models.

There are other survival regression models as well, which combine hazard-based and odds-based regression models and are built by taking into account both hazard rate and odds functions. For instance, Yang and Prentice (2005) developed the Yang-Prentice model, a semi-parametric survival regression model that can include crossover survival curves. In order to describe survival data with crossed survival

curves, Demarqui and Mayrink (2021) modified the Yang-Prentice (YP) model using a piecewise exponential baseline distribution. Both the PH and PO models are included as sub-models in the YP model. A generalized odds-rate hazards model was developed by Banerjee et al. (2007) and includes the PH, PO, and AFT models as special cases.

For censored lifetime data, Royston and Parmar (2002) presented a flexible parametric model based on the PH and PO models. On the other hand, Huang et al. (2008) also introduced a general class of regression model called the PH-PO model, which includes PH and PO models as sub-models. Huang and Jiang (2009) proposed an extension of the PH-PO model into a more generalized model that takes into account time scale changing effects and time varying coefficient effects. A semi-parametric super model containing six popular survival regression models, including the PH, PO, AFT, AH, YP, and GH models, was recently proposed by Zhang et al. (2019). Davis (2018) has recommended the development of further new families that combine hazard-based and odds-based regression models. For additional details, please see (Zhou and Hanson, 2015).

The absence of a general class of odds-based and hazard-based regression models that encompasses all hazard-based and odds-based regression frameworks is an issue that needs to be addressed. Each of the hazard-based and odds-based regression model classes mentioned above can capture different aspects of survival data. On the other hand, choosing which hazard-based or odds-based regression model is the most suitable and precise in reflecting the link between baseline (hazard or odds) and covariates, is an issue and an important research problem that must be addressed. To explicitly nest simpler models and to address the issue, we propose a novel, general, flexible, fully parametric class of hazard-based and odds-based regression framework named "Amoud class (AM)".

In contrast, there are three categories of survival regression model classes: non-parametric, semi-parametric, and parametric models. Compared to non-parametric

and semi-parametric methods, parametric models are more informative. They can be used to forecast survival times, hazard rates, as well as mean and median survival times in addition to computing relative effect estimates (Lawless, 2011). They can also be used to plot covariate-adjusted survival curves and forecast absolute risk over time. Semi-parametric models lack the power of parametric models when the parametric form is incorrectly stated. Additionally, they are more effective, resulting in estimates with reduced standard errors and greater accuracy (Khan, 2018; Khan and Khosa, 2016). In addition, parametric techniques use full maximum likelihood to estimate parameters. Parametric model residuals often take the form of the discrepancy between what was observed and what was expected (Collett, 2015).

Considering the discussion above, the current study proposes a fully parametric class of regression models that comprises formally nested special cases of the PH, PO, AH, AO,AFT, GH, and GO survival regression models. As a result, model selection among these models can be accomplished by conducting approximate likelihood ratio tests using the frequentist approach. To describe baseline hazard or baseline odds, a generalized log-logistic (GLL) distribution containing some of the most frequent parametric baseline distributions used in survival analysis, such as the log-logistic (LL), Bur-XII, exponential, and Weibull distributions, is employed. A right-censoring mechanism is considered, and the proposed model's parameters are evaluated using maximum likelihood estimation and Bayesian estimation techniques. A real-world right-censored survival dataset with a crossing survival curve is utilized to demonstrate how the proposed AM class can be employed.

Hence, the novelty of this research chapter is to introduce and investigate a novel, general, tractable, fully-parametric class of hazard-based and odds-based regression model for dealing with right-censored survival data with or without crossing survival curves. This is accomplished by assuming the GLL distribution in the proposed class to cope with the baseline distribution. To the author's best knowledge, no one has ever contemplated employing the parametric AM class of parametric hazard-

based and odds-based regression models in general, and with GLL baseline hazard in particular. This class is an extension of the most common hazard-based and odds-based regression frameworks in the literature. On the other hand, another area of interest that has yet to be addressed in the context of the AM class is the use of both the inferential procedures, Bayesian and frequentist approaches. As a result, the strategies are investigated utilizing the frequentist approach using the MLE method and the Bayesian approach using non-informative priors.

The structure of the chapter is as follows: Section 8.2 presents a review of the hazard-based, and odds-based regression models in the context of survival, duration, and reliability analysis. The formulation of the AM Class, its associated probabilistic functions, and sub-models of the class are discussed in Section 8.3. Section 8.4 presents the baseline distribution under examination in this study, as well as some of its special circumstances. Section 8.5 presents the estimation of the proposed class parameters using both classical and Bayesian estimation approaches. Section 8.6 shows a real-world, right-censored cancer data set with crossing survival curves. Section 8.7 finishes the study with a farewell address and recommendations for future research.

8.2 Recent Literature Review and State of Art

In this section, we review the studies completed in the framework of the hazard-based and odds-based regression models that are closely related to the proposed class in order to illustrate the state of scientific development in the context of current survival, duration, and reliability models.

8.2.1 Hazard-Based Regression Models

In general, survival datasets are highly skewed and can be censored for some subjects, possibly even the most. Standard linear regression models cannot fit them,

and they also only allow for the interpretation of regression coefficients in terms of the mean of time. However, different models can be applied to survival data to generate different interpretations. Observed times' functions rather than the observed times themselves are used for this. The hazard rate and the odds functions, in particular, are two probabilistic functions that are extremely important practically in survival analysis.

There are four major types of hazard-based regression models proposed in the literature to fit survival time data in medical investigations, namely, PH, AH, AFT, and GH models. These models can be used to analyze real-world data in domains other than medicine, such as economics, marketing, engineering, social science, criminology, and education. The modeling approach differs depending on the researcher's event of interest; the general notion is to watch time until the event occurs; however, for some subjects, the event never occurs.

The formulation and construction of four hazard-based regression models are reviewed and discussed in this section. We define the alternative structures below using the hazard rate function (hrf), odds function, survival (complementary distribution) function (sf), and cumulative (or integrated) hazard function (chf) in relation to time t and a vector of covariates x. We suppose that the vector of covariates lacks an intercept to avoid concerns about identifiability. The unknown regression coefficients are represented by the vector β.

8.2.1.1 PH Model

The semi-parametric PH model introduced by Cox (1972) is one of the most well-known hazard-based regression models in survival analysis. The hrf is multiplicatively affected by the impact of the covariates in this model. Different researchers have examined and analyzed studies relating to the parametric PH model utilizing various baseline distributions and inferential techniques. A parametric PH model, with an extended exponential geometric baseline distribution was developed and

evaluated by Rezaei et al. (2014). A parametric PH with GLL baseline distribution was also proposed by (Khan and Khosa, 2016). A modified PH model and a reversed PH model employing the Marshal-Olkin baseline distribution were examined by (Balakrishnan et al., 2018). Muse et al. (2022g) have investigated the Bayesian analysis of the PH model with a GLL baseline distribution.

The PH model's hrf, odds, sf, and the chf can be stated as follows:

$$h_{PH}(t;\beta,x) = H_0(t)e^{x'\beta}, \tag{8.1}$$

$$R_{PH}(t;\beta,x) = R_0(t)^{e^{x'\beta}}, \tag{8.2}$$

$$S_{PH}(t;\beta,x) = S_0(t)^{e^{x'\beta}}, \tag{8.3}$$

$$H_{PH}(t;\beta,x) = H_0(t)e^{x'\beta}, \tag{8.4}$$

where H_0, R_0, S_0, and H_0 are the baseline hazard rate, odds, survival and cumulative hazard functions.

8.2.1.2 AFT Model

The PH model is the most popular hazard-based regression model in survival analysis, but it can only be used in situations in which the PH assumption holds. An alternative to the PH model is the AFT model (Kalbfleisch, 1978; Buckley and James, 1979). The AFT model is analogous to a hazard-based regression model in which covariates measured on an individual are assumed to act multiplicatively on the time-scale, influencing the rate at which the individual advances along the time axis. Numerous scholars have studied and discussed studies involving the parametric AFT model using various baseline hazards and statistical inference techniques. A parametric AFT model with an exponentiated Weibull baseline distribution was presented and analyzed by (Khan, 2018). A parametric AFT with a log-exponential power baseline distribution was also proposed by (Olosunde and EJIOFOR, 2021).

Ashraf-Ul-Alam and Khan (2021) used a generalized Top-leone-Weibull baseline distribution to study a parametric AFT model. A parametric AFT model with a GLL baseline distribution was recently proposed by (Muse et al., 2022a).

The hrf, odds, sf, and chf of the AFT model are defined by:

$$h_{AFT}(t;\beta,x) = H_0\left(te^{x'\beta}\right)e^{x'\beta}, \tag{8.5}$$

$$R_{AFT}(t;\beta,x) = R_0\left(te^{x'\beta}\right), \tag{8.6}$$

$$S_{AFT}(t;\beta,x) = S_0\left(te^{x'\beta}\right), \tag{8.7}$$

$$H_{AFT}(t;\beta,x) = H_0\left(te^{x'\beta}\right). \tag{8.8}$$

8.2.1.3 AH Model

AFT and PH models have been widely applied to deal with lifetime data in different disciplines of knowledge. Despite being widely used, such hazard-based regression models are not suitable to handle survival data with crossing survival curves. Chen and Wang (2000) proposed a semi-parametric hazard-based regression model, named the AH model, allowing the analysis of crossing survival curves. In the context of a parametric AH model, different baseline hazards are available in the AHSurv package (Muse et al., 2022f).

The hrf, odds, sf, and chf of the AH model are defined by:

$$h_{AH}(t;\beta,x) = H_0\left(te^{x'\beta}\right), \tag{8.9}$$

$$R_{AH}(t;\beta,x) = R_0\left(te^{x'\beta}\right)^{e^{-x'\beta}}, \tag{8.10}$$

$$S_{AH}(t;\beta,x) = S_0\left(te^{x'\beta}\right)^{e^{-x'\beta}}, \tag{8.11}$$

$$H_{AH}(t;\beta,x) = H_0\left(te^{x'\beta}\right)e^{-x'\beta}. \tag{8.12}$$

8.2.1.4 GH Model

Ciampi and Etezadi-Amoli (1985) introduced a general model for testing the PH and the AFT hypothesis in the analysis of censored lifetime data with the presence of covariates. Then, Etezadi-Amoli and Ciampi (1987) extended their work by the application of the splines as a baseline function. Chen and Jewell (2001) introduced a general class of semi-parametric hazard-based regression models by following the same procedure to (Ciampi and Etezadi-Amoli, 1985) just by adding the AH framework.

The hrf, odds, sf, and chf of the GH model are expressed as follows:

$$h_{GH}\left(t;\beta_1,\beta_2,x\right)=H_0\left(te^{x'\beta_1}\right)e^{x'\beta_2}, \tag{8.13}$$

$$R_{GH}\left(t;\beta_1,\beta_2,x\right)=R_0\left(te^{x'\beta_1}\right)^{e^{x'(\beta_2-\beta_1)}}, \tag{8.14}$$

$$S_{GH}\left(t;\beta_1,\beta_2,x\right)=S_0\left(te^{x'\beta_1}\right)^{e^{x'(\beta_2-\beta_1)}}, \tag{8.15}$$

$$H_{GH}\left(t;\beta_1,\beta_2,x\right)=H_0\left(te^{x'\beta_1}\right)e^{x'(\beta_2-\beta_1)}, \tag{8.16}$$

where β_1 and β_2 denote the unknown regression parameters.

8.2.1.5 Special Cases of the GH Model

All of the hazard-based regression models listed above are incorporated into the GH model of hazard-based models as special cases. The GH model can be used to derive the PH,AH, and AFT, models, according to the following theorem.

Theorem 8.2.1 *Suppose $h(t;x)$ is given by Equation (8.13). Then, we have the following results:*

1. If $\beta_2=\beta_1$, then

$$h\left(t;\beta,x\right)=h_o\left(te^{x'\beta}\right)e^{x'\beta}$$

giving the ***AFT*** *model.*

2. If $\beta_1 = 0$*, then*

$$h\left(t;\beta,x\right) = h_o\left(t\right)e^{x'\beta}$$

giving the ***PH*** *model.*

3. If $\beta_2 = 0$*, then*

$$h\left(t;\beta,x\right) = h_o\left(te^{x'\beta}\right)$$

giving the ***AH*** *model.*

Proof 8.2.1 (Proof of Theorem 1) *The proof of Theorem 1 is straightforward.*

Figure 8.1 illustrates the relationship between the hazard-based regression models and their formulations using hazard rate and cumulative hazard functions.

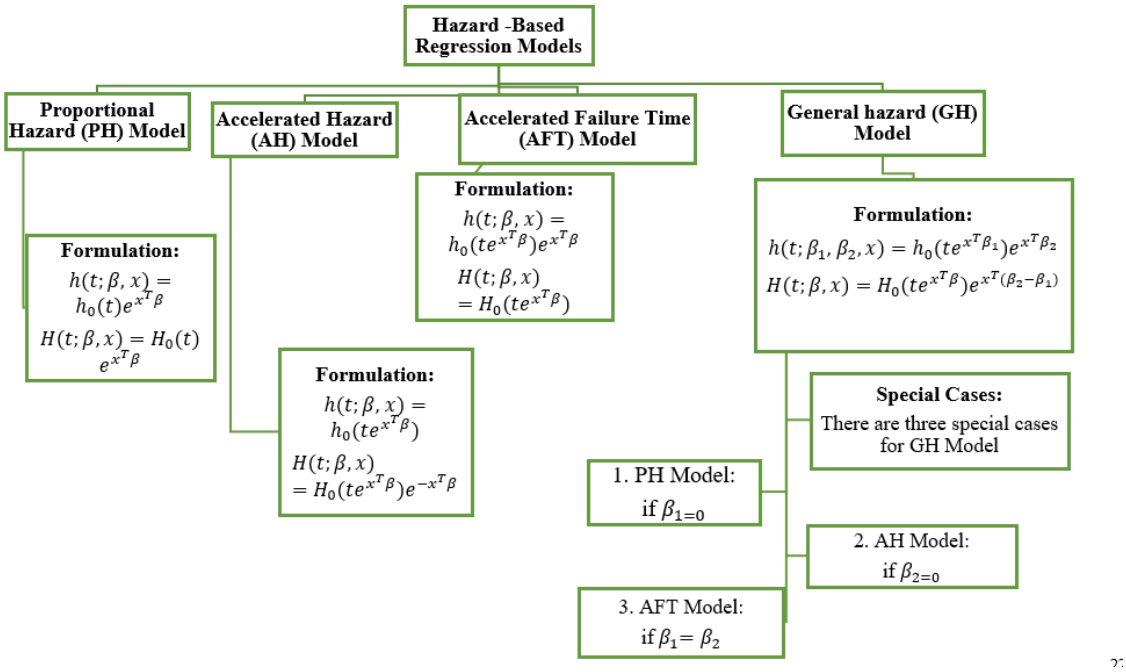

Figure 8.1: Visual graph illustrating the relationship between the hazard-based regression models and their formulation in terms of hazard function including the general hazard (GH), accelerated hazard (AH), accelerated failure time (AFT), and proportional hazard (PH) models

8.2.2 Odds-Based Regression Models

To fit survival time data in medical research, two primary types of odds-based regression models have been proposed in the literature: proportional odds and AFT models. Two more innovative odds-based regression models proposed in this study are the accelerated odds and general odds models. In fields other than medicine, such as economics, marketing, engineering, social science, criminology, and education, these models can be utilized to examine actual data.

The odds function indicates how much more likely it is that a particular event will occur for a given period t. As a result, the odds function is denoted by $R(t;\theta)$, and its mathematical expression is given by the relationship between the cumulative distribution function and its complementary (survival function):

$$R(t;\theta) = \frac{F(t;\theta)}{S(t;\theta)} = \frac{1-\exp[-H(t;\theta)]}{\exp[-H(x;\theta)]} = \exp[H(t;\theta)] - 1, \tag{8.17}$$

where $R(t;\theta), F(t;\theta), S(t;\theta)$, and $H(t;\theta)$ are the odds, cdf, survival and cumulative hazard functions respectively, and the $\theta =$ is the vector of distributional parameters. The associated derivative of the odds function is expressed as follows:

$$r(t;\theta) = \frac{dR(t;\theta)}{d(t)} = \frac{h(t;\theta)}{S(t;\theta)} = \frac{f(t;\theta)}{S(t;\theta)^2}, \tag{8.18}$$

where $r(t;\theta), h(t;\theta)$, and $f(t;\theta)$ are the derivative of odds, hrf, and pdf functions respectively.

In this section, we review two odds-based regression models that have been explored in the literature along with their formulation. On the other hand, based on the author's knowledge, we present two novel odds-based regression models that have never been used before in the literature. We define the alternative structures below with respect to time t and a vector of covariates x using the odds function $R(.)$, derivative of odds function $r(.)$, hazard function $h(.)$, and survival function $S(.)$.

We assume that the covariate vector is free of an intercept. to ease issues about identifiability. The vector β is used to represent the unknown regression coefficients.

8.2.2.1 Proportional Odds Model

The proportional odds (PO) model originally introduced by Bennett (1983a), is an odds-based regression model. According to Bennett (1983a), the PO model is structurally similar to the proportional hazards model of Cox and may be used in similar situations. Although the PO model represents an attractive alternative to the PH model.

The odds function of this model is expressed as follows:

$$R_{PO}(t;\beta,x) = R_0(t)e^{x'\beta}, \tag{8.19}$$

where $R_0(t)$ is the baseline odds function. The associated derivative of the odds function of the PO model is computed as follows:

$$r_{PO}(t;\beta,x) = r_0(t)e^{x'\beta}, \tag{8.20}$$

where $r_0(t)$ is the baseline derivative odds function.

The hrf, and sf of the PO model are computed as follows:

$$h_{PO}(t;\beta,x) = \frac{r_0(t)e^{x'\beta}}{1+R_0(t)e^{x'\beta}}, \tag{8.21}$$

In terms of the baseline hazard, it can be expressed as follows using Equation (8.18):

$$h_{PO}(t;\beta,x) = \frac{\left[\frac{h_0(t)}{S_0(t)}\right]e^{x'\beta}}{1+\left[\frac{F_0(t)}{S_0(t)}\right]e^{x'\beta}} = \frac{h_0(t)e^{x'\beta}}{F_0(t)e^{x'\beta}+S_0(t)} \tag{8.22}$$

$$S_{PO}(t;\beta,x) = \left[\frac{1}{1+R_0(t)e^{x'\beta}}\right] = \left[\frac{1}{1+\left[\frac{F_0(t)}{S_0(t)}\right]e^{x'\beta}}\right] = \left[1+\left[\frac{F_0(t)}{S_0(t)}\right]e^{x'\beta}\right]^{-1} \tag{8.23}$$

Now, we'll put forth two new models that employ methods related to the hazard-based regression models. All of the odds-based regression models in this section will be generalized as well. The model formulation put forward by Chen and Wang (2000) served as inspiration for the initial proposed approach. Their model includes accelerated hazards, but we propose a model with accelerated odds. This is how our models differ from theirs. The model formulation proposed by Chen and Jewell (2001) served as the basis for the second proposed model. The PH, AH, and AFT models are included in their model as sub-models. In contrast to their model, ours includes the PO, AFT, and AO models as sub-models. General odds model is the name of this model.

8.2.2.2 Accelerated Forms

The second parametric method of taking into account the effect of covariates, known as the accelerated form, presupposes that the covariates directly rescale time. Accelerated effects of covariates come in two varieties: Two examples of this are the:

i. Accelerated failure time (AFT) model and

ii. Accelerated odds model.

The accelerated types of the odds-based regression models can be formulated in two different ways, the first of which is similar to the AFT model. The AFT model is the only parametric survival regression framework that belongs to both the hazard-based and odds-based regression models, and both the continuous probability distributions that are closed under the hazard-based regression models and those closed under the odds-based regression models are consistent with the AFT model. For instance, the Weibull and Log-logistic distributions, we will explore these distributions in Section

8.4 of this study. Based on what the authors know, the formulation 2, give us a new survival regression model that has never been used previously.

The formulation one belongs to the AFT framework and can be expressed as follows:

$$R_{AF}(t;\beta,x) = R_0\left(te^{x'\beta}\right) \tag{8.24}$$

The associated derivative of the odds function of the AFT model is computed as follows:

$$r_{AF}(t;\beta,x) = r_0\left(te^{x'\beta}\right)e^{x'\beta} \tag{8.25}$$

The hrf and sf are expressed as follows respectively

$$\begin{aligned} h_{AF}(t;\beta,x) &= \frac{R_0'\left(te^{x'\beta}\right)}{1+R_0\left(te^{x'\beta}\right)} \\ &= \frac{\left[\frac{h_0\left(te^{x'\beta}\right)}{S_0\left(te^{x'\beta}\right)}\right]e^{x'\beta}}{1+\left[\frac{F_0\left(te^{x'\beta}\right)}{S_0\left(te^{x'\beta}\right)}\right]} \\ &= \frac{h_0\left(te^{x'\beta}\right)e^{x'\beta}}{F_0\left(te^{x'\beta}\right)+S_0\left(te^{x'\beta}\right)} \\ &= h_0\left(te^{x'\beta}\right)e^{x'\beta} \end{aligned} \tag{8.26}$$

$$S_{AF}(t;\beta,x) = \left[\frac{1}{1+R_0\left(te^{x'\beta}\right)}\right]^{-1} = \left[\frac{1}{1+\left[\frac{F_0\left(te^{x'\beta}\right)}{S_0\left(te^{x'\beta}\right)}\right]}\right]^{-1} = \left[S_0\left(te^{x'\beta}\right)\right] \tag{8.27}$$

This model as you can see after its derivation and simplification is similar to the AFT model. As a result, we can remark that the AFT model is the only one of the survival regression models that holds true for both hazard-based and odds-based regression models.

8.2.2.3 Accelerated Odds Model

A novel parametric odds-based regression model that can incorporate censored lifetime data sets with crossing survival curves is introduced here and named "accelerated odds (AO)" model. This model is formulated using the odds function and by using the same procedure to the AH Model, we got the following parametric odds-based regression model that is a new one and not done in the literature so far:

$$R_{AO}(t;\beta,x) = R_0\left(te^{x'\beta}\right)e^{-x'\beta}, \tag{8.28}$$

The associated derivative of the odds function of the AO model is computed as follows:

$$r_{AO}(t;\beta,x) = r_0\left(te^{x'\beta}\right), \tag{8.29}$$

The hrf and sf are expressed as follows:

$$h_{AO}(t;\beta,x) = \frac{R_0'\left(te^{x'\beta}\right)}{1+R_0\left(te^{x'\beta}\right)} = \frac{h_0\left(te^{x'\beta}\right)}{F_0\left(te^{x'\beta}\right)e^{-x'\beta}+S_0\left(te^{x'\beta}\right)} \tag{8.30}$$

$$S_{AO}(t;\beta,x) = \left[\frac{1}{1+e^{-x'\beta}\left[\frac{F_0\left(te^{x'\beta}\right)}{S_0\left(te^{x'\beta}\right)}\right]}\right] = \left\{1+e^{-x'\beta}\left[\frac{F_0\left(te^{x'\beta}\right)}{S_0\left(te^{x'\beta}\right)}\right]\right\}^{-1} \tag{8.31}$$

8.2.2.4 General Odds Model

Another novel general survival regression model, termed the "general odds (GO)" model, is introduced here and consists of three odds-based regression models as special cases, namely: proportional odds (PO), accelerated failure time (AFT), and accelerated odds (AO) models.

The odds function of this model can be computed as follows:

$$R_{GO}\left(t;\beta_1,\beta_2,x\right) = R_0\left(te^{x'\beta_1}\right)e^{x'(\beta_2-\beta_1)} \tag{8.32}$$

The associated derivative of the odds function of the GO model corresponding to the odds function in Equation (8.32) is computed as follows:

$$r_{GO}\left(t;\beta_1,\beta_2,x\right)=r_0\left(te^{x'\beta_1}\right)e^{x'\beta_2} \tag{8.33}$$

The hrf and sf of the GO model corresponding to Equation (8.32) are expressed as follows:

$$h_{GO}\left(t;\beta_1,\beta_2,x\right)=\frac{h_0\left(te^{x'\beta_1}\right)e^{x'\beta_2}}{e^{x'(\beta_2-\beta_1)}F_0\left(te^{x'\beta_1}\right)+S_0\left(te^{x'\beta_1}\right)} \tag{8.34}$$

$$S_{GO}\left(t;\beta_1,\beta_2,x\right)=\left[1+e^{x'(\beta_2-\beta_1)}\frac{F_o\left(te^{x'\beta_1}\right)}{S_0\left(te^{x'\beta_1}\right)}\right]^{-1} \tag{8.35}$$

In terms of the odds function, the survivor function in Equation (8.35) of the GO model can be computed as follows:

$$S_{GO}\left(t;\beta_1,\beta_2,x\right)=\left[1+R_0(t)e^{x'(\beta_2-\beta_1)}\right]^{-1} \tag{8.36}$$

8.2.2.5 Special Cases of the GO Model

All of the odds-based regression models listed above are incorporated into the GO model of odds-based models as special cases. The GO model can be used to derive the PO, AO, and AFT, models, according to the following theorem.

Theorem 8.2.2 *Suppose $r(t;x)$ is given by Equation (8.33). Then, we have the following results:*

1. If $\beta_1=\beta_2$, then

$$r\left(t;\beta,x\right)=r_o\left(te^{x'\beta}\right)e^{\beta x'}$$

*giving the **AFT** model.*

2. If $\beta_2=0$, then

$$r\left(t;\beta,x\right)=r_o\left(t\right)e^{x'\beta}$$

giving the ***PO*** *model.*

3. If $\beta_1 = 0$*, then*

$$r\left(t;\beta,x\right) = r_o\left(te^{x'\beta}\right)$$

giving the ***AO*** *model.*

Proof 8.2.2 (Proof of Theorem 2) *The proof of Theorem 2 is straightforward.*

Figure 8.2 illustrates the relationship between the odds-based regression models and their formulations using odds function and the derivative of the odds function.

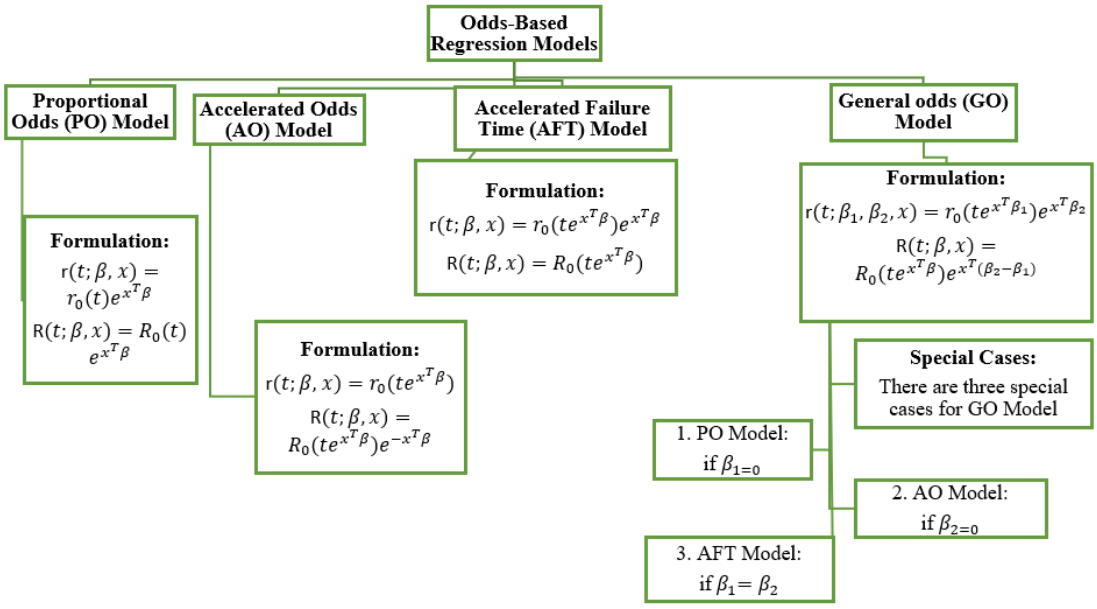

Figure 8.2: Visual graph illustrating the relationship between the odds-based regression models and their formulation in terms of odds function including the general odds (GO), accelerated failure time (AFT), accelerated odds (AO), and proportional odds (PO) models.

8.3 The Proposed Class

8.3.1 Why AM Class of Hazard-Based and Odds-Based Regression Models?

All of the hazard-based and odds-based regression models discussed in the preceding Section 8.2 can model different aspects of time-to-event data. However, determining which model is the most accurate and precise in revealing the correlation between explanatory variables and the baseline hazard (or the baseline odds) is a challenging issue and a significant research question that must be addressed.

In real life, we must decide between hazard-based regression models and odds-based regression models when provided with a dataset. A popular technique would be to fit one model to each of them, and then test the model to determine where it falls well short. However, the possibility of verifying the model assumptions may be constrained due to the finite sample size and other data characteristics. Additionally, if the right time-dependent covariates are taken into account, both the hazard-based models, such as the PH, AFT, AH, and GH models, and the odds-based models, such as the PO, AO, AFT, and GO models, may be able to fit the data relatively well.

Another issue with time-to-event data is that lifetimes can be censored in a variety of ways, including left, right, interval, double, and middle censoring, as well as survival data with crossover survival or hazard curves. Furthermore, a general class containing all of the preceding eight hazard-based and odds-based regression models is required. As a result, it is difficult to address all of the aforementioned open topics using both frequentist and Bayesian methods.

To address the aforementioned problems and to fill the gap, we introduce the AM class of hazard-based and odds-based survival regression models, a unique, novel, tractable, universal, parametric class of survival regression models that encompasses all hazard-based and odds-based regression models to help applied statisticians to

decide which model to fit in a given censored survival dataset. We estimate the model parameters using both frequentist and Bayesian approaches, and we evaluate the proposed model's nested structure using a likelihood ratio test.

In this section, we introduce the new survival regression model, its main probabilistic functions, and some special cases.

8.3.2 Model Formulation

Let T be a non-negative random variable that represents the length of time until an event of interest occurs. As already sketched, a universal class for hazard-based and odds-based regression models called the "Amoud Class (AM)" has the following closed form in order to accommodate survival data with or without the crossover of the hazard and survival curves:

$$R_{AM}\left(t;\beta_1,\beta_2,\beta_3,x\right)=e^{x'(\beta_2-\beta_1)}R_o\left(te^{x'\beta_1}\right)^{e^{x'(\beta_3-\beta_1)}}, \tag{8.37}$$

where $R_o(.)$ is the baseline odds function. This generality is attained using a structure resembling the general class of hazard-based regression models, with the addition that the baseline odds function is multiplied to a link function (i.e., log-linear function) for the covariates.

The sf for the AM model corresponding to the odds function in Equation (8.37) is expressed as follows:

$$S_{AM}\left(t;\beta_1,\beta_2,\beta_3,x\right)=\left[1+e^{x'(\beta_2-\beta_1)}\frac{F_o\left(te^{x'\beta_1}\right)}{S_0\left(te^{x'\beta_1}\right)}\right]^{-e^{x'(\beta_3-\beta_1)}}. \tag{8.38}$$

The hrf for the AM model corresponding to Equation (8.37) is computed as follows:

$$h_{AM}\left(t;\beta_1,\beta_2,\beta_3,x\right)=\frac{e^{x'(\beta_2+\beta_3-\beta_1)}H_0\left(te^{x'\beta_1}\right)}{e^{x'(\beta_2-\beta_1)}F_0\left(te^{x'\beta_1}\right)+S_0\left(te^{x'\beta_1}\right)}. \tag{8.39}$$

8.3.3 Probabilistic Functions for the Amoud Class Model

In terms of odds function, the sf for the AM model in Equation (8.38) can be expressed as follows:

$$S_{AM}\left(t;\beta_1,\beta_2,\beta_3,x\right)=\left[1+e^{x'(\beta_2-\beta_1)}R_o\left(te^{\beta_1 x'}\right)\right]^{-e^{x'(\beta_3-\beta_1)}}. \tag{8.40}$$

The derivative of the odds function for the AM model is expressed as follows:

$$r_{AM}\left(t;\beta_1,\beta_2,\beta_3,x\right)=r_o\left(te^{x'\beta_1}\right)e^{x'(\beta_2+\beta_1)}R_o\left(te^{\beta_1 x'}\right)^{e^{x'(\beta_3-\beta_1)}-1}. \tag{8.41}$$

The cdf for the AM model is computed as follows:

$$F_{AM}\left(t;\beta_1,\beta_2,\beta_3,x\right)=1-\left[1+e^{x'(\beta_2-\beta_1)}\frac{F_o\left(te^{x'\beta_1}\right)}{S_0\left(te^{x'\beta_1}\right)}\right]^{-e^{x'(\beta_3-\beta_1)}}, \tag{8.42}$$

where the baseline hazard, odds, survival, cumulative distribution, and the derivative of the odds functions are $H_0(.)$, $R_0(.)$, $S_0(.)$, $F_0(.)$, and $r_0(.)$, respectively.

8.3.4 Special Sub-Models of the Proposed Class

All of the hazard-based and odds-based regression models listed above are incorporated into the AM Class of hazard-based and odds-based survival models as special cases. The AM class can be used to derive the PH, PO, AH, AO, AFT, GH, and GO models, according to the following theorem:

Theorem 8.3.1 *Suppose $R(t;x)$ is given by Equation (8.38). Then, we have the following results:*

1. If $\beta_2=\beta_1$, then

$$R\left(t;\beta_1,\beta_2,x\right)=R_o\left(te^{x'\beta_1}\right)^{e^{x'(\beta_2-\beta_1)}}$$

*giving the **GH** model.*

2. If $\beta_3 = \beta_1$, *then*

$$R\left(t;\beta_1,\beta_2,x\right) = R_o\left(te^{x'\beta_1}\right)e^{x'(\beta_2-\beta_1)}$$

which is the ***GO*** *model.*

3. If $\beta_3 = \beta_2 = \beta_1$, *then*

$$R\left(t;\beta,x\right) = R_o\left(te^{x'\beta}\right)$$

giving the ***AFT*** *model.*

4. If $\beta_3 = \beta_1 = 0$, *then*

$$R\left(t;\beta,x\right) = R_o\left(t\right)e^{x'\beta}$$

which is the ***PO*** *model.*

5. If $\beta_2 = \beta_1 = 0$, *then*

$$R\left(t;\beta,x\right) = R_o\left(t\right)^{e^{x'\beta}}$$

giving the ***PH*** *model.*

6. If If $\beta_3 - \beta_1 = 0, \beta_2 = 0$, *then*

$$R\left(t;\beta,x\right) = R_o\left(te^{x'\beta}\right)^{e^{-x'\beta}}$$

which is the ***AO*** *model.*

7. If $\beta_2 - \beta_1 = 0, \beta_3 = 0$, *then*

$$R\left(t;\beta,x\right) = R_o\left(te^{x'\beta}\right)e^{-x'\beta}$$

giving the ***AH*** *model.*

Proof 8.3.1 (Proof of Theorem 3) *The proof of Theorem 3 is straightforward.*

Figure 8.3 illustrates the relationship between the proposed AM class and its sub-models including the GH, GO, AFT, AO, AH, PO, and PH models.

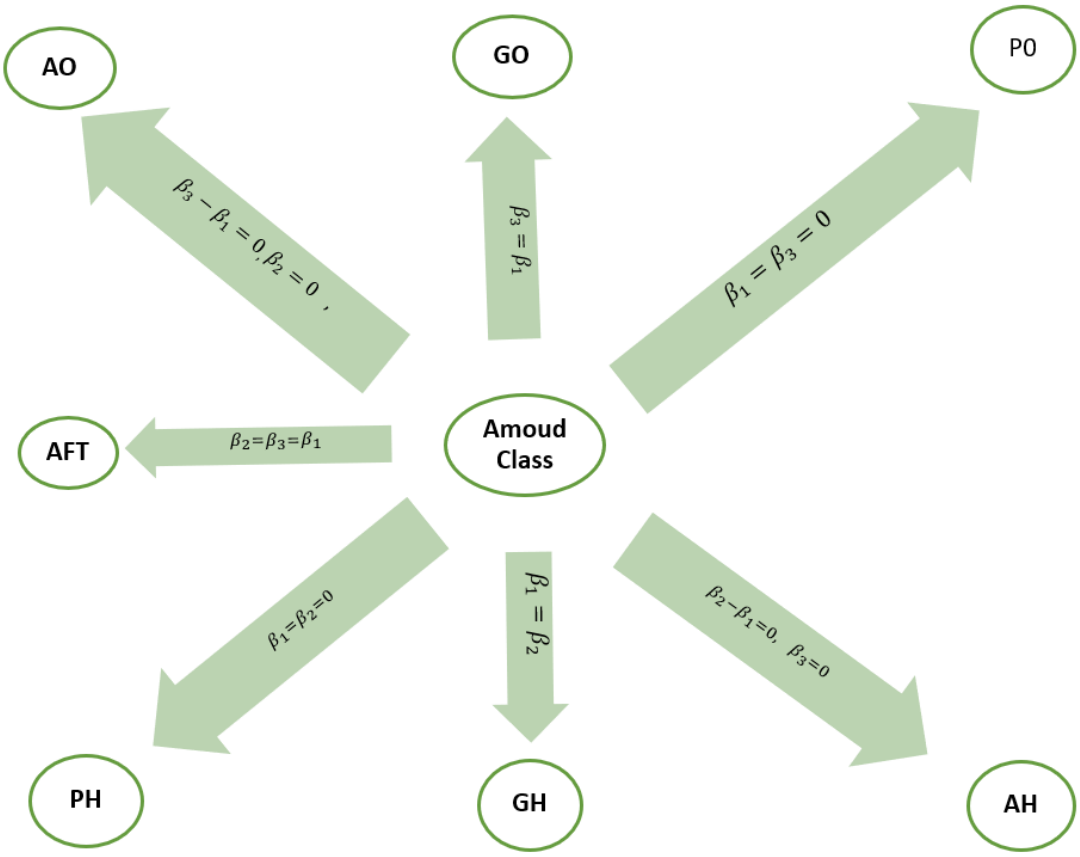

Figure 8.3: Visual graph illustrating the relationship between the proposed Amoud Class (AM) and its sub-models including the proportional hazard (PH), general odds (GO), general hazard (GH), accelerated failure time (AFT), accelerated odds (AO), and proportional odds (PO).

8.4 Baseline Distribution

The Weibull distribution, LL distribution, and a GLL distribution that combines both of them are three different baseline distributions that are presented in this section. The closeness of the Weibull distribution under hazard-based regression models and the closeness of the LL distribution under odds-based regression models were also proven. When applied to censored survival data, the closeness of the distributions is what causes the regression models to produce comparable findings. We proposed the use of a modified baseline distribution that demonstrates the differences between the survival regression models taken into consideration in this study

because the Weibull and LL distributions have limitations and give the same results under different survival regression models.

8.4.1 Weibull Baseline for Hazard-Based Regression Models

The Weibull distribution is widely used as a baseline distribution in survival and reliability regression models. Its hrf is monotone. Moreover, hrf and sf can be derived analytically, and as such, censored data can be analyzed easily. Because of the tractability and flexibility of hazard and survival functions, the Weibull model is popular among researchers in survival and reliability analysis. However, the Weibull distribution has the limitation of not capable of accommodating non-monotone unimodal and bathtub-shaped hazard functions (**?**Alkhairy et al., 2021). Another issue is that Weibull distribution is not a PO model, but this is the only distribution closed under the hazard-based regression models. This means that the PH, AH, and AFT models coincide when the baseline hrf is that of the Weibull distribution. This also means that the GH model is not identifiable.

The hrf and chf of the Weibull distribution are expressed as follows:

$$h_W(t;k,\alpha) = \alpha k(kt)^{\alpha-1}, \quad t \geq 0, \quad k,\alpha > 0, \tag{8.43}$$

$$H_W(t;k,\alpha) = (kt)^{\alpha}, \quad t \geq 0, \quad k,\alpha > 0, \tag{8.44}$$

where $k > 0$ and $\alpha > 0$ are the rate and shape parameters, respectively.

The odds function for the Weibull distribution is expressed as follows:

$$R_{LL}(t;k,\alpha) = \exp\left[H_W(t;k,\alpha)\right] - 1 = \exp((kt)^{\alpha}) - 1, \quad t \geq 0, \quad k,\alpha > 0. \tag{8.45}$$

The associated derivative of the odds function of the Weibull distribution is as

follows:

$$r_W(t;k,\alpha) = h_W(t;\theta)\exp\left[H_W(t;\theta)\right] = \alpha k(kt)^{\alpha-1}\exp((kt)^{\alpha}), \quad t \geq 0, \quad k,\alpha > 0. \tag{8.46}$$

The Weibull accelerated failure time (W-AFT) model is defined as follows:

$$h_{W-AFT}(t;\beta,x) = H_0\left(te^{x'\beta}\right)e^{x'\beta} = e^{x'\beta}\alpha k\left(kte^{x'\beta}\right)^{\alpha-1} = \left(e^{x'\beta}k\right)^{\alpha}\alpha t^{\alpha-1} = \alpha k^* t^{\alpha-1}. \tag{8.47}$$

In Equation (8.47), we can observe the hrf for the Weibull in Equation (8.43). As mentioned, the scale parameter differs between groups and we can write it as: $T_i \sim Weibull\left(k^* = \left(e^{x'\beta}k\right)^{\alpha}, \alpha\right)$, with scale k^* and shape α.
On the other hand, if a Weibull distribution is assumed for T_i under the PH framework (W-PH) in Equation (8.1), it then follows that $T_i \sim$ Weibull $\left(k^* = e^{x'\beta}k^{\alpha}, \alpha\right)$, with scale k^* and shape α. The hrf for the W-PH model is rewritten as follows:

$$h_{W-PH}(t;\beta,x) = r_0(t)e^{x'\beta} = \alpha k(kt)^{\alpha-1}e^{x'\beta} = e^{x'\beta}k^{\alpha}\alpha t^{\alpha-1} = \alpha k^* t^{\alpha-1}. \tag{8.48}$$

Technically, it is possible to compare both models in terms of the resulting scale parameter k^*. This proves that the Weibull baseline is the only baseline distribution that is closed under all hazard-based regression models.

8.4.2 Log-Logistic Baseline for Odds-Based Regression Models

The LL distribution is a frequently used baseline distribution in survival and reliability regression models. Its hrf is monotone decreasing hazard and non-monotone unimodal. The LL model hazard and density shapes are similar to those of the log-normal distribution. but it has explicit algebraic expressions for the hazard rate and survival functions which makes it more suitable for the analysis of censored lifetime

data than the log-normal distribution (Muse et al., 2021b,c). The LL distribution has the limitation of not being capable of accommodating monotone increasing and bathtub shaped hrfs. Another issue is that the LL distribution is not a PH model, but is the only distribution closed under the odds-based regression models. This means that PO, AO, and AFT models coincide when the baseline hazard is LL. This also makes the GO model not identifiable.

The cdf and sf of the LL distribution are expressed as follows:

$$F_{LL}(t;k,\alpha) = \frac{(kt)^{\alpha}}{1+(kt)^{\alpha}}, \quad t \geq 0, \quad k,\alpha > 0, \tag{8.49}$$

$$S_{LL}(t;k,\alpha) = \frac{1}{1+(kt)^{\alpha}}, \quad t \geq 0, \quad k,\alpha > 0, \tag{8.50}$$

where $k > 0$, and $\alpha > 0$ are the rate and shape parameters, respectively.

The odds function for the LL distribution is expressed as follows:

$$R_{LL}(t;k,\alpha) = \frac{F_{LL}(t;k,\alpha)}{S_{LL}(t;k,\alpha)} = \frac{\frac{(kt)^{\alpha}}{1+(kt)^{\alpha}}}{\frac{1}{1+(kt)^{\alpha}}} = (kt)^{\alpha}, \quad t \geq 0, \quad k,\alpha > 0. \tag{8.51}$$

The associated derivative of the odds function of the LL distribution is as follows:

$$r_{LL}(t;k,\alpha) = R'_{LL}(t;k,\alpha) = \alpha k(kt)^{\alpha-1}, \quad t \geq 0, \quad k,\alpha > 0. \tag{8.52}$$

It is obvious that the odds function for the LL distribution and its derivative are comparable to the chf and hrf functions for the Weibull distribution, respectively. Therefore, it is simple to illustrate that the odds-based regression models simply consider the LL distribution as a closed baseline distribution. the PO and AFT models, as examples.

According to Lawless Lawless (2011), the LL distribution can be used to support a parametric AFT model, allowing scale parameter to differ between groups. For this, we need to keep the AFT structure that we mentioned above in the odds-based

regression model formulation and adopt the derivative odds function of the LL distribution for Equation (8.52) for the reference group.

The log-logistic AFT (LL-AFT) is defined as follows:

$$r_{LL-AFT}(t;\beta,x) = r_0\left(te^{x'\beta}\right)e^{x'\beta} = e^{x'\beta}\alpha k\left(kte^{x'\beta}\right)^{\alpha-1} = \left(e^{x'\beta}k\right)^{\alpha}\alpha t^{\alpha-1} = \alpha k^* t^{\alpha-1}. \tag{8.53}$$

In Equation (8.53), we can observe the derivative of the odds structure for the LL in (8.52). As mentioned, the scale parameter differs between groups and we can write it as $T_i \sim \log-\text{logistic}\left(k^* = \left(e^{x'\beta}k\right)^{\alpha}, \alpha\right)$, with scale k^* and shape α.

On the other hand, if a LL distribution is assumed for T_i under the PO framework (LL-PO) in Equation (8.20) it then follows that $T_i \sim \log-\text{logistic}\left(k^* = e^{x'\beta}k^{\alpha}, \alpha\right)$, with scale k^* and shape α. The derivative of the odds function for the LL-PO model is rewritten as follows:

$$r_{LL-pO}(t;\beta,x) = r_0(t)e^{x'\beta} = \alpha k(kt)^{\alpha-1}e^{x'\beta} = e^{x'\beta}k^{\alpha}\alpha t^{\alpha-1} = \alpha k^* t^{\alpha-1}. \tag{8.54}$$

Technically, it is possible to compare both models in terms of the resulting scale parameter k^*. This proves that the log-logistic distribution is the only baseline distribution that is closed under all odds-based regression models.

8.4.3 Generalized Log-Logistic Baseline for All Models

The GLL distribution (Al-Aziz et al., 2022; Muse et al., 2022a,g, 2021a; Khan and Khosa, 2016) is an example of a baseline distribution that can incorporate both monotone and non-monotone hrfs, as well as be closed under both odds-based and hazard-based regression models, and has the benefit of including both the Weibull and LL models as sub-models (Muse et al., 2021a).

The hrf and the odds function of the GLL distribution are expressed as follows:

$$h_{GLL}(t;k,\alpha,\eta)=\frac{\alpha k(kt)^{\alpha-1}}{[1+(\eta t)^{\alpha}]},\quad t\geq 0,\quad k,\alpha,\eta>0, \tag{8.55}$$

$$R_{GLL}(t;k,\alpha,\eta)=[1+(\eta t)^{\alpha}]^{\frac{k^{\alpha}}{\eta^{\alpha}}}-1,\quad t\geq 0,\quad k,\alpha,\eta>0, \tag{8.56}$$

where $k>0$, $\alpha>0$, and $\eta>0$ are the distributional rate and shape parameters, respectively.

The hrf in Equation (8.55) consists of different sub-models of the GLL distribution (Muse et al., 2021a).

- Log-logistic (LL) distribution: when $k=\eta$, Equation (8.55) reduces to the hrf of an LL distribution, which is

$$h_{LL}(t;k,\alpha)=\frac{\alpha k(kt)^{\alpha-1}}{[1+(kt)^{\alpha}]},\quad t\geq 0,\quad k,\alpha>0. \tag{8.57}$$

-Burr-XII (BXII) distribution: when $\eta=1$, Equation (8.55) reduces to the hrf of a BXII-2 distribution, which is

$$h_{BXII}(t;k,\alpha)=\frac{\alpha k(kt)^{\alpha-1}}{[1+t^{\alpha}]},\quad t\geq 0,\quad k,\alpha>0. \tag{8.58}$$

- Weibull (W) distribution: when $\eta\to 0$, Equation (8.55) reduces to the hrf of the W distribution, which is

$$h_{W}(t;k,\alpha)=\alpha k(kt)^{\alpha-1},\quad t\geq 0,\quad k,\alpha>0. \tag{8.59}$$

8.5 Estimation Based on Frequentist and Bayesian Approaches

In this section, the unknown parameters of the proposed fully parametric AM class with GLL, LL and Weibull baseline distributions are estimated using frequentist (maximum likelihood estimation (MLE)) and Bayesian approaches.

8.5.1 MLE for Right-Censored Data

As was earlier indicated, not always an observed time will be a survival time: the subject is observed up to a particular time and is no longer followed up for a reason unrelated to the event occurrence. This is an illustration of a right-censored observed time, which was taken into consideration in this work and is the most common type of censoring in oncology studies. The same survival likelihood functions are reached despite the fact that there are many right-censoring techniques (Lawless, 2011). This ensures the identifiability of the distribution of the observed times under the further assumption that the survival times are independent random variables for all subjects (random censoring) and that the censoring times depend on no parameter associated with the survival function (non-informative censoring) (Collett, 2015). These presumptions allow for the formulation of a general expression for the survival likelihood function. Assuming that a survival time $T_i = t_i$ or a censored time $C_i = c_i$ are recorded for each subject, $i, 1 \leq i \leq n$. Assume also that survival (censoring) times are independent among all subjects, i.e., $T_1, \ldots, T_n \sim F_T(t; \theta_T)$ $(C_1, \ldots, C_n \sim F_C(c; \theta_C))$. The actual observable time is defined by $Y_i = \min(T_i, C_i)$, whose distribution is indexed by a vector $\theta(\theta_T, \theta_C)$ of parameters. Then, the information of a subject i is given by the pair (Y_i, δ_i), where $\delta_i = \mathbb{I}_{T_i < c_i}$ being the censoring indicator random variable. For a pair $(Y_i = t_i, \delta_i = 1\ )$ (a survival

observed time), the likelihood contribution is given by:

$$\begin{aligned}\lim_{\varepsilon\to 0^+}\frac{1}{2\varepsilon}P\left(y_i-\varepsilon<Y_i<y_i+\varepsilon,\delta_i=1;\theta\right)&=\lim_{\varepsilon\to 0^+}\frac{1}{2\varepsilon}P\left(y_i-\varepsilon<T_i<y_i+\varepsilon,T_i\leq C_i;\theta\right)\\&=\lim_{\varepsilon\to 0^+}\frac{1}{2\varepsilon}\int_{y_i-\varepsilon}^{y_i+\varepsilon}\int_t^{\infty}dF_C\left(c;\theta_C\right)dF_T\left(t;\theta_T\right)\ \text{(independence)}\\&=\lim_{\varepsilon\to 0^+}\frac{1}{2\varepsilon}\int_{y_i-\varepsilon}^{y_i+\varepsilon}\left[1-F_C\left(c;\theta_C\right)\right]dF_T\left(t;\theta_T\right)\\&=\left[1-F_C\left(y_i;\theta_C\right)\right]f_T\left(y_i;\theta_T\right).\end{aligned}\tag{8.60}$$

On the other hand, the likelihood contribution for a pair $(Y_i=c_i,\delta_i=0)$ (right censored observed time), the likelihood contribution is provided by

$$\begin{aligned}\lim_{\varepsilon\to 0^+}\frac{1}{2\varepsilon}P\left(y_i-\varepsilon<Y_i<y_i+\varepsilon,\delta_i=0;\theta\right)&=\lim_{\varepsilon\to 0^+}\frac{1}{2\varepsilon}P\left(y_i-\varepsilon<C_i<y_i+\varepsilon,T_i>C_i;\theta\right)\\&=\left[1-F_T\left(y_i;\theta_T\right)\right]f_C\left(y_i;\theta_C\right).\end{aligned}\tag{8.61}$$

Thus, under a random right censoring, the survival likelihood function for a sample $y=(y_1,...,y_n)$ of size n has the following expression:

$$L(\theta;y)=\prod_{i=1}^{n}\left\{\left[1-F_C\left(y_i;\theta_C\right)\right]f_T\left(y_i;\theta_T\right)\right\}^{\delta_i}\left\{\left[1-F_T\left(y_i\mid\theta_T\right)\right]f_C\left(y_i;\theta_C\right)\right\}^{1-\delta_i}.\tag{8.62}$$

Assuming that censoring is non-informative, i.e., the distribution of the censoring times does not depend on the parameters θ_T from the survival function, the factors $\left[1-F_C\left(y_i;\theta_C\right)\right]^{\delta_i}$ and $\left[f_C\left(y_i;\theta_C\right)\right]^{1-\delta_i}$ do not give any information for inference and can be dropped from Equation (8.62). Thereby, $\theta=\theta_T$ and a simpler survival

likelihood function is given by

$$\begin{aligned} L(\theta,\beta;D) &= \prod_{i=1}^{n} \left[f\left(t_i;\theta,\beta,x\right)\right]^{\delta_i} \left[S\left(t_i;\theta,\beta,x\right)\right]^{1-\delta_i} \\ &= \prod_{i=1}^{n} \left[\frac{h\left(t_i;\theta,\beta,x\right)}{S\left(t_i;\theta,\beta,x\right)}\right]^{\delta_i} \left[S\left(t_i;\theta,\beta,x\right)\right]^{1-\delta_i} \\ &= \prod_{i=1}^{n} \left[h\left(t_i;\theta,\beta,x\right)\right]^{\delta_i} S\left(t_i;\theta,\beta,x\right) \\ &= \prod_{i=1}^{n} \left[h\left(t_i;\theta,\beta,x\right)\right]^{\delta_i} \exp\left[-H\left(t_i;\theta,\beta,x\right)\right], \end{aligned} \tag{8.63}$$

where $D = (t_i, \delta_i, x_i, i = 1, 2, \ldots, n)$ represents the observed data including t_i = survival time, δ_i = censoring time, θ is the vector of baseline distributional parameters, and x_i = covariates. The maximum likelihood estimation can be generated via an iterative optimization method (e.g., the Newton-Raphson algorithm).

The above formulation in Equation (8.63) is useful for modelling hazard-based regression models, like the PH, AH, and GH models. An alternative version can be obtained only in terms of the odds function and its derivative as follows:

$$\begin{aligned} L(\theta,\beta;D) &= \prod_{i=1}^{n} \left[f\left(t_i;\theta,\beta,x\right)\right]^{\delta_i} \left[S\left(t_i;\theta,\beta,x\right)\right]^{1-\delta_i} \\ &= \prod_{i=1}^{n} \left[r\left(t_i;\theta,\beta,x\right) S\left(t_i;\theta,\beta,x\right)^2\right]^{\delta_i} \left[S\left(t_i;\theta,\beta,x\right)\right]^{1-\delta_i} \\ &= \prod_{i=1}^{n} \left[r\left(t_i;\theta,\beta,x\right)\right]^{\delta_i} \left[S\left(t_i;\theta,\beta,x\right)\right]^{1-\delta_i} \\ &= \prod_{i=1}^{n} \left[\frac{r\left(t_i;\theta,\beta,x\right)}{1+R\left(t_i;\theta,\beta,x\right)}\right]^{\delta_i} \left[S\left(t_i;\theta,\beta,x\right)\right], \\ &= \prod_{i=1}^{n} \left[\frac{r\left(t_i;\theta,\beta,x\right)}{1+R\left(t_i;\theta,\beta,x\right)}\right]^{\delta_i} \exp\left[-H\left(t_i;\theta,\beta,x\right)\right] \\ &= \prod_{i=1}^{n} \left[\frac{r\left(t_i;\theta,\beta,x\right)}{1+R\left(t_i;\theta,\beta,x\right)}\right]^{\delta_i} \left[\frac{1}{1+R\left(t_i;\theta,\beta,x\right)}\right]. \end{aligned} \tag{8.64}$$

The log-likelihood function corresponding to Equation (8.63) is written as follows:

$$\ell(\theta, \beta; D) = \sum_{i=1}^{n} \delta_i \log \left[H_0 \left(t_i; \theta, \beta, x \right) \right] - \sum_{i=1}^{n} H_0 \left(t_i; \theta, \beta, x \right). \tag{8.65}$$

8.5.2 The Log-Likelihood Functions

Let $\theta = (\alpha, k)', \Xi = (\beta_1', \beta_2', \beta_3')', \Omega = (\theta', \Xi')', a_i = e^{x_i'(\beta_2+\beta_3-\beta_1)}, b_i = e^{x_i'(\beta_2-\beta_1)}, c_i = e^{x_i'(\beta_3-\beta_1)}$, and $d_i = e^{k.t_i.x_i'\beta_1}$ and assume the Weibull baseline distribution, then the log-likelihood function for the Weibull-AM (W-AM) model is

$$\begin{aligned} \ell(\mathbf{\Omega}) = & \sum_{i=1}^{n} \delta_i \alpha + \sum_{i=1}^{n} \delta_i k + (\alpha - 1) \sum_{i=1}^{n} \delta_i \log d_i \\ & + \sum_{i=1}^{n} \delta_i \log a_i - \sum_{i=1}^{n} \delta_i \log \left[\{ b_i \left(1 - d_i \right) \} + d_i \right] \\ & - \sum_{i=1}^{n} \log \left[c_i \left(1 + b_i \left\{ \frac{1 - d_i}{d_i} \right\} \right) \right]. \end{aligned} \tag{8.66}$$

The log-likelihood function for the LL baseline distribution under the AM class can be expressed as follows:

$$\begin{aligned} \ell(\mathbf{\Omega}) = & \sum_{i=1}^{n} \delta_i \alpha + \sum_{i=1}^{n} \delta_i k + (\alpha - 1) \sum_{i=1}^{n} \delta_i \log \left(\log d_i \right) - \sum_{i=1}^{n} \delta_i \log \left[1 + \log d_i \right] \\ & + \sum_{i=1}^{n} \delta_i \log a_i - \sum_{i=1}^{n} \delta_i \log \left[\left\{ b_i \left(\frac{\log d_i}{1 + \log d_i} \right) \right\} + (1 + \log d_i) \right] \\ & - \sum_{i=1}^{n} \log \left[c_i \left(1 + b_i \left\{ \log d_i \right\} \right) \right]. \end{aligned} \tag{8.67}$$

Moreover, assuming $\theta = (\alpha, k, \eta)'$, and regarding the GLL baseline distribution

under the AM class, the log-likelihood function can be expressed as follows:

$$\begin{aligned}\ell(\boldsymbol{\Omega}) = &\sum_{i=1}^{n}\delta_i\alpha + \sum_{i=1}^{n}\delta_i k + (\alpha-1)\sum_{i=1}^{n}\delta_i \log\left(\log d_i\right) - \sum_{i=1}^{n}\delta_i \log\left[1+m_i\right] \\ &+ \sum_{i=1}^{n}\delta_i \log a_i - \sum_{i=1}^{n}\delta_i \log\left[\left\{b_i\left(1-\left\{1+m_i\right\}^{-\left(\frac{\kappa}{\eta}\right)^{\alpha}}\right)\right\} + \left(\left\{1+m_i\right\}^{-\left(\frac{\kappa}{\eta}\right)^{\alpha}}\right)\right] \\ &- \sum_{i=1}^{n}\log\left[c_i\left\langle 1+b_i\left\{1+m_i\right\}^{\left(\frac{\kappa}{\eta}\right)^{\alpha}}-1\right\rangle\right].\end{aligned} \tag{8.68}$$

8.5.3 Bayesian Inference

In this section, we offer general guidelines for prior selection of the regression coefficients associated with covariates and baseline distribution parameters. We examined a prior independent scenario between the baseline parameters in H_0(t) (baseline hazard) or $R_0(t)$(baseline odds) and the regression coefficients. Additionally, we determined the prior independence of the regression coefficients in a non-informative scenario with normal distributions of zero mean and a large known variance (Lázaro et al., 2021) as

$$\pi\left(H_0,\beta_1,\beta_2,\beta_3\right) = \pi\left(H_0\right)\pi(\beta_1,\beta_2,\beta_3) = \pi\left(H_0\right)\prod_{j=1}^{J}\mathrm{N}\left(\beta_j \mid 0,\sigma_j^2\right), \tag{8.69}$$

where $\pi\left(H_0\right)$ is the prior distribution of all baseline parameters and hyperparameters in H_0(t).

For the baseline hazard parameter θ in baseline distributions, we consider the following priors:

$$\pi(\alpha) \sim G\left(a_1,b_1\right) = \frac{b_1^{a_1}}{\Gamma\left(a_1\right)}\alpha^{a_1-1}e^{-b_1\alpha}; a_1,b_1,\alpha>0, \tag{8.70}$$

$$\pi(\eta) \sim G\left(a_2,b_2\right) = \frac{b_2^{a_2}}{\Gamma\left(a_2\right)}\eta^{a_2-1}e^{-b_2\eta}; a_2,b_2,\eta>0, \tag{8.71}$$

$$\pi(k) \sim G\left(a_3, b_3\right) = \frac{b_3^{a_3}}{\Gamma\left(a_3\right)} k^{a_3-1} e^{-b_3 k}; a_3, b_3, k > 0. \tag{8.72}$$

The values of the hyper-parameters values of the prior distributions are selected from the historical data of the baseline distribution (Muse et al., 2021a).

For the regression coefficients prior, we have

$$\pi\left(\beta_1'\right) \sim N\left(a_4, b_4\right), \tag{8.73}$$

$$\pi\left(\beta_2'\right) \sim N\left(a_5, b_5\right), \tag{8.74}$$

$$\pi\left(\beta_3'\right) \sim N\left(a_6, b_6\right). \tag{8.75}$$

The joint prior distribution for the distributional parameters and coefficient of regression expressed as follows:

$$\pi\left(\alpha, k, \eta, \beta_1', \beta_2'\right) = \pi(\alpha)\pi(\eta)\pi(k)\pi\left(\beta_1'\right)\pi\left(\beta_2'\right)\pi\left(\beta_3'\right). \tag{8.76}$$

The model must be supplied with data $\mathcal{D} = \left\{\left(t_i, \delta_i, \boldsymbol{x}_i\right), i = 1, \ldots, n\right\}$, where t_i is the observed lifetime time for the i th individual, δ_i is the censoring status taking 1 if the event of interest has occurred and 0 otherwise, and $\boldsymbol{x}_i$ are the explanotory variables. Prior knowledge and experimental data are combined in the posterior distribution via the Bayes' theorem, and we get

$$\pi\left(H_0, \beta_1, \beta_2, \beta_3; \mathcal{D}\right) \propto \mathcal{L}\left(H_0, \beta_1, \beta_2, \beta_3\right)\pi\left(H_0, \beta_1, \beta_2, \beta_3\right), \tag{8.77}$$

where $\mathcal{L}\left(H_0, \beta\right)$ is the likelihood function of $\left(H_0, \beta_1, \beta_2, \beta_3\right)$ given in Equation (8.63). This study uses Markov chain Monte Carlo (McMC) techniques for Bayesian inference, and the Metropolis within the Gibbs algorithm is used to sample from the posterior distribution (Smith and Roberts, 1993). In our implementation, the independence sampler is used to update each parameter component (Vines et al., 1996).

8.6 Practical Illustrations

A clinical trial right-censored oncology dataset is examined in this section to demonstrate the applicability and tractability of the proposed models, including the fully-parametric AM Class, GO, and AO models with three different baseline distributions, including Weibull, LL, and GLL baseline distributions in modelling right-censored survival data with crossing survival curves. We compared the proposed AM class with its sub-models that contain both hazard-based regression models, including PH, AH, AFT, and GH models, and the odds-based regression models, including PO, AO, AFT, and GO models, using both the MLE frequentist approach (MLE) estimation technique and Bayesian approaches using noninformative priors. The class and its sub-models were compared using different information criteria, including the classical ones (AIC, BIC, BCAIC, CAIC, and HQIC), Bayesian model selection (WAIC, and LOOIC), and checking the nested structure of the AM class using the LRT test.

8.6.1 IPASS Clinical Trial Data Set

In order to show the applicability of the proposed models, we re-analyzed a large dataset from a randomized clinical trial called IPASS for this study. In a randomized controlled trial, gefitinib vs. carboplatin-paclitaxel was compared for progression-free survival in patients with advanced pulmonary adenocarcinoma. An unadjusted PH model was used to examine the main outcome. Despite the implicit violation of the PH assumption represented by the crossing of the two survival curves, the study's findings were published using this model (Mok et al., 2009).

Argyropoulos and Unruh (2015) reconstructed and re-published the IPASS dataset, and it is now freely available in an AHSurv R package (Muse et al., 2022e). The features stated in the references are all still there in this reconstructed dataset,

which is also accessible to the clinical trial's results. The months of March 2006 through April 2008 are covered by the database. The main objective of the trial is to evaluate the effects of gefitinib versus carboplatin/paclitaxel doublet chemotherapy on progression-free survival (in months) in a subset of patients with non-small-cell lung cancer (NSCLC). According to the trial's design, $n = 1207$ previously untreated individuals in East Asia with advanced lung adenocarcinoma and who were non-smokers or previous light smokers were randomly assigned to either carboplatin + paclitaxel (608 patients) or gefitinib (608 patients) (609 patients). The observations show 965 occurrences of the event of interest (79.3 percent), with 449 (73.7 percent) relating to patients receiving gefitinib and 516 (84.9 percent) related to patients receiving carboplatin+paclitaxel.

The primary goal of this section is to appropriately assess the rebuilt IPASS data and estimate the regression coefficients using the proposed fully-parametric AM class provided in Section 8.3. For the proposed model, we evaluate both the maximum likelihood and the Bayesian estimating approaches to achieve this goal.

We fit all hazard-based and odds-based regression models as well as the general proposed AM class using three different baseline distributions, namely Weibull, LL, and GLL distributions, letting $x_i = I(\text{treatment} = \text{chemotherapy})$, which equals 1 if the treatment involves gefitinib and 0 if the treatment involves carboplatin/paclitaxel. Tables 8.1–8.3 provide a summary of the numerical results.

Figure 8.4 displays the total time on test (TTT) plot for the survival time and the survival curves of the two types of drugs where crossing between the curves can be seen, which confirms the efficacy of the proposed novel models in this study, including the AM, GO, and AO models, plus some other existing models in the literature, including the GH, and AH models, and that it is appropriate for the analysis of survival data with crossover survival curves.

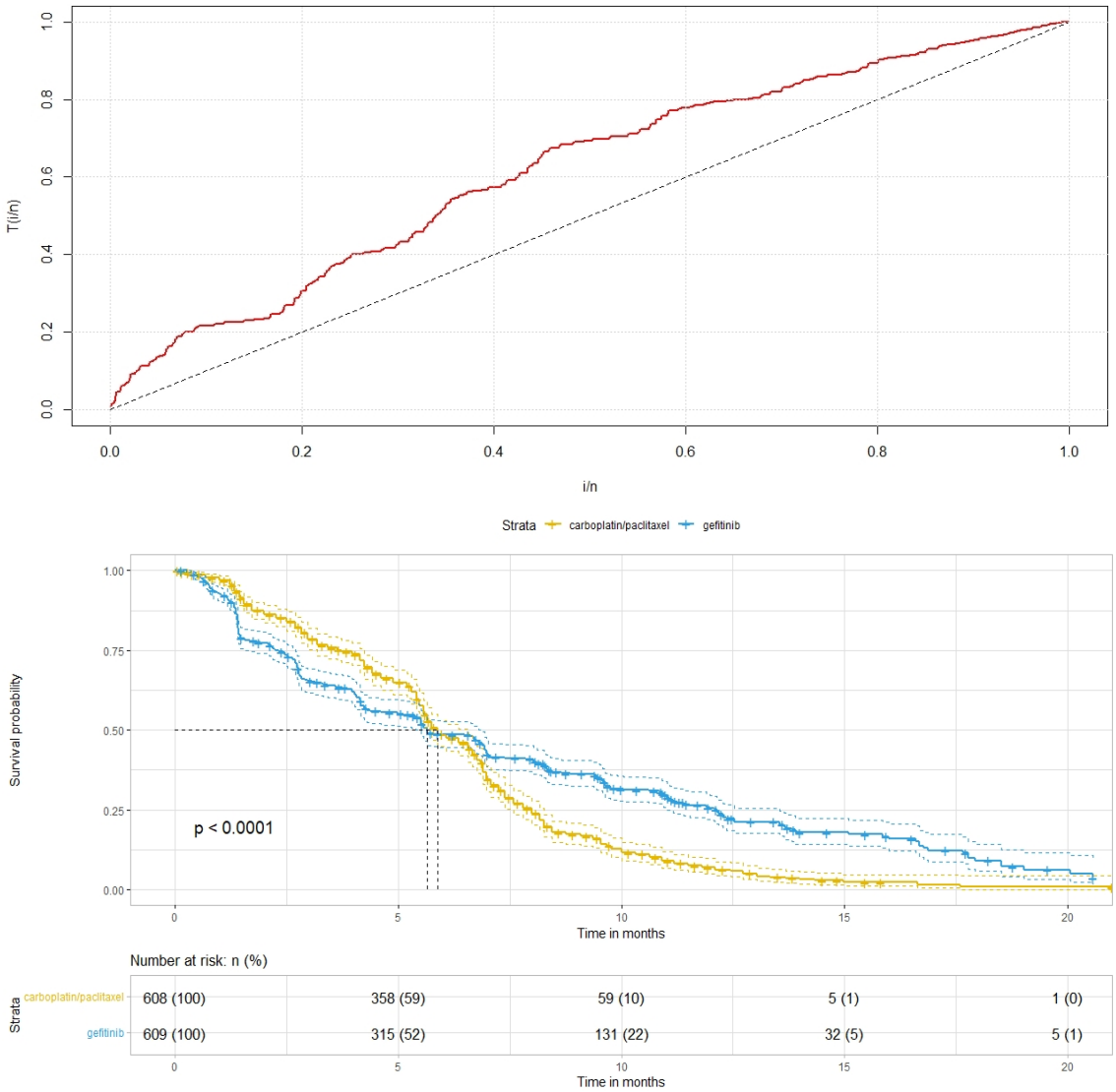

Figure 8.4: Illustrating the total time on test (TTT) plot, and the crossing survival curves for the two types of drugs determined using the Kaplan-Meier method for the IPASS dataset.

The parameters and related standard errors from the various hazard- and odds-based regression models employing the Weibull baseline distribution and five different information criterion estimations are shown in Table 8.1. The TTT plot in Figure 8.4 shows that the data's increasing hazard rate points to the theoretical use of the Weibull baseline distribution. According to the findings, the Weibull-Amoud class has the lowest values for each information criterion when compared to all other competing regression models, demonstrating its superiority over other hazard-based

and odds-based regression models. Another crucial aspect that restricts the use of the Weibull baseline distribution is the fact that all hazard-based regression models yield the same result, which is a weakness of the Weibull baseline.

All hazard-based regression models, including the AH, PH, AFT, and GH models, produce the same findings when compared to the Weibull baseline as illustrated in Table 8.1. The LL baseline distribution was used to fit and compare all of the regression models after we looked at an alternate baseline distribution. Using the LL baseline distribution and five different information criteria, Table 8.2 provides estimates of the parameters and related standard errors from the various hazard-based and odds-based regression models. According to the AIC values in Table 8.2, there is no clear preference for one model over the other. The fact that all odds-based regression models yield the same result is a significant factor that restricts the applicability of the LL baseline distribution.

Table 8.1: Results from the fitted proposed fully-parametric odds-based and hazard-based regression models with W baseline distribution to IPASS data set

Models	Parameter(s)	Estimate	SE	AIC	BCAIC	BIC	CAIC	HQIC
W-AM	β_1	0.167	0.008	5552.888	5583.409	5578.409	5552.849	5562.495
	β_2	-1.346	0.109					
	β_3	2.815	0.129					
	α	1.845	0.002					
	κ	7.003	0.009					
W-GH	β_1	1.398	0.007	5686.968	5711.385	5707.385	5686.943	5694.653
	β_2	-0.862	0.129					
	α	1.362	0.129					
	κ	6.613	0.129					
W-GO	β_1	-2.574	0.008	5609.716	5634.133	5630.133	5609.691	5617.402
	β_2	1.921	0.011					
	α	1.596	0.003					
	κ	6.870	0.005					
W-AH	β	-0.875	0.005	5684.474	5702.787	5699.787	5684.460	5690.239
	α	1.365	0.004					
	κ	6.762	0.003					
W-PH	β	-0.320	0.007	5684.474	5702.787	5699.787	5684.460	5690.239
	α	1.365	0.004					
	κ	6.762	0.003					
W-AO	β	-0.567	0.001	5652.746	5671.058	5668.058	5652.731	5658.510
	α	1.419	0.003					
	κa	6.515	0.002					
W-PO	β	-0.003	0.017	5708.640	5726.952	5723.952	5708.625	5714.404
	α	1.341	0.010					
	κ	7.594	0.006					
W-AFT	β	-0.234	0.009	5684.474	5702.787	5699.787	5684.460	5690.239
	α	1.365	0.004					
	κ	6.762	0.003					

When the Weibull distribution is used as the baseline distribution, all hazard-based regression models exhibit coincidence as shown in Table 8.1. On the other hand, when the baseline distribution is an LL distribution, all odds-based regression models exhibit coincidence as illustrated in Table 8.2.

These two points recommend looking for and utilizing a modified baseline distribu-

tion, which can provide us with various results for all survival regression models, regardless of whether they are hazard-based, odds-based, or a combination of both. We used the GLL baseline distribution, where a sub-model of the Weibull distribution that yields different results for all the regression models taken into consideration in this study, to close the gap and compare the seven different hazard-based and odds-based regression models that are currently in use.

Table 8.2: Results from the fitted proposed fully-parametric odds-based and hazard-based regression models with LL baseline distribution to IPASS data set

Models	Parameter(s)	Estimate	SE	AIC	BCAIC	BIC	CAIC	HQIC
LL-AM	β_1	2.185	0.030	5690.051	5720.572	5715.572	5690.014	5699.658
	β_2	-0.361	0.009					
	β_3	-0.171	0.011					
	α	2.291	0.022					
	κ	5.498	0.006					
LL-GH	β_1	1.074	0.018	5688.051	5712.468	5708.468	5688.027	5695.737
	β_2	-0.088	0.019					
	α	2.291	0.029					
	κ	5.498	0.109					
LL-GO	β_1	-0.845	0.038	5757.552	5781.968	5777.968	5757.527	5765.237
	β_2	0.726	0.129					
	α	1.801	0.129					
	κ	5.478	0.129					
LL-AH	β	1.119	0.014	5687.976	5706.289	5703.289	5687.961	5693.740
	α	2.259	0.008					
	κ	5.614	0.009					
LL-PH	β	-0.128	0.013	5751.598	5769.911	5766.911	5751.584	5757.362
	α	1.832	0.010					
	κ	5.156	0.009					
LL-AO	β	0.061	0.007	5755.552	5773.864	5770.864	5755.537	5761.316
	α	1.801	0.009					
	κ	5.478	0.010					
LL-PO	β	0.049	0.004	5755.552	5773.864	5770.864	5755.537	5761.316
	α	1.801	0.009					
	κ	5.478	0.010					
LL-AFT	β	0.027	0.009	5755.552	5773.864	5770.864	5755.537	5761.316
	α	1.801	0.009					
	κ	5.478	0.010					

Table 8.3: Results from the fitted proposed fully-parametric odds-based and hazard-based regression models with GLL baseline distribution to IPASS data set

Models	Parameter(s)	Estimate	SE	AIC	BCAIC	BIC	CAIC	HQIC
GLL-AM	β_1	0.179	0.003	**5554.864**	**5591.489**	**5585.489**	**5554.810**	**5566.392**
	β_2	-1.344	0.001					
	β_3	2.824	0.002					
	α	1.853	0.129					
	κ	0.143	0.129					
	η	0.010	0.129					
$GLL-GH$	β_1	2.049	0.001	5605.079	5635.600	5630.600	5605.041	5614.686
	β_2	-0.835	0.002					
	α	1.818	0.129					
	κ	0.151	0.129					
	η	0.045	0.129					
$GLL-GO$	β_1	-0.635	0.002	5645.507	5676.028	5671.028	5645.470	5655.114
	β_2	0.339	0.001					
	α	1.364	0.129					
	κ	0.145	0.129					
	η	0.000	0.129					
$GLL-AH$	β	1.545	0.001	5656.576	5680.993	5676.993	5656.551	5664.262
	α	1.886	0.129					
	κ	0.154	0.129					
	η	0.101	0.129					
$GLL-PH$	β	-0.064	0.014	5686.187	5710.604	5706.604	5686.162	5693.872
	α	-0.171	0.129					
	κ	-0.171	0.129					
	η	-0.171	0.129					
$GLL-AO$	β	-0.708	0.002	5661.339	5685.755	5681.755	5661.314	5669.024
	α	1.482	0.129					
	κ	0.158	0.129					
	η	0.000	0.129					
$GLL-PO$	β	-0.008	0.021	5708.832	5733.249	5729.249	5708.808	5716.518
	α	1.406	0.129					
	κ	0.139	0.129					
	η	0.031	0.129					
GLL-AFT	β	-0.166	0.014	5688.582	5712.999	5708.999	5688.557	5696.267
	α	1.359	0.129					
	κ	0.143	0.129					
	η	0.000	0.129					

For the proposed AM class and seven different hazard- and odds-based regression models with GLL baseline distribution, the parameter estimates and their associated

standard errors are shown in Table 8.3. We see that for the eight competing models, the estimates of the baseline distribution parameters and their standard errors are quite similar and within a reasonable range. The GLL-AM model appears to be preferred above the other competing models and provided the best-fitting model, according to the values of the five distinct information criteria. The results also showed that the GH and GO models are preferred over their sub-models. Finally, the results indicate that the only basic survival regression models that can be used to model and analyze survival data with crossing survival curves are the AH and AO models.

The GLL-AM regression model is the best model compared to the others, according to the likelihood ratio test (LRT) results in Table 8.4. The previously stated result is supported by the plots of the estimated hazards in Figure8.7.

Table 8.4: LRT Values for the Amoud class and its sub-models using IPASS Data Set

Model	Hypothesis	LRT Statistic	P-value
GH	$H_0 : \beta_2 = \beta_1, H_1 : H_0$ is false,	52.214	<0.0001
GO	H_0: $\beta_3 = \beta_1$, $H_1 : H_0$ i is false,	92.644	<0.0001
AH	H_0: $\beta_2 - \beta_1 = 0, \beta_3 = 0$, $H_1 : H_0$ is false,	105.712	<0.0001
AO	$H_0 : \beta_3 - \beta_1$=0,β_2=0, $H_1 : H_0$ is false,	135.322	<0.0001
PH	$H_0 : \beta_1 = \beta_2$=0, $H_1 : H_0$ i s false,	110.474	<0.0001
PO	H_0: $\beta_1 = \beta_3$=0, $H_1 : H_0$ is false,	157.968	<0.0001
AFT	$H_0 : \beta_1 = \beta_2 = \beta_3$, $H_1 : H_0$ is false,	137.718	<0.0001

According to the likelihood ratio test (LRT) results in Table 8.5, the GLL-GH regression model is the most effective of the alternatives hazard-based regression models.

Table 8.5: LRT Values for the GH Model and its sub-models using IPASS Data Set

Models	Hypothesis	LRT Statistic	P-value
GH vs. AH	$H_0 : \beta_2$=0, $H_1 : H_0$ i is false,	53.498	<0.0001
GH vs. PH	$H_0 : \beta_1$=0, $H_1 : H_0$ i is false,	83.108	<0.0001
GH vs. AFT	$H_0\ \beta_1 = \beta_2$, $H_1 : H_0$ is false,	85.504	<0.0001

According to the likelihood ratio test (LRT) results in Table 8.6, the GLL-GO regression model is the most effective of the alternatives odds-based regression models.

Table 8.6: LRT values for the GO Model and its sub-models using IPASS Data Set

Models	Hypothesis	LRT Statistic	P-value
GO vs. AO	$H_0 : \beta_2$=0, $H_1 : H_0$ is false,	17.830	<0.0001
GO vs. PO	$H_0 : \beta_1$=0, $H_1 : H_0$ is false,	65.324	<0.0001
GO vs. AFT	$H_0 : \beta_1 = \beta_2$, $H_1 : H_0$ is false,	45.074	<0.0001

The W-AM model is the best model among the others, according to the hazard plots of the estimated hazard rates in Figure 8.5, while other hazard-based models with Weibull baseline distributions fitted similarly and there was no more difference at all. According to the plots in Figure 8.6, there is no model that is preferred over the others when the baseline distribution is the LL distribution, and there is no difference between any of the odds-based regression models that were fitted. Finally, as illustrated in Figure 8.7, the GLL-AM is the superior compared to the other competitive models.

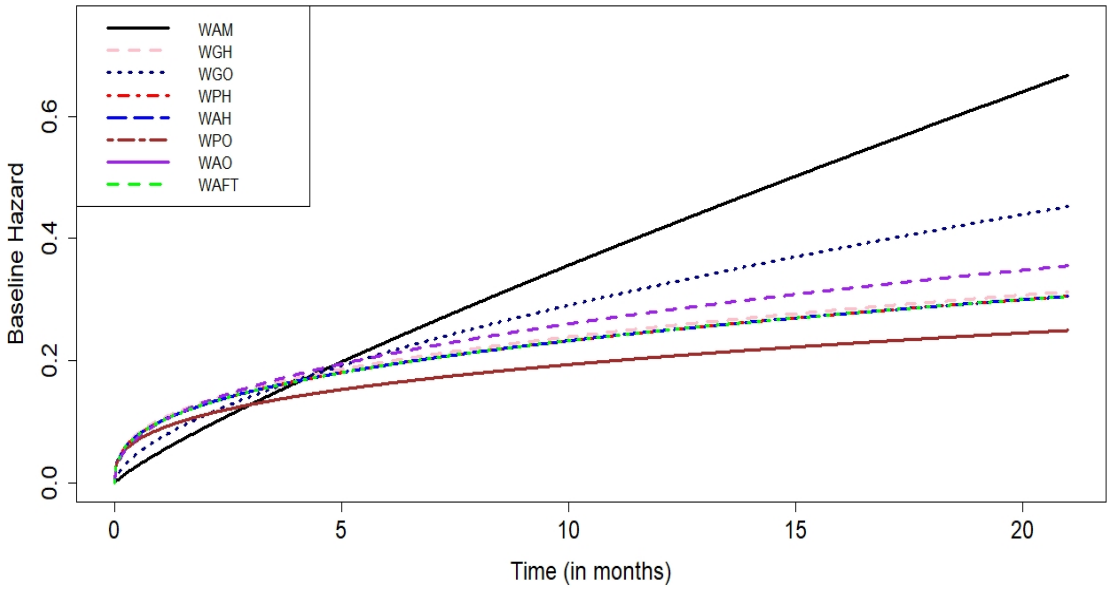

Figure 8.5: Estimated hrfs for the competitive models of the IPASS dataset.

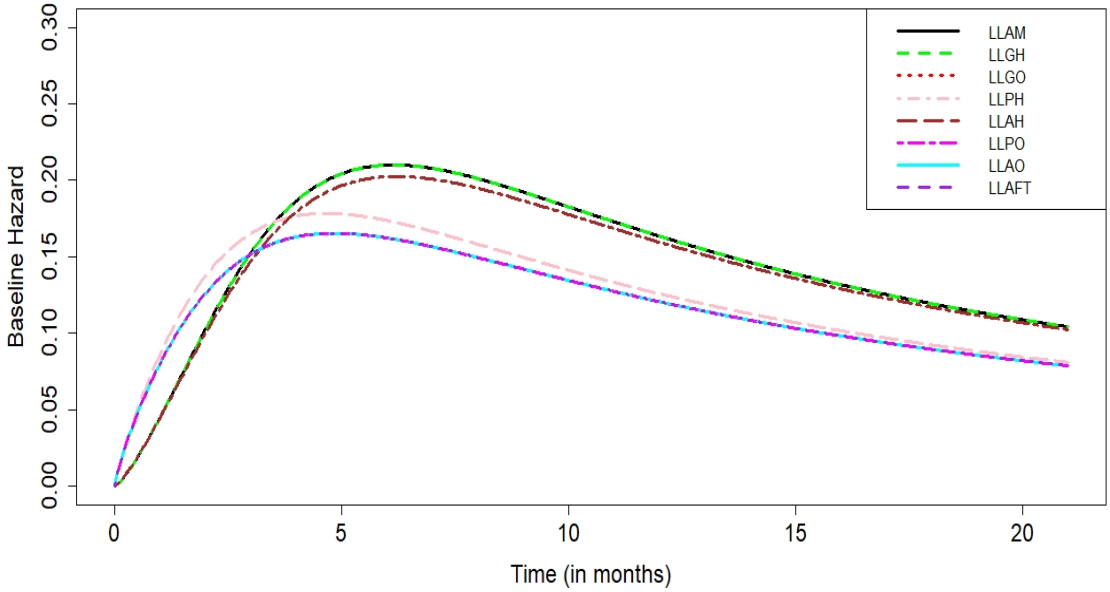

Figure 8.6: Estimated hrfs for the competitive models of the IPASS dataset.

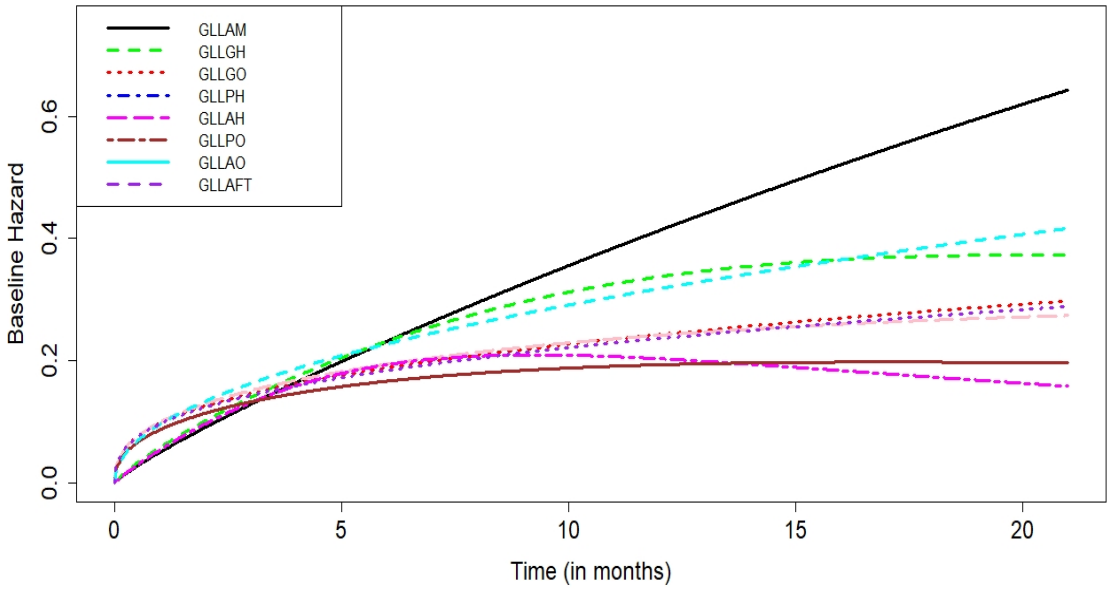

Figure 8.7: Estimated hrfs for the competitive models of the IPASS dataset.

8.6.2 Bayesian Analysis

We performed all Bayesian survival inferential procedures resulting from the combination of the aforementioned generic baseline distribution specifications with various prior scenarios in baseline parameters as well as regression coefficients related to explanatory variables. Using the Rstan package in R (Carpenter et al., 2017), the joint posterior distribution for each model was approximated. We performed four parallel chains with 3,000 iterations and a burn-in of 1,000 for each estimated model. To lessen autocorrelation in the sample, chains were also trimmed by storing after every fifth iteration. With a prospective scale reduction factor close to 1 and an actual number of separate simulation draws more than 400, convergence to the joint posterior distribution was assured (Ashraf-Ul-Alam and Khan, 2021; Alvares et al., 2021).

The posterior distribution's numerical summary characteristics are displayed in Ta-

ble 8.7. According to the summary results, the McMC algorithm has converged to the joint posterior distribution because the potential scale reduction factor ($\hat{R}$) is 1, the effective sample size ($n-eff$) is greater than 400, and the Monte Carlo standard error (SE) is less than 5percent of the posterior standard deviations (SD) for all of the parameters. For visually examining convergence, use trace graphs. The trace plots in Figures 8.8 to 8.16 demonstrate a stationary pattern fluctuating inside a band, demonstrating convergence of the McMC algorithm. For the proposed Amoud class, density and autocorrelation graphs are also employed in Figures 8.17 and 8.17 respectively, and both show that the McMC algorithm has converged. Table 8.8 displays the computed models' WAIC and LOOIC values. In comparison to the other fitted models, including the GLL-AH, GLL-AO, GLL-PO, GLL-AFT, GLL-GH, GLL-GO, and GLL-PH models, the GLL-AM model performs better based on the WAIC and LOOIC values. The worst performance is displayed by the most popular survival regression models, such as the PH, PO, and AFT models. This proves that, despite their frequent application, these models are not appropriate for handling survival data with crossing survival curves.

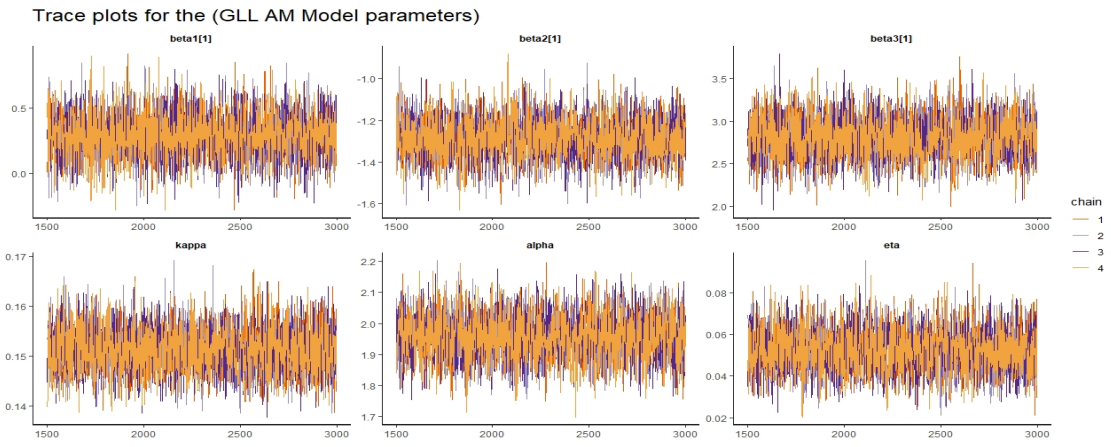

Figure 8.8: Trace plots for the GLL-AM model parameters.

Table 8.7: Results for the posterior properties of the competitive models.

Models	Par (s)	Estimate	SE	SD	2.5%	Medium	97.5%	$n-eff$	$\hat{R}$
GLL-AM	β_1	0.293	0.003	0.166	-0.032	0.291	0.618	4108	1.000
	β_2	-1.287	0.002	0.096	-1.473	-1.288	-1.096	3309	1.000
	β_3	2.810	0.004	0.242	2.349	2.811	3.279	3431	1.000
	α	1.953	0.001	0.072	1.818	1.952	2.095	3477	1.000
	κ	0.151	0.000	0.004	0.143	0.151	0.160	4309	1.001
	η	0.052	0.000	0.010	0.032	0.052	0.073	3613	1.001
GLL-GH	β_1	1.839	0.005	0.219	1.444	1.828	2.296	2068	1.001
	β_2	-0.672	0.003	0.139	-0.968	-0.663	-0.425	2232	1.000
	α	1.851	0.001	0.075	1.712	1.849	2.004	2672	1.002
	κ	0.155	0.000	0.005	0.145	0.155	0.165	2431	1.001
	η	0.059	0.000	0.013	0.036	0.059	0.087	1739	1.002
GLL-GO	β_1	-0.395	0.001	0.065	-0.524	-0.395	-0.263	3169	1.001
	β_2	0.157	0.001	0.077	0.005	0.158	0.307	3303	1.001
	α	1.450	0.001	0.041	1.371	1.450	1.533	3576	1.000
	κ	0.145	0.000	0.004	0.137	0.145	0.154	3523	1.001
	η	0.008	0.000	0.007	0.000	0.007	0.027	3664	1.000
GLL-AH	β	1.448	0.003	0.143	1.163	1.445	1.728	3090	1.000
	α	1.912	0.002	0.089	1.742	1.911	2.087	2446	1.000
	κ	0.159	0.000	0.006	0.148	0.159	0.170	2406	1.000
	η	0.112	0.000	0.013	0.088	0.112	0.137	1989	1.000
GLL-PH	β	-0.294	0.001	0.066	-0.427	-0.294	-0.168	3606	1.000
	α	1.526	0.001	0.054	1.425	1.524	1.635	3214	1.001
	κ	0.167	0.000	0.007	0.155	0.167	0.180	3082	1.000
	η		0.000	0.016	0.042	0.071	0.104	2693	1.002
GLL-AO	β	-0.557	0.001	0.092	-0.740	-0.556	-0.378	4181	1.000
	α	1.481	0.001	0.041	1.403	1.480	1.561	3743	1.001
	κ	0.163	0.000	0.005	0.153	0.163	0.174	3575	1.001
	η	0.042	0.000	0.011	0.022	0.042	0.067	3654	1.002
GLL-PO	β	-0.011	0.002	0.105	-0.214	-0.011	0.196	2226	1.001
	α	1.412	0.002	0.059	1.308	1.408	1.537	1028	1.003
	κ	0.140	0.000	0.007	0.128	0.140	0.156	837	1.003
	η	0.033	0.001	0.019	0.003	0.032	0.075	726	1.004
GLL-AFT	β	-0.188	0.001	0.054	-0.291	-0.190	-0.078	3662	1.000
	α	1.481	0.001	0.051	1.388	1.479	1.584	3041	1.000
	κ	0.161	0.000	0.006	0.150	0.161	0.174	2976	1.000
	η	0.061	0.000	0.016	0.032	0.060	0.095	2601	1.000

Table 8.8: Bayesian Model selection between the proposed AM class and its sub-models using the GLL baseline distribution.

Model	WAIC	LOOIC
GLL-AM	**5559.50**	**5559.54**
GLL-GH	5608.30	5608.28
GLL-GO	5651.76	5651.69
GLL-AH	5657.40	5697.43
GLL-PH	5692.50	5692.51
GLL-AO	5666.60	5666.61
GLL-PO	5708.60	5708.62
GLL-AFT	5698.00	5698.04

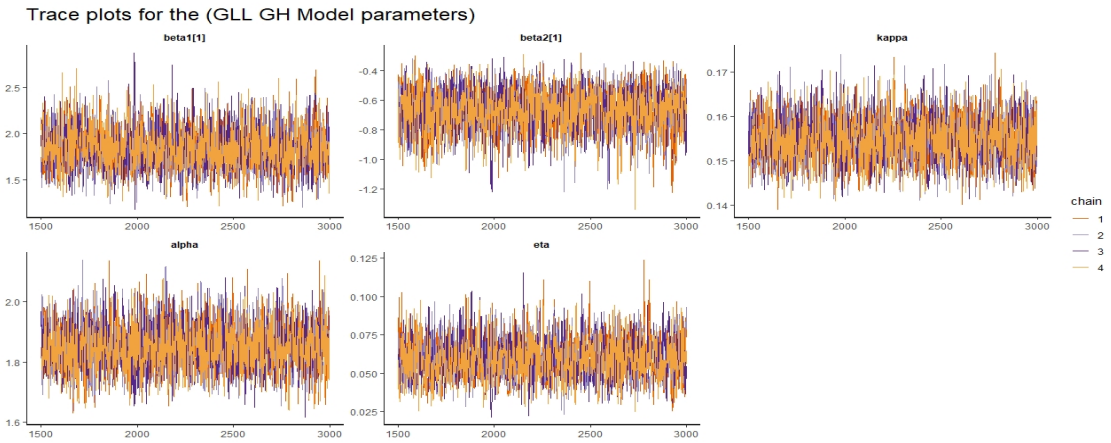

Figure 8.9: Trace plots for the GLL-GH model parameters.

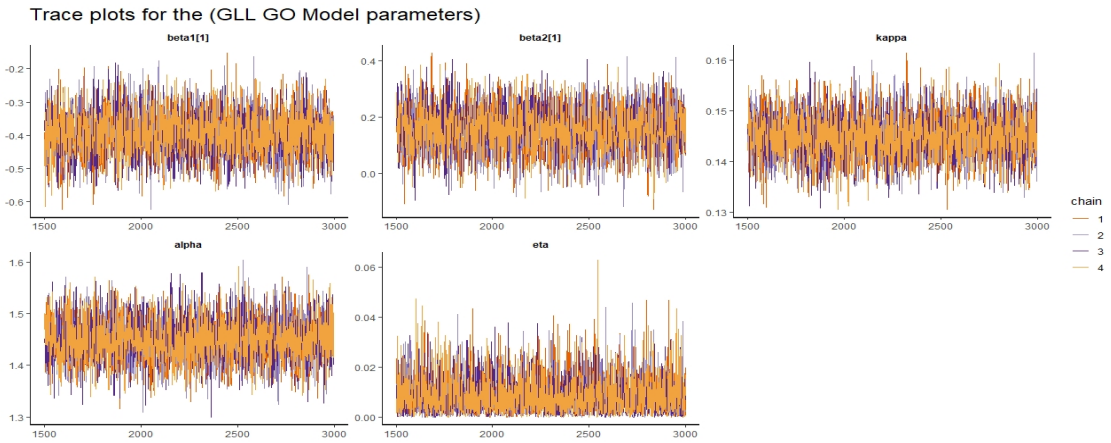

Figure 8.10: Trace plots for the GLL-GO model parameters.

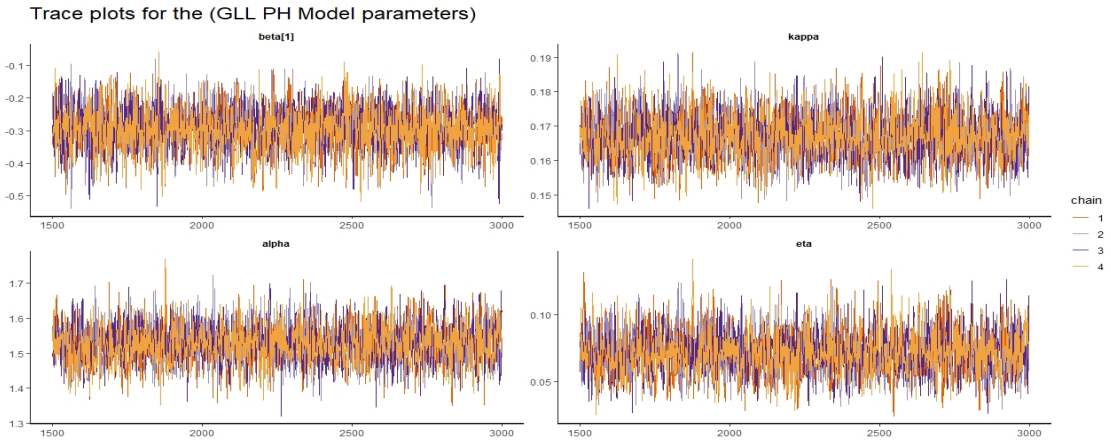

Figure 8.11: Trace plots for the GLL-PH model parameters.

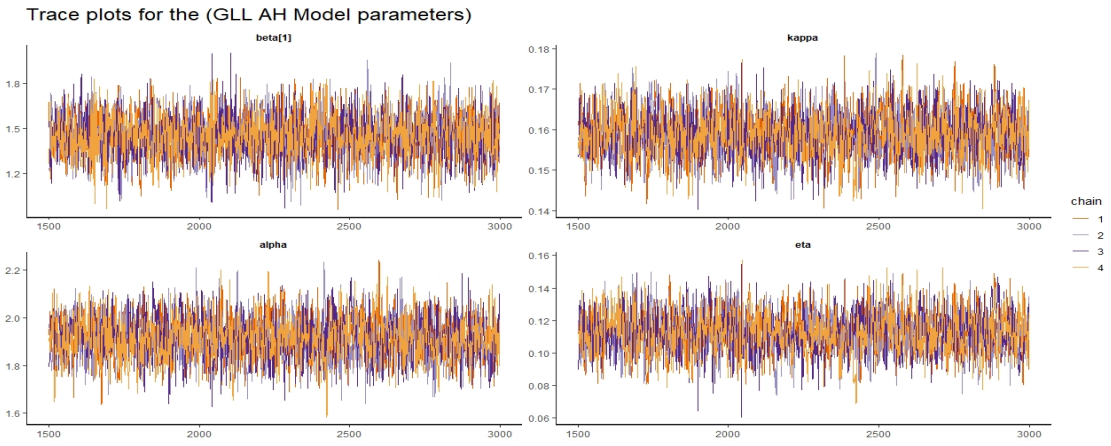

Figure 8.12: Trace plots for the GLL-AH model parameters.

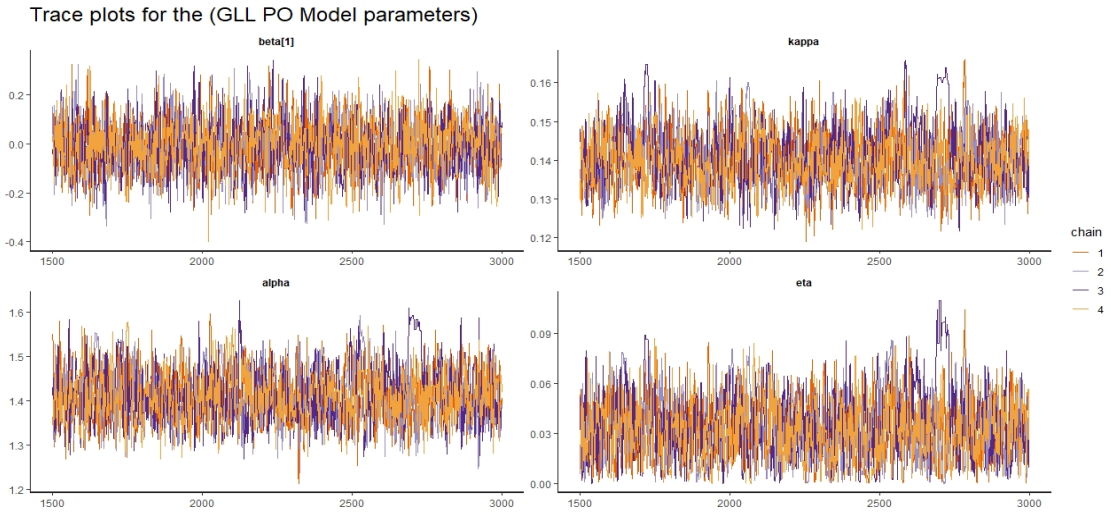

Figure 8.13: Trace plots for the GLL-PO model parameters..

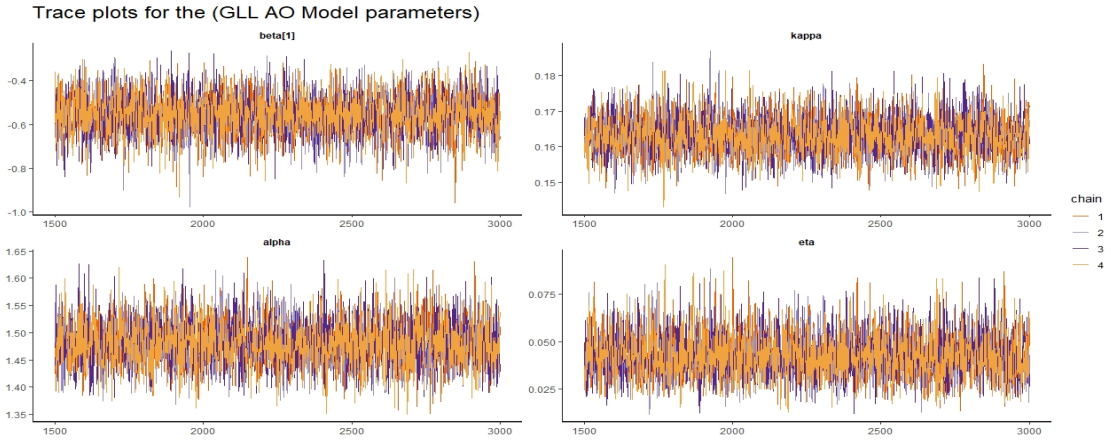

Figure 8.14: Trace plots for the GLL-AO model parameters..

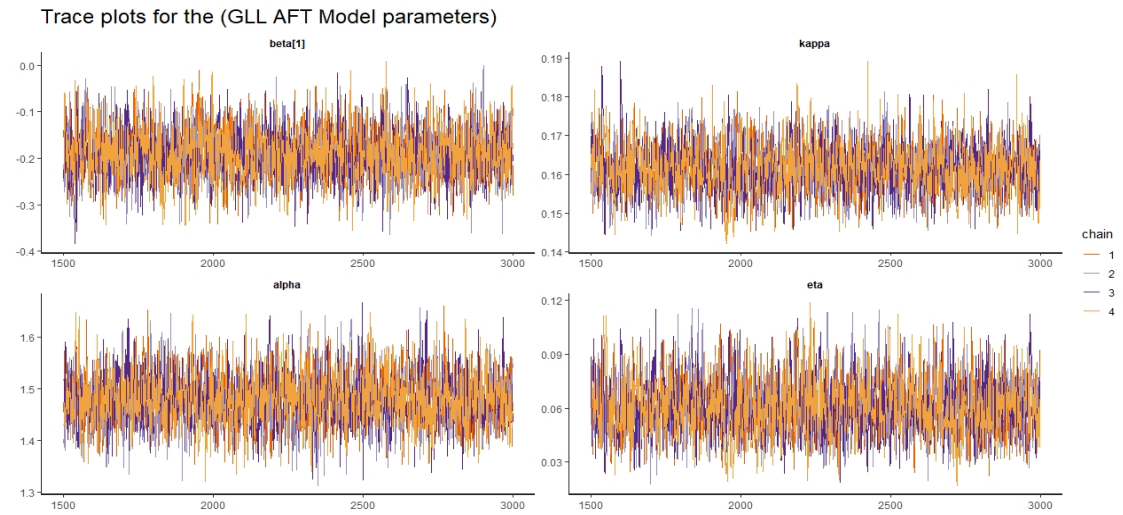

Figure 8.15: Trace plots for the GLL-AFT model parameters.

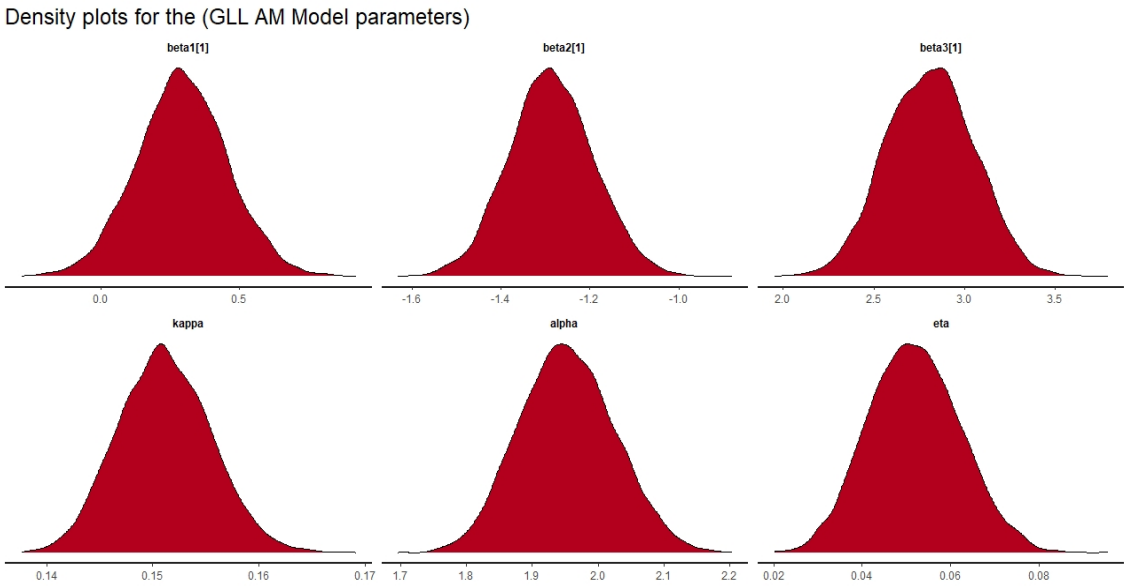

Figure 8.16: Density plots for the GLL-AM model parameters.

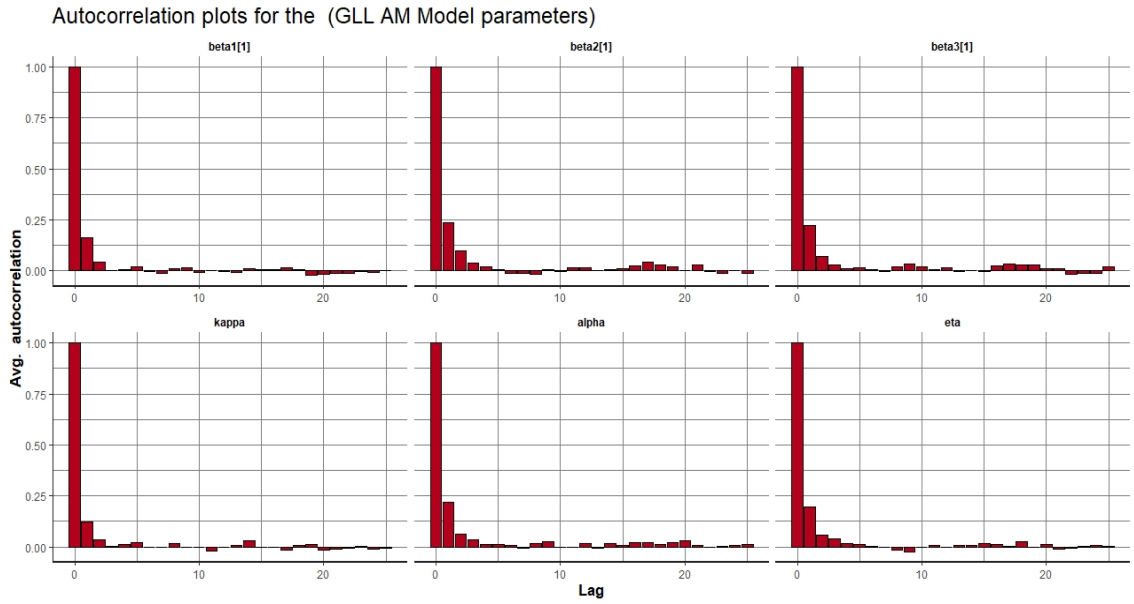

Figure 8.17: Autocorrelation plots for the GLL-AM model parameters.

8.7 Conclusions

We investigated a novel, general, flexible, fully parametric class for hazard-based and odds-based regression models, named the AM class, with a GLL baseline distribution that can incorporate the basic shapes of the failure rate and contains, as specific cases, the main survival regression models of interest in time-to-event analysis: PO, PH, AO, AH, AFT, GO, and GH models. However, the AH, AO, and GO models' restricted utility is mostly due to a lack of reliable and efficient estimating methods. We demonstrated that both classical and Bayesian inference may be performed using existing optimization techniques by adopting a flexible parametric baseline distribution.

The proposed AM class framework is quite adaptable and can easily be applied to a wide range of reliability and survival analysis applications. This framework specifically incorporates and generalizes the practically significant PH, AFT, AH, GH, PO, AO, and GO survival regression models. Additionally, the GLL baseline model, which only requires one additional parameter, accounts for the main hrf

shapes (monotone and non-monotone) within some of the most common baseline distributions (Burr type XII, LL, Weibull, and exponential distributions).

The combination of such adaptable parametric odds-based and hazard-based regression models with the AM class structure is a potent tool for modeling survival times. Although we concentrated on overall survival models, the proposed tractable fully parametric AM class is equally useful in excess hazard (relative survival) models. In the AM class, we used the GLL distribution as a baseline distribution; however, other versatile parametric distributions, such as the generalized Weibull, exponentiated Weibull, power generalized Weibull, and generalized gamma distributions, can also accommodate the basic shapes of the hrf including constant, monotone and non-monotone shapes.

We only used the GLL distribution in this case since it allows for a simple implementation, makes parameter interpretation easier, and the accompanying MLEs and Bayesian estimators are consistent and asymptotically normal in the presence of right-censored observations. Finally, an R package called AmoudSurv was developed to fit the odds-based regression models (Muse et al., 2022b).

In the future, we want to develop an R package to fit the most common parametric hazard-based and odds-based regression models, such as the AH, AO, AFT, PH, PO, GO, GH, and AM models, with different baseline distributions that can represent varied hazard rates. This study's technique can also be extended to numerous event scenarios, such as the multi-state model, competing risk model, and to include lifetime data with cure proportion rate and frailty characteristics. It is also possible to adapt it to joint model frameworks, spatial models, mixed effects models, and excess hazard models. Other strategies for censoring observations, such as interval censoring, left censoring, middle-censoring, and double-censoring, could be utilized in future investigations. This is beyond the focus of this study, but it will be covered in many others.

CHAPTER 9

Conclusions, Contributions and Future Work

9.1 Introduction

This chapter presents the summary of the dissertation, main contributions and recommendations for future work.

9.2 Conclusions

Right-censored survival data arise naturally in many oncology studies. The existing literature in right-censored survival data analysis mainly focuses on the estimation of the distribution function of the lifetime variate without presence of covariates, small studies used the PH and AFT models. There are many situations where the lifetime data are observed along with covariates.

The existing statistical methods for various censoring schemes such as right, left, interval, middle and double censoring are not enough in such contexts and therefore new flexible Bayesian parametric hazard-based regression models along with modified baseline distributions are required for the analysis of such a data. In view of this, in this thesis, we have proposed several flexible Bayesian Parametric hazard-based regression models for right-censored survival data.

In chapter 3, we studied the mathematical and statistical properties of the baseline distribution considered in this thesis. The proposed baseline hazard parametesr are estimated using both classical via MLE and Bayesian approaches. An extensive Monte Carlo simulation is used to assess the performance of the estimators. A real-

life oncology data set related to bladder cancer patients is used to demonstrate the applicability and flexibility of the baseline distribution.

In chapter 4, we presented a flexible Bayesian parametric proportional hazard model and the model parameters are estimated via the Bayesian approach using non-informative priors. Extensive simulation studies were carried out to asses the performance of the proposed model's estimators under finite sample setting. The proposed model was applied to a right-censored cancer data set.

In chapter 5, we have proposed accelerated failure time model in parametric context using generalized log-logistic baseline and the model parameters are estimated using both Bayesian (via non-inforamtive priors) and frequentist approach (via MLE). Extensive simulation studies were carried out to assess the performance of the proposed model's estimators under finite sample setting. The proposed model was applied to a right-censored larynx cancer patients data.

In chapter 6, we have introduced a flexible Bayesian parametric accelerated hazard model with generalized log-logistic baseline hazard. The proposed model's parameters are estimated using both classical and Bayesian inference. An extensive simulation study was carried out to assess the performance of the proposed model's estimators. The proposed model was well demonstrated using right-censored cancer data set with crossover survival curves that the PH and AFT models arenot appropriate to model.

Furthermore, in Chapter 7, We have extended the proposed models in chapters 4,5, and 6 to a general class of hazard-based regression models in parametric context. The proposed model's parameters were estimated using both Classical (via MLE), and Bayesian approach via non-informative priors. Simulation studies were carried out to assess the proposed model's estimators and the proposed model was applied to real-life right-censored cancer data set.

Finally, in Chapter 8, we combined the models described in Chapters 4, 5, 6, and 7, as well as three more prevalent techniques, proportional odds, accelerated odds

and general odds models, to build a univeral class of survival regression models in a parametric framework named Amoud Class. The parameters of the proposed class were estimated using both Bayesian and frequentist approaches, and the nested structure of the Amoud class was evaluated using the likelihood ration test. The proposed model was applied to real-life right-censored cancer data set.

9.3 Contributions

The initial goal was to provide simply a flexible Bayesian parametric AFT model with a generalized log-logistic (GLL) baseline distribution, but we recognize that with a larger number of parametric hazard-based regression models, there would be a bigger range of applicability. Since we obtained the formulation, interpretation, inference, and applications for each parametric hazard-based regression model using GLL baseline distribution.

This thesis contributes to the statistical science with some novel parametric hazard-based and odds-based regression models:

i. The Bayesian parametric PH model with GLL baseline hazard is the first contributed model.

ii. The second contribution is a computational study in which we created an R package called "CoxSurv" to fit a flexible parametric proportional hazard (PH) utilizing at least 15 distinct baseline distributions that can accommodate diverse hazard rate shapes and mixed effects proportional hazard (MEPH) models for the application of clustered and spatial survival data sets.

iii. The third contribution comes from the AFT, in which we applied both Bayesian and frequentist techniques and used the GLL as a baseline hazard.

iv. The fourth contribution comes from the AH framework, where we used the GLL as a baseline hazard, similar to the previous two. We used both Bayesian

and classical inferences. Whereas the PH and AFT cannot be used to examine survival data with crossing survival curves, the AH model can.

v. The fifth contribution is based on computational work in which we created an R package named "AHSurv" to fit flexible parametric accelerated hazards (AH) regression models with 13 different baseline hazard distributions for both net and relative survival frameworks.

vi. The sixth contribution stems from the extension and modification of the previous three (PH, AFT, and AH) models into a nested general class of hazard-based regression models that all use the same GLL baseline distribution.

vii. The seventh contribution is also based on computational work in which we created an R package named "AmoudSurv" to fit tractable parametric odds-based regression models, including the proportional odds, accelerated failure time, and two novel models developed in this study in Chapter 8, named accelereted odds (AO), and general odds (GO) models with 17 different baseline distributions that can accommodate diverse hazard rate shapes. This package enables other researchers to use the approaches in their own datasets that are related to the dissertation aims.

viii. Last but not least, we looked for a universal class named Amoud Class that nested most of the survival regression models, including hazard-based regression (PH, AH, AFT, and GH) models, odds-based regression (PO, AO, AFT, GO) models, to make it easier to compare them and choose which one is acceptable for the different survival data sets..

9.4 Future Work

This thesis has attempted to relax the assumptions in Section 1.3. We would suggest some simple modifications to the parametric hazard-based regression models we

employed in the thesis that may give place to future work. These, as well as some additional extensions are summarized below:

i. The data sets used in this dissertation were right-censored samples. However, other types of censoring data sets may arise in different disciplines of studies. Therefore, further studies should be consider the use of other types of censored data in demonstrating the applications of the developed hazard-based regression models.

ii. Another important area is when we have more than one event times. In many occasions, there may arise multiple causes of failure, which can be observed along with covariates. Competing risks models can be employed in such contexts. Hence, the proposed hazard-based regression models can be extended to this setup, which merits a future research.

iii. The paradigm described in Chapter 3 is applicable to other distributions. In fact, we recommend the development of new flexible distributions using parameter induction technique. Also, assessing the flexibility of the proposed distribution using censored data point out another research line.

iv. The parametric hazard-based regression models which we have considered in the chapters 4, 5, 6, 7, and 8 can be extended by choosing other modified baseline distributions that can incorporate all the basic shapes for the hazard rate function.

v. Chapter 8, suggests a couple of extensions. For instance, the study of other general parametric classes of hazard-based and odds-based regression models.

vi. Another area for future work could be using the reversed hazard rate function instead of hazard rate. The reverse hazard rate function is applicable for modeling and analysis of left censored survival data sets.

vii. In many applications of time-to-event data analysis, the patients are either treated at various medical facilities or are part of various clusters based on geographic or administrative regions. Accounting for between-cluster variability is necessary for the study of such data. Ignoring this variability would force the analysis to rely on irrational assumptions and might have an impact on how the statistical models are influenced. Consequently, another future research avenue could be the development of novel parametric mixed-effects hazard- and odds-based models, which are especially well suited for the study of clustered survival data.

viii. Throughout this dissertation, we assumed that the survival time variate and censoring interval are independently distributed, given covariates. In practice, this assumption may seem to be restrictive and in such instances one can relax this requirement, which leads to a dependent setup. The parametric hazard-based regression models for right-censored data under such a dependent setup can be explored in future research.

Bibliography

Abbas, K. and Tang, Y. (2016). Objective bayesian analysis for log-logistic distribution. *Communications in Statistics-Simulation and Computation*, 45(8):2782–2791.

Adeyinka, F. S. and Olapade, A. K. (2019). On transmuted four parameters generalized log-logistic distribution. *International Journal of Statistical Distributions and Applications*, 5(2):32–37.

Ahmad, Z., Hamedani, G. G., and Butt, N. S. (2019). Recent developments in distribution theory: a brief survey and some new generalized classes of distributions. *Pakistan Journal of Statistics and Operation Research*, pages 87–110.

Akhtar, M., Khan, A., Akhtar, M., and Khan, A. (2014). A. log-logistic distribution as a reliability model: A bayesian analysis. *American Journal of Mathematics and Statistics*, 4(3):162–170.

Al-Aziz, S. N., Muse, A. H., Jawad, T. M., Sayed-Ahmed, N., Aldallal, R., and Yusuf, M. (2022). Bayesian inference in a generalized log-logistic proportional hazards model for the analysis of competing risk data: An application to stem-cell transplanted patients data. *Alexandria Engineering Journal*, 61(12):13035–13050.

Al-Shomrani, A. A., Shawky, A., Arif, O. H., and Aslam, M. (2016). Log-logistic distribution for survival data analysis using mcmc. *SpringerPlus*, 5(1):1–16.

Aldahlan, M. A. (2020). Alpha power transformed log-logistic distribution with application to breaking stress data. *Advances in Mathematical Physics*, 2020.

Alfaer, N. M., Gemeay, A. M., Aljohani, H. M., and Afify, A. Z. (2021). The extended log-logistic distribution: inference and actuarial applications. *Mathematics*, 9(12):1386.

Alkhairy, I., Nagy, M., Muse, A. H., and Hussam, E. (2021). The arctan-x family of distributions: Properties, simulation, and applications to actuarial sciences. *Complexity*, 2021.

Alvares, D., Lázaro, E., Gómez-Rubio, V., and Armero, C. (2021). Bayesian survival analysis with bugs. *Statistics in Medicine*, 40(12):2975–3020.

Alvares, D. and Rubio, F. J. (2021). A tractable bayesian joint model for longitudinal and survival data. *Statistics in Medicine*, 40(19):4213–4229.

Argyropoulos, C. and Unruh, M. L. (2015). Analysis of time to event outcomes in randomized controlled trials by generalized additive models. *PLoS One*, 10(4):e0123784.

Aryal, G. R. (2013). Transmuted log-logistic distribution. *Journal of Statistics Applications & Probability*, 2(1):11.

Ashraf-Ul-Alam, M. and Khan, A. A. (2021). Generalized topp-leone-weibull aft modelling: A bayesian analysis with mcmc tools using r and stan. *Austrian Journal of Statistics*, 50(5):52–76.

Austin, P. C. (2012). Generating survival times to simulate cox proportional hazards models with time-varying covariates. *Statistics in medicine*, 31(29):3946–3958.

Balakrishnan, N., Barmalzan, G., and Haidari, A. (2018). Modified proportional hazard rates and proportional reversed hazard rates models via marshall-olkin distribution and some stochastic comparisons. *Journal of the Korean Statistical Society*, 47(1):127–138.

Banerjee, T., Chen, M.-H., Dey, D. K., and Kim, S. (2007). Bayesian analysis of generalized odds-rate hazards models for survival data. *Lifetime data analysis*, 13(2):241–260.

Barlow, R. E. and Doksum, K. A. (1972). Isotonic tests for convex orderings. In *Proceedings of the sixth Berkeley Symposium on mathematical statistics and probability*, volume 1, pages 293–323. Univ of California Press.

Basharat, N. (2019). *Joint analysis of longitudinal and time to event data using accelerated failure time models: A Bayesian approach*. PhD thesis, University of Saskatchewan.

Bender, R., Augustin, T., and Blettner, M. (2005). Generating survival times to simulate cox proportional hazards models. *Statistics in medicine*, 24(11):1713–1723.

Bennett, S. (1983a). Analysis of survival data by the proportional odds model. *Statistics in medicine*, 2(2):273–277.

Bennett, S. (1983b). Log-logistic regression models for survival data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 32(2):165–171.

Block, H. W., Savits, T. H., and Singh, H. (1998). The reversed hazard rate function. *Probability in the Engineering and Informational Sciences*, 12(1):69–90.

Bogaerts, K., Komárek, A., and Lesaffre, E. (2017). *Survival analysis with interval-censored data: A practical approach with examples in R, SAS, and BUGS*. Chapman and Hall/CRC.

Breslow, N. E., Edler, L., and Berger, J. (1984). A two-sample censored-data rank test for acceleration. *Biometrics*, pages 1049–1062.

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika*, 66(3):429–436.

Burke, K. (2014). Multi-parameter regression survival models.

Burnham, K. P. and Anderson, D. R. (2004). Multimodel inference: understanding aic and bic in model selection. *Sociological methods & research*, 33(2):261–304.

Burr, I. W. (1942). Cumulative frequency functions. *The Annals of mathematical statistics*, 13(2):215–232.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, 76(1).

Chaudhary, A. K. (2019). Bayesian analysis of two-parameter exponentiated log-logistic distribution. *Pravaha*, 25(1):1–12.

Chen, Y., Hanson, T., and Zhang, J. (2014). Accelerated hazards model based on parametric families generalized with bernstein polynomials. *Biometrics*, 70(1):192–201.

Chen, Y. Q. and Jewell, N. P. (2001). On a general class of semiparametric hazards regression models. *Biometrika*, 88(3):687–702.

Chen, Y. Q., Jewell, N. P., and Yang, J. (2003). Accelerated hazards model: method, theory and applications. *Handbook of Statistics*, 23:431–441.

Chen, Y. Q. and Wang, M.-C. (2000). Analysis of accelerated hazards models. *Journal of the American Statistical Association*, 95(450):608–618.

Chesneau, C. and Jamal, F. (2020). The sine kumaraswamy-g family of distributions. *Journal of Mathematical Extension*, 15.

Christensen, R., Johnson, W., Branscum, A., and Hanson, T. E. (2010). *Bayesian ideas and data analysis: an introduction for scientists and statisticians*. CRC press.

Ciampi, A. and Etezadi-Amoli, J. (1985). A general model for testing the proportional hazards and the accelerated failure time hypotheses in the analysis of censored survival data with covariates. *Communications in Statistics-Theory and Methods*, 14(3):651–667.

Co, C. A. T. (2010). Investigating the use of the accelerated hazards model for survival analysis.

Collett, D. (2015). *Modelling survival data in medical research*. CRC press.

Colosimo, E. A. and Giolo, S. R. (2006). *Análise de sobrevivência aplicada*. Editora Blucher.

Consul, J. I. (2016). *a Flexible Baysian modelling of covariate effects on survival*. PhD thesis, Newcastle University.

Cordeiro, G. M., Silva, R. B., and Nascimento, A. D. (2020). Recent advances in lifetime and reliability models.

Cowles, M. K. and Carlin, B. P. (1996). Markov chain monte carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, 91(434):883–904.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202.

Cox, D. R. and Oakes, D. (2018). *Analysis of survival data*. Chapman and Hall/CRC.

Crowther, M. J. (2014). *Development and application of methodology for the parametric analysis of complex survival and joint longitudinal-survival data in biomedical research*. PhD thesis, University of Leicester.

Danish, M. Y. (2016). *Bayesian Analysis of Some Randomly Censored Lifetime Distributions*. PhD thesis, ALLAMA IQBAL OPEN UNIVERSITY, ISLAMABAD.

Danish, M. Y. and Arshad, I. A. (2017). Parametric survival analysis assuming the proportional hazard functions of survival and censoring times distributions. *Journal of Statistics*, 24(1).

Davis, A. (2018). *Modelling techniques for time-to-event data analysis*. PhD thesis, University of Bath.

Demarqui, F. et al. (2022). Semiparametric modeling for multivariate survival data via copulas. *arXiv preprint arXiv:2203.03325*.

Demarqui, F. N. and Mayrink, V. D. (2021). Yang and prentice model with piecewise exponential baseline distribution for modeling lifetime data with crossing survival curves. *Brazilian Journal of Probability and Statistics*, 35(1):172–186.

Demarqui, F. N., Mayrink, V. D., and Ghosh, S. K. (2019). An unified semiparametric approach to model lifetime data with crossing survival curves. *arXiv preprint arXiv:1910.04475*.

Denwood, M. J. (2016). runjags: An r package providing interface utilities, model templates, parallel computing methods and additional distributions for mcmc models in jags. *Journal of statistical software*, 71:1–25.

Dey, S., Kumar, D., Anis, M. Z., Nadarajah, S., and Okorie, I. (2021). A review of transmuted distributions. *Journal of the Indian Society for Probability and Statistics*, 22(1):47–111.

Diao, G., Zeng, D., and Yang, S. (2013). Efficient semiparametric estimation of short-term and long-term hazard ratios with right-censored data. *Biometrics*, 69(4):840–849.

Dos Santos, C. A., Granzotto, D. C., Tomazella, V. L., and Louzada, F. (2018). Hierarchical transmuted log-logistic model: a subjective bayesian analysis. *Journal of Risk and Financial Management*, 11(1):13.

Egge, K. and Zahl, P.-H. (1999). Survival of glaucoma patients. *Acta Ophthalmologica Scandinavica*, 77(4):397–401.

Elsayed, E. A., Liao, H., and Wang, X. (2006). An extended linear hazard regression model with application to time-dependent dielectric breakdown of thermal oxides. *Iie Transactions*, 38(4):329–340.

Etezadi-Amoli, J. and Ciampi, A. (1987). Extended hazard regression for censored survival data with covariates: a spline approximation for the baseline hazard function. *Biometrics*, pages 181–192.

Feng, C., Han, Y., Wu, P., and Tu, X. (2012). Survival analysis. In *Modern Clinical Trial Analysis*, pages 1–23. Springer.

Fernández-i Marín, X. (2016). ggmcmc: Analysis of mcmc samples and bayesian inference. *Journal of Statistical Software*, 70:1–20.

Gelman, A. (1996). Inference and monitoring convergence, markov chain monte carlo in practice, ed. gilks r, richardson s, spiegelhalter dj.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian data analysis*. Chapman and Hall/CRC.

Gelman, A. et al. (2013). Bayesian data analysis gelman. *J Chem Inf Model*, 53:1689–99.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, pages 457–472.

Geweke, J. F. et al. (1991). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Technical report, Federal Reserve Bank of Minneapolis.

Gilks, W. R., Thomas, A., and Spiegelhalter, D. J. (1994). A language and program for complex bayesian modelling. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 43(1):169–177.

Gompertz, B. (1825). Xxiv. on the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. in a letter to francis baily, esq. frs &c. *Philosophical transactions of the Royal Society of London*, (115):513–583.

Granzotto, D. C. T. and Louzada, F. (2015). The transmuted log-logistic distribution: modeling, inference, and an application to a polled tabapua race time up to first calving data. *Communications in Statistics-Theory and Methods*, 44(16):3387–3402.

Gu, E. (2020). *Flexible Regression Models for Survival Data*. PhD thesis, University of South Carolina.

Gui, W. (2013). Marshall-olkin extended log-logistic distribution and its application in minification processes. *Appl Math Sci*, 7(80):3947–3961.

Gupta, R. C. and Wu, H. (2001). Analyzing survival data by proportional reversed hazard model. *International Journal of Reliability and Applications*, 2(1):1–26.

Guure, C. B., Ibrahim, N. A., Dwomoh, D., and Bosomprah, S. (2015). Bayesian statistical inference of the loglogistic model with interval-censored lifetime data. *Journal of Statistical Computation and Simulation*, 85(8):1567–1583.

Hamedani, G. (2013). The zografos-balakrishnan log-logistic distribution: Properties and applications. *Journal of Statistical Theory and Applications*.

Heidelberger, P. and Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, 31(6):1109–1144.

Hjort, N. L. (1992). On inference in parametric survival data models. *International Statistical Review/Revue Internationale de Statistique*, pages 355–387.

Huang, T., Elsayed, E., and Jiang, T. (2008). An alt proportional hazard-proportional odds model. In *The Proceedings of the 14th ISSAT International Conference on Reliability and Quality in Design*.

Huang, T. and Jiang, T. (2009). An extended proportional hazards-proportional odds model in accelerated life testing. In *2009 8th International Conference on Reliability, Maintainability and Safety*, pages 1173–1176. IEEE.

Ibrahim, J. G., Chen, M.-H., Sinha, D., Ibrahim, J., and Chen, M. (2001). *Bayesian survival analysis*, volume 2. Springer.

Jammalamadaka, S. R. and Mangalam, V. (2003). Nonparametric estimation for middle-censored data. *Journal of nonparametric statistics*, 15(2):253–265.

Kalbfleisch, J. D. (1978). Non-parametric bayesian analysis of survival time data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 40(2):214–221.

Kalbfleisch, J. D. and Prentice, R. L. (1973). Marginal likelihoods based on cox's regression and life model. *Biometrika*, 60(2):267–278.

Kalbfleisch, J. D. and Prentice, R. L. (2011). *The statistical analysis of failure time data*. John Wiley & Sons.

Kang, S. G., Lee, K., and Lee, W. D. (2014). Noninformative priors for the generalized half-normal distribution. *Journal of the Korean Statistical Society*, 43(1):19–29.

Kardaun, O. (1983). Statistical survival analysis of male larynx-cancer patients-a case study. *Statistica neerlandica*, 37(3):103–125.

Kayid, M. (2021). Further results on the proportional vitalities model. *Entropy*, 23(9):1201.

Khan, S. A. (2018). Exponentiated weibull regression for time-to-event data. *Lifetime data analysis*, 24(2):328–354.

Khan, S. A. and Basharat, N. (2021). Accelerated failure time models for recurrent event data analysis and joint modeling. *Computational Statistics*, pages 1–29.

Khan, S. A. and Basharat, N. (2022). Accelerated failure time models for recurrent event data analysis and joint modeling. *Computational Statistics*, 37(4):1569–1597.

Khan, S. A. and Khosa, S. K. (2016). Generalized log-logistic proportional hazard model with applications in survival analysis. *Journal of Statistical Distributions and Applications*, 3(1):1–18.

Khosa, S. K. (2019). *Parametric Proportional Hazard Models with Applications in Survival analysis*. PhD thesis, University of Saskatchewan.

Klein, J. P., Van Houwelingen, H. C., Ibrahim, J. G., and Scheike, T. H. (2014). *Handbook of survival analysis*. CRC Press Boca Raton, FL:.

Komárek, A. and Lesaffre, E. (2008). Bayesian accelerated failure time model with multivariate doubly interval-censored data and flexible distributional assumptions. *Journal of the American Statistical Association*, 103(482):523–533.

Laurie, J. A., Moertel, C. G., Fleming, T. R., Wieand, H. S., Leigh, J. E., Rubin, J., McCormack, G. W., Gerstner, J. B., Krook, J. E., and Malliard, J. (1989). Surgical adjuvant therapy of large-bowel carcinoma: an evaluation of levamisole and the combination of levamisole and fluorouracil. the north central cancer treatment group and the mayo clinic. *Journal of Clinical Oncology*, 7(10):1447–1456.

Lawless, J. F. (2011). *Statistical models and methods for lifetime data*. John Wiley & Sons.

Lázaro, E., Armero, C., and Alvares, D. (2021). Bayesian regularization for flexible baseline hazard functions in cox survival models. *Biometrical Journal*, 63(1):7–26.

Lázaro Hervás, E. (2018). Flexible bayesian survival models with application in biometric studies.

Lee, E. T. and Wang, J. (2003). *Statistical methods for survival data analysis*, volume 476. John Wiley & Sons.

Lee, S.-H. (2009). Some estimators and tests for accelerated hazards model using weighted cumulative hazard difference. *Journal of Applied Statistics*, 36(5):473–482.

Lee, S.-H. (2016). On the estimators and tests for the semiparametric hazards regression model. *Lifetime data analysis*, 22(4):531–546.

Leemis, L. M., Shih, L.-H., and Reynertson, K. (1990). Variate generation for accelerated life and proportional hazards models with time dependent covariates. *Statistics & probability letters*, 10(4):335–339.

Legrand, C. (2021). *Advanced Survival Models*. Chapman and Hall/CRC.

Lemonte, A. J. (2014). The beta log-logistic distribution. *Brazilian journal of probability and statistics*, 28(3):313–332.

Li, L., Hanson, T., and Zhang, J. (2015). Spatial extended hazard model with application to prostate cancer survival. *Biometrics*, 71(2):313–322.

Lima, S. R. and Cordeiro, G. M. (2017). The extended log-logistic distribution: properties and application. *Anais da Academia Brasileira de Ciências*, 89:03–17.

Lin, D. (1994). Cox regression analysis of multivariate failure time data: the marginal approach. *Statistics in medicine*, 13(21):2233–2247.

Louzada-Neto, F. (1997). Extended hazard regression model for reliability and survival analysis. *Lifetime Data Analysis*, 3(4):367–381.

Louzada-Neto, F. (2001). Bayesian analysis for hazard models with non-constant shape parameter. *Computational Statistics*, 16(2):243–254.

Magnusson, M., Vehtari, A., Jonasson, J., and Andersen, M. (2020). Leave-one-out cross-validation for bayesian model comparison in large data. In *International Conference on Artificial Intelligence and Statistics*, pages 341–351. PMLR.

Mahmood, Z., M Jawa, T., Sayed-Ahmed, N., Khalil, E., Muse, A. H., and Tolba, A. H. (2022). An extended cosine generalized family of distributions for reliability modeling: Characteristics and applications with simulation study. *Mathematical Problems in Engineering*, 2022.

Malik, A. S. and Ahmad, S. (2020). An extension of log-logistic distribution for analyzing survival data. *Pakistan Journal of Statistics and Operation Research*, pages 789–801.

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC.

Mendoza, N. V., Ortega, E. M., and Cordeiro, G. M. (2016). The exponentiated-log-logistic geometric distribution: Dual activation. *Communications in Statistics-Theory and Methods*, 45(13):3838–3859.

Midhu, N., Sankaran, P., and Nair, N. U. (2013). A class of distributions with the linear mean residual quantile function and it's generalizations. *Statistical Methodology*, 15:1–24.

Moertel, C. G., Fleming, T. R., Macdonald, J. S., Haller, D. G., Laurie, J. A., Goodman, P. J., Ungerleider, J. S., Emerson, W. A., Tormey, D. C., Glick, J. H., et al. (1990). Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. *New England Journal of Medicine*, 322(6):352–358.

Moertel, C. G., Fleming, T. R., Macdonald, J. S., Haller, D. G., Laurie, J. A., Tangen, C. M., Ungerleider, J. S., Emerson, W. A., Tormey, D. C., Glick, J. H., et al. (1995). Fluorouracil plus levamisole as effective adjuvant therapy after resection of stage iii colon carcinoma: a final report. *Annals of internal medicine*, 122(5):321–326.

Mok, T. S., Wu, Y.-L., Thongprasert, S., Yang, C.-H., Chu, D.-T., Saijo, N., Sunpaweravong, P., Han, B., Margono, B., Ichinose, Y., et al. (2009). Gefitinib or carboplatin–paclitaxel in pulmonary adenocarcinoma. *New England Journal of Medicine*, 361(10):947–957.

Mudholkar, G. S. and Hutson, A. D. (1996). The exponentiated weibull family: some properties and a flood data application. *Communications in Statistics–Theory and Methods*, 25(12):3059–3083.

Mudholkar, G. S. and Srivastava, D. K. (1993). Exponentiated weibull family for analyzing bathtub failure-rate data. *IEEE transactions on reliability*, 42(2):299–302.

Muse, A. H., Chesneau, C., Ngesa, O., and Mwalili, S. (2022). Flexible parametric accelerated hazard model: Simulation and application to censored lifetime data with crossing survival curves. *Mathematical and Computational Applications*, 27(6):104.

MUSE, A. H., MWALILI, S., and NGESA, O. (2022). Amoud-gfamily of lifetime distributions: Properties, hazard-based regression models and applications to survival data. *Engineering (ICMASE 2022) 4-7 July 2022, Technical University of Civil Engineering Bucharest, Romania*, page 78.

Muse, A. H., Mwalili, S., Ngesa, O., Almalki, S. J., and Abd-Elmougod, G. A. (2021a). Bayesian and classical inference for the generalized log-logistic distribution with applications to survival data. *Computational intelligence and neuroscience*, 2021.

Muse, A. H., Mwalili, S., Ngesa, O., Alshanbari, H. M., Khosa, S. K., and Hussam, E. (2022a). Bayesian and frequentist approach for the generalized log-logistic accelerated failure time model with applications to larynx-cancer patients. *Alexandria Engineering Journal*, 61(10):7953–7978.

Muse, A. H., Mwalili, S., Ngesa, O., and Chesneau, C. (2022b). 'amoudsurv:an r package for tractable parametric odds-based regression models'. https://cran.r-project.org/web//packages/AmoudSurv/index.html.

Muse, A. H., Mwalili, S., Ngesa, O., Chesneau, C., Al-Bossly, A., and El-Morshedy, M. (2022c). Bayesian and frequentist approaches for a tractable parametric general class of hazard-based regression models: An application to oncology data. *Mathematics*, 10(20):3813.

Muse, A. H., Mwalili, S., Ngesa, O., Chesneau, C., Alshanbari, H. M., and El-Bagoury, A.-A. H. (2022d). Amoud class for hazard-based and odds-based regression models: Application to oncology studies. *Axioms*, 11(11):606.

Muse, A. H., Mwalili, S., Ngesa, O., and Kilai, M. (2022e). 'ahsurv:an r package for flexible parametric accelerated hazards (ah) regression models'. https://cran.r-project.org/web/packages/AHSurv/index.html.

Muse, A. H., Mwalili, S., Ngesa, O., and Kilai, M. (2022f). 'ahsurv:an r package for flexible parametric accelerated hazards (ah) regression models'. *CRAN*.

Muse, A. H., Mwalili, S. M., Ngesa, O., et al. (2021b). On the log-logistic distribution and its generalizations: a survey. *International Journal of Statistics and Probability*, 10(3):93.

Muse, A. H., Ngesa, O., Mwalili, S., Alshanbari, H. M., and El-Bagoury, A.-A. H. (2022g). A flexible bayesian parametric proportional hazard model: Simulation and applications to right-censored healthcare data. *Journal of Healthcare Engineering*, 2022.

Muse, A. H., Tolba, A. H., Fayad, E., Abu Ali, O. A., Nagy, M., and Yusuf, M. (2021c). Modelling the covid-19 mortality rate with a new versatile modification of the log-logistic distribution. *Computational Intelligence and Neuroscience*, 2021.

Mwalili, S. M. (2006). Bayesian and frequentist approaches to correct for misclassification error with applications to caries research.

Nasiru, S. (2018). *A new generalization of transformed-transformer family of distributions*. PhD thesis, JKUAT.

Nasiru, S., Mwita, P. N., and Ngesa, O. (2019). Alpha power transformed frechet distribution. *Applied Mathematics & Information Sciences*, 13(1):129–141.

Oakes, D. and Dasu, T. (1990). A note on residual life. *Biometrika*, 77(2):409–410.

Olosunde, A. A. and EJIOFOR, C. (2021). Log-exponential power distribution for accelerated failure time model in survival analysis and its application. *Afrika Statistika*, 16(1):2587–2603.

Oluyede, B., Foya, S., Warahena-Liyanage, G., and Huang, S. (2016). The log-logistic weibull distribution with applications to lifetime data. *Austrian Journal of Statistics*, 45(3):43–69.

Owino, N. O. (2014). *Bayesian spatial models with application to HIV, TB and STI modeling in Kenya.* PhD thesis.

Panaro, R. V. (2020). spsurv: An r package for semi-parametric survival analysis. *arXiv preprint arXiv:2003.10548*.

Pang, M. (2020). *Flexible Accelerated Failure Time modeling of multivariable time-to-event data*. McGill University (Canada).

Plummer, M. et al. (2003). Jags: A program for analysis of bayesian graphical models using gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*, volume 124, pages 1–10. Vienna, Austria.

Plummer, M., Stukalov, A., Denwood, M., and Plummer, M. M. (2019). Package 'rjags'. *update*, 1.

Prasad, S. and Sankaran, P. (2017). *Regression Models for the Analysis of Middle-censored Lifetime Data*. PhD thesis, Cochin University of Science and Technology.

Prentice, R. L. (1976). A generalization of the probit and logit methods for dose response curves. *Biometrics*, pages 761–768.

Putter, H., Sasako, M., Hartgrink, H., Van de Velde, C., and Van Houwelingen, J. (2005). Long-term survival with non-proportional hazards: Results from the dutch gastric cancer trial. *Statistics in medicine*, 24(18):2807–2821.

Qi, X., Zhou, S., and Plummer, M. (2020). A note on bayesian modeling specification of censored data in jags. *arXiv preprint arXiv:2012.02074*.

Qing Chen, Y. (2001). Accelerated hazards regression model and its adequacy for censored survival data. *Biometrics*, 57(3):853–860.

Raftery, A. E. and Lewis, S. M. (1992). [practical markov chain monte carlo]: comment: one long run with diagnostics: implementation strategies for markov chain monte carlo. *Statistical science*, 7(4):493–497.

Raftery, A. E. and Lewis, S. M. (1995). The number of iterations, convergence diagnostics and generic metropolis algorithms. *Practical Markov Chain Monte Carlo*, 7(98):763–773.

Ramos, P. L. (2018). Bayesian and classical inference for the generalized gamma distribution and related models.

Ramos, P. L., Moala, F. A., and Achcar, J. A. (2014). Objective priors for estimation of extended exponential geometric distribution. *Journal of Modern Applied Statistical Methods*, 13(2):12.

Reath, J., Dong, J., and Wang, M. (2018). Improved parameter estimation of the log-logistic distribution with applications. *Computational Statistics*, 33(1):339–356.

Rezaei, S., Hashami, S., and Najjar, L. (2014). Extended exponential geometric proportional hazard model. *Annals of Data Science*, 1(2):173–189.

Rosaiah, K., Nagarjuna, K. M., Kumar, D., and Rao, B. S. (2014). Exponential—log logistic additive failure rate model. *International Journal of Scientific and Research Publications*, 4(1):2250–3153.

Royston, P. and Parmar, M. K. (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in medicine*, 21(15):2175–2197.

Rubio, F. J., Alvares, D., Redondo-Sanchez, D., Marcos-Gragera, R., Sánchez, M.-J., and Luque-Fernandez, M. A. (2022). Bayesian variable selection and survival modeling: assessing the most important comorbidities that impact lung and colorectal cancer survival in spain. *BMC Medical Research Methodology*, 22(1):1–14.

Rubio, F. J., Rachet, B., Giorgi, R., Maringe, C., and Belot, A. (2021). On models for the estimation of the excess mortality hazard in case of insufficiently stratified life tables. *Biostatistics*, 22(1):51–67.

Rubio, F. J., Remontet, L., Jewell, N. P., and Belot, A. (2019). On a general structure for hazard-based regression models: an application to population-based cancer research. *Statistical methods in medical research*, 28(8):2404–2417.

Rubio Alvarez, F. J. and Drikvandi, R. (2022). Megh: A parametric class of general hazard models for clustered survival data. *Statistical Methods in Medical Research*.

Schruben, L., Singh, H., and Tierney, L. (1983). Optimal tests for initialization bias in simulation output. *Operations Research*, 31(6):1167–1178.

Schruben, L. W. (1982). Detecting initialization bias in simulation output. *Operations Research*, 30(3):569–590.

Shakhatreh, M. K. (2018). A new three-parameter extension of the log-logistic distribution with applications to survival data. *Communications in Statistics-Theory and Methods*, 47(21):5205–5226.

Shrahili, M., Albabtain, A. A., Kayid, M., and Kaabi, Z. (2020). Stochastic aspects of proportional vitalities model. *Mathematics*, 8(10):1823.

Shyur, H.-J., Elsayed, E., and Luxhøj, J. T. (1999). A general model for accelerated life testing with time-dependent covariates. *Naval Research Logistics (NRL)*, 46(3):303–321.

Singh, K. P., Lee, C. M.-S., and George, E. O. (1988). On generalized log-logistic model for censored survival data. *Biometrical Journal*, 30(7):843–850.

Singh, V. (1998). *Entropy-based parameter estimation in hydrology*, volume 30. Springer Science & Business Media.

Smith, A. F. and Roberts, G. O. (1993). Bayesian computation via the gibbs sampler and related markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(1):3–23.

Souza, L., de Oliveira, W. R., de Brito, C. C. R., Chesneau, C., Fernandes, R., and Ferreira, T. A. (2022). Sec-g class of distributions: Properties and applications. *Symmetry*, 14(2):299.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van der Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(3):485–493.

Tahir, M., Mansoor, M., Zubair, M., and Hamedani, G. (2014). Mcdonald log-logistic distribution with an application to breast cancer data. *Journal of Statistical Theory and Applications*.

Tahir, M. H. and Cordeiro, G. M. (2016). Compounding of distributions: a survey and new generalized classes. *Journal of Statistical Distributions and Applications*, 3(1):1–35.

Tahir, M. H. and Nadarajah, S. (2015). Parameter induction in continuous univariate distributions: Well-established g families. *Anais da Academia Brasileira de Ciências*, 87:539–568.

Team, R. C. (2019). 2020. *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria: Available at: https://www. R-project. org/.[Google Scholar]*.

Team, R. C. et al. (2013). R: A language and environment for statistical computing.

Therneau, T. and Lumley, T. (2013). R survival package. *R Core Team*.

Tong, X., Zhu, L., Leng, C., Leisenring, W., and Robison, L. L. (2013). A general semiparametric hazards regression model: efficient estimation and structure selection. *Statistics in medicine*, 32(28):4980–4994.

Turkman, M. A. A., Paulino, C. D., and Müller, P. (2019). *Computational Bayesian Statistics: An Introduction*, volume 11. Cambridge University Press.

van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J., et al. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers*, 1(1):1–26.

Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and computing*, 27(5):1413–1432.

Vehtari, A., Gelman, A., Gabry, J., and Yao, Y. (2021). Package 'loo'. *Efficient Leave-One-Out Cross-Validation and WAIC for Bayesian Models*.

Vines, S., Gilks, W., and Wild, P. (1996). Fitting bayesian multiple random effects models. *Statistics and Computing*, 6(4):337–346.

Wang, K., Ye, X., and Ma, J. (2018a). An empirical analysis of post-work grocery shopping activity duration using modified accelerated failure time model to differentiate time-dependent and time-independent covariates. *PloS one*, 13(11):e0207810.

Wang, X., Yue, Y., and Faraway, J. J. (2018b). *Bayesian regression modeling with INLA*. Chapman and Hall/CRC.

Watanabe, S. (2013). A widely applicable bayesian information criterion. *Journal of Machine Learning Research*, 14(27):867–897.

Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of applied mechanics*.

Yahaya, A. and Dewu, M. M. (2017). Bayesian estimation of scale parameter of the log-logistic distribution under the assumption of chi-square and maxwell priors. *ATBU Journal of Science, Technology and Education*, 4(3):39–46.

Yang, S. and Prentice, R. (2005). Semiparametric analysis of short-term and long-term hazard ratios with two-sample survival data. *Biometrika*, 92(1):1–17.

Zhang, H., Wang, P., and Sun, J. (2018). Regression analysis of interval-censored failure time data with possibly crossing hazards. *Statistics in medicine*, 37(5):768–775.

Zhang, J., Hanson, T., and Zhou, H. (2019). Bayes factors for choosing among six common survival models. *Lifetime data analysis*, 25(2):361–379.

Zhang, J. and Peng, Y. (2009). Crossing hazard functions in common survival models. *Statistics & probability letters*, 79(20):2124–2130.

Zhou, H. and Hanson, T. (2015). Bayesian spatial survival models. *Nonparametric Bayesian Inference in Biostatistics*, pages 215–246.

APPENDIX A
Paper 1

The first published manuscript

International Journal of Statistics and Probability; Vol. 10, No. 3; May 2021
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

On the Log-Logistic Distribution and Its Generalizations: A Survey

Abdisalam Hassan Muse[1], Samuel M. Mwalili[1,2], Oscar Ngesa[1,2,3]

[1] Mathematics Department, Pan African University, Institute of Basic Sciences, Technology, and Innovation, Kenya

[2] Statistics and Actuarial Sciences Department, Jomo Kenyatta University of Agriculture and Technology, Kenya

[3] Mathematics and Informatics Department, Taita-Taveta University, Kenya

Correspondence: Abdisalam Hassan Muse, Mathematics Department (Statistics Option), Pan African University, Institute of Basic Sciences, Technology, and Innovation, (PAUSTI), P.O. Box 62000 00200 Nairobi, Kenya.

Received: February 4, 2021 Accepted: April 7, 2021 Online Published: April 20, 2021

doi:10.5539/ijsp.v10n3p93 URL: https://doi.org/10.5539/ijsp.v10n3p93

Abstract

In this paper, we present a review on the log-logistic distribution and some of its recent generalizations. We cite more than twenty distributions obtained by different generating families of univariate continuous distributions or compounding methods on the log-logistic distribution. We reviewed some log-logistic mathematical properties, including the eight different functions used to define lifetime distributions. These results were used to obtain the properties of some log-logistic generalizations from linear representations. A real-life data application is presented to compare some of the surveyed distributions.

Keywords: log-logistic distribution, log-logistic generalizations, generalized classes of distributions, construction of new families, censored data, survival analysis

APPENDIX B
Paper 2

The second published manuscript

Hindawi
Computational Intelligence and Neuroscience
Volume 2021, Article ID 5820435, 24 pages
https://doi.org/10.1155/2021/5820435

Research Article

Bayesian and Classical Inference for the Generalized Log-Logistic Distribution with Applications to Survival Data

Abdisalam Hassan Muse,[1] **Samuel Mwalili**,[2] **Oscar Ngesa**,[3] **Saad J. Almalki**,[4] **and Gamal A. Abd-Elmougod**[5]

[1]*Department of Mathematics (Statistics Option) Programme, Pan African University, Institute for Basic Science, Technology and Innovation (PAUSTI), Nairobi 6200-00200, Kenya*
[2]*Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi 6200-00200, Kenya*
[3]*Department of Mathematics and Physical Sciences, Taita Taveta University, Voi 635-80300, Kenya*
[4]*Department of Mathematics and Statistics, College of Science, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia*
[5]*Department of Mathematics, Faculty of Science, Damanhour University, Damanhour, Egypt*

Correspondence should be addressed to Abdisalam Hassan Muse; abdisalam.hassan@amoud.edu.so

Received 19 August 2021; Accepted 14 September 2021; Published 11 October 2021

Academic Editor: Ahmed Mostafa Khalil

The generalized log-logistic distribution is especially useful for modelling survival data with variable hazard rate shapes because it extends the log-logistic distribution by adding an extra parameter to the classical distribution, resulting in greater flexibility in analyzing and modelling various data types. We derive the fundamental mathematical and statistical properties of the proposed distribution in this paper. Many well-known lifetime special submodels are included in the proposed distribution, including the Weibull, log-logistic, exponential, and Burr XII distributions. The maximum likelihood method was used to estimate the unknown parameters of the proposed distribution, and a Monte Carlo simulation study was run to assess the estimators' performance. This distribution is significant because it can model both monotone and nonmonotone hazard rate functions, which are quite common in survival and reliability data analysis. Furthermore, the proposed distribution's flexibility and usefulness are demonstrated in a real-world data set and compared to its submodels, the Weibull, log-logistic, and Burr XII distributions, as well as other three-parameter parametric survival distributions, such as the exponentiated Weibull distribution, the three-parameter log-normal distribution, the three-parameter (or the shifted) log-logistic distribution, the three-parameter gamma distribution, and an exponentiated Weibull distribution. The proposed distribution is plausible, according to the goodness-of-fit, log-likelihood, and information criterion values. Finally, for the data set, Bayesian inference and Gibb's sampling performance are used to compute the approximate Bayes estimates as well as the highest posterior density credible intervals, and the convergence diagnostic techniques based on Markov chain Monte Carlo techniques were used.

APPENDIX C
Paper 3

The third published manuscript

Hindawi
Journal of Healthcare Engineering
Volume 2022, Article ID 2051642, 28 pages
https://doi.org/10.1155/2022/2051642

Research Article

A Flexible Bayesian Parametric Proportional Hazard Model: Simulation and Applications to Right-Censored Healthcare Data

Abdisalam Hassan Muse,[1] **Oscar Ngesa**,[2] **Samuel Mwalili**,[3] **Huda M. Alshanbari**,[4] **and Abdal-Aziz H. El-Bagoury**[5]

[1]*Department of Mathematics (Statistics Option), Pan African University Institute for Basic Science Technology and Innovation (PAUSTI), Nairobi 62000 00200, Kenya*
[2]*Department of Mathematics and Physical Sciences, Taita Taveta University, Voi 635-80300, Kenya*
[3]*Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi, Kenya*
[4]*Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia*
[5]*Basic Science Department, Higher Institute of Engineering and Technology, El-Mahala El-Kubra, Egypt*

Correspondence should be addressed to Abdisalam Hassan Muse; abdisalam.hassan@amoud.edu.so

Received 2 February 2022; Revised 24 March 2022; Accepted 18 April 2022; Published 2 June 2022

Academic Editor: Mohamed Elhoseny

Survival analysis is a collection of statistical techniques which examine the time it takes for an event to occur, and it is one of the most important fields in biomedical sciences and other variety of scientific disciplines. Furthermore, the computational rapid advancements in recent decades have advocated the application of Bayesian techniques in this field, giving a powerful and flexible alternative to the classical inference. The aim of this study is to consider the Bayesian inference for the generalized log-logistic proportional hazard model with applications to right-censored healthcare data sets. We assume an independent gamma prior for the baseline hazard parameters and a normal prior is placed on the regression coefficients. We then obtain the exact form of the joint posterior distribution of the regression coefficients and distributional parameters. The Bayesian estimates of the parameters of the proposed model are obtained using the Markov chain Monte Carlo (McMC) simulation technique. All computations are performed in Bayesian analysis using Gibbs sampling (BUGS) syntax that can be run with Just Another Gibbs Sampling (JAGS) from the R software. A detailed simulation study was used to assess the performance of the proposed parametric proportional hazard model. Two real-survival data problems in the healthcare are analyzed for illustration of the proposed model and for model comparison. Furthermore, the convergence diagnostic tests are presented and analyzed. Finally, our research found that the proposed parametric proportional hazard model performs well and could be beneficial in analyzing various types of survival data.

APPENDIX D
Paper 4

The fourth published manuscript

Alexandria Engineering Journal (2022) 61, 7953–7978

HOSTED BY

ELSEVIER

Alexandria University

Alexandria Engineering Journal

www.elsevier.com/locate/aej
www.sciencedirect.com

Bayesian and frequentist approach for the generalized log-logistic accelerated failure time model with applications to larynx-cancer patients

Abdisalam Hassan Muse [a,*], Samuel Mwalili [b], Oscar Ngesa [c], Huda M. Alshanbari [d], Saima Khan Khosa [e], Eslam Hussam [f]

[a] *Department of Mathematics (Statistics Option) Programme, Pan African University, Institute for Basic Sciences, Technology and Innovation (PAUSTI), Nairobi 62000-00200, Kenya*
[b] *Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi 62000-00200, Kenya*
[c] *Department of Mathematics and Physical Sciences, Taita Taveta University, Voi 635-80300, Kenya*
[d] *Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, P.O.Box 84428, Riyadh 11671, Saudi Arabia*
[e] *Department of Statistics, Bahauddin Zakariya University, Multan, Pakistan*
[f] *Department of Mathematics, Faculty of Science, Helwan University, Cairo, Egypt*

Received 24 November 2021; revised 27 December 2021; accepted 10 January 2022
Available online 31 January 2022

KEYWORDS

Bayesian inference;
Accelerated failure time model;
Generalized log-logistic distribution;
JAGS;
Hazard-based regression models;
Maximum likelihood estimation;
Censored data

Abstract The log-normal, log-logistic and Weibull distributions are commonly utilized to model survival data. Unimodal (or non-monotone) failure rate functions are modeled using the log-normal and the log-logistic families, whereas monotone failure rate functions are modeled using the Weibull family. The growing availability of survival data with a variety of features encourages statisticians to propose more flexible parametric models that can accommodate both monotone (increasing or decreasing), and non-monotone (unimodal or bathtub) failure rate functions. One such model is the generalized log-logistic distribution which not only accommodates unimodal failure rates but also allows for a monotone and non-monotone failure rate functions. This distribution has shown to have a lot of potential in univariate analysis of survival data. However, many studies are primarily concerned with determining the link between survival time and one or more explanatory variables. This leads to the study of hazard-based regression models in survival and reliability analysis, which can be formulated in a variety of ways. One such method concerns formulating

Abbreviations: AFT, Accelerated failure time; AH, Accelerated Hazard; AIC, Akaike Information Criterion; BUGS, Bayesian analysis Using Gibbs Sampling; CDF, Cumulative distribution function; CHF, Cumulative hazard function; CI, Credibility Intervals; DIC, Deviance Information Criterion; GLL, Generalized log-logistic; HPD, Highest Probability density; HRF, Hazard rate function; JAGS, Just Another Gibbs Sampling; LL, Log-logistic; McMC, Markov chain Monte Carlo; MLE, Maximum likelihood Estimate; MSE, Mean Square Error; PDF, Probability density function; PH, Proportional hazard; PO, Proportional Odds; RM, Relative Medium; SF, Survival Function; TTT, Total time on test

* Corresponding author.
E-mail addresses: abdisalam.hassan@amoud.edu.so (A.H. Muse), hmalshanbari@pnu.edu.sa (H.M. Alshanbari).

Peer review under responsibility of Faculty of Engineering, Alexandria University.

https://doi.org/10.1016/j.aej.2022.01.033
1110-0168

APPENDIX E
Paper 5

The fifth published manuscript

Mathematical and Computational Applications

MDPI

Article

Flexible Parametric Accelerated Hazard Model: Simulation and Application to Censored Lifetime Data with Crossing Survival Curves

Abdisalam Hassan Muse [1,2,*], Christophe Chesneau [3,*], Oscar Ngesa [1,4] and Samuel Mwalili [1,5]

[1] Institute for Basic Sciences, Technology and Innovation (PAUSTI), Pan African University, Nairobi 62000-00200, Kenya
[2] Faculty of Science and Humanities, School of Postgraduate Studies and Research, Amoud University, Borama 25263, Somalia
[3] Department of Mathematics, LMNO, CNRS-Université de Caen, Campus II, Science 3, 14032 Caen, France
[4] Department of Mathematics and Physical Sciences, Taita Taveta University, Voi 635-80300, Kenya
[5] Department of Statistics and Actuarial Science, Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi 62000-00200, Kenya
* Correspondence: abdisalam.hassan@amoud.edu.so (A.H.M.); christophe.chesneau@unicaen.fr (C.C.)

Abstract: This study aims to propose a flexible, fully parametric hazard-based regression model for censored time-to-event data with crossing survival curves. We call it the accelerated hazard (AH) model. The AH model can be written with or without a baseline distribution for lifetimes. The former assumption results in parametric regression models, whereas the latter results in semi-parametric regression models, which are by far the most commonly used in time-to-event analysis. However, under certain conditions, a parametric hazard-based regression model may produce more efficient estimates than a semi-parametric model. The parametric AH model, on the other hand, is inappropriate when the baseline distribution is exponential because it is constant over time; similarly, when the baseline distribution is the Weibull distribution, the AH model coincides with the accelerated failure time (AFT) and proportional hazard (PH) models. The use of a versatile parametric baseline distribution (generalized log-logistic distribution) for modeling the baseline hazard rate function is investigated. For the parameters of the proposed AH model, the classical (via maximum likelihood estimation) and Bayesian approaches using noninformative priors are discussed. A comprehensive simulation study was conducted to assess the performance of the proposed model's estimators. A real-life right-censored gastric cancer dataset with crossover survival curves is used to demonstrate the tractability and utility of the proposed fully parametric AH model. The study concluded that the parametric AH model is effective and could be useful for assessing a variety of survival data types with crossover survival curves.

Keywords: Bayesian inference; hazard-based regression model; survival analysis; accelerated hazard model; generalized log-logistic distribution; crossover survival curves; censored data; maximum likelihood estimation.

check for updates

Citation: Muse, A.H.; Chesneau, C.; Ngesa, O.; Mwalili, S. Flexible Parametric Accelerated Hazard Model: Simulation and Application to Censored Lifetime Data with Crossing Survival Curves. *Math. Comput. Appl.* **2022**, *27*, 104. https://doi.org/10.3390/mca27060104

Academic Editor: Sandra Ferreira

Received: 2 November 2022
Accepted: 29 November 2022
Published: 30 November 2022

APPENDIX F
Paper 6

The sixth published manuscript

mathematics

MDPI

Article

Bayesian and Frequentist Approaches for a Tractable Parametric General Class of Hazard-Based Regression Models: An Application to Oncology Data

Abdisalam Hassan Muse [1,2,*], Samuel Mwalili [3], Oscar Ngesa [4], Christophe Chesneau [5,*], Afrah Al-Bossly [6] and Mahmoud El-Morshedy [6,7]

[1] Institute for Basic Sciences, Technology and Innovation (PAUSTI), Pan African University, Nairobi 62000-00200, Kenya
[2] Faculty of Science and Humanities, School of Postgraduate Studies and Research, Amoud University, Borama 25263, Somalia
[3] Department of Statistics and Actuarial Science, Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi 62000-00200, Kenya
[4] Department of Mathematics and Physical Sciences, Taita Taveta University, Voi 635-80300, Kenya
[5] Department of Mathematics, LMNO, CNRS-Université de Caen, Campus II, Science 3, 14032 Caen, France
[6] Department of Mathematics, College of Science and Humanities in AL-Kharj, Prince Sattam Bin Abdulaziz University, AL-Kharj 11942, Saudi Arabia
[7] Department of Mathematics, Faculty of Science, Mansoura University, Mansoura 35516, Egypt
* Correspondence: abdisalam.hassan@amoud.edu.so (A.H.M.); christophe.chesneau@unicaen.fr (C.C.)

check for updates

Citation: Muse, A.H.; Mwalili, S.; Ngesa, O.; Chesneau, C.; Al-Bossly, A.; El-Morshedy, M. Bayesian and Frequentist Approaches for a Tractable Parametric General Class of Hazard-Based Regression Models: An Application to Oncology Data. *Mathematics* **2022**, *10*, 3813. https://doi.org/10.3390/math10203813

Academic Editors: Araceli Queiruga-Dios, Fatih Yilmaz, Ion Mierlus-Mazilu, Deolinda M. L. Dias Rasteiro and Jesús Martín Vaquero

Received: 19 September 2022
Accepted: 10 October 2022
Published: 15 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract: In this study, we consider a general, flexible, parametric hazard-based regression model for censored lifetime data with covariates and term it the "general hazard (GH)" regression model. Some well-known models, such as the accelerated failure time (AFT), and the proportional hazard (PH) models, as well as the accelerated hazard (AH) model accounting for crossed survival curves, are sub-classes of this general hazard model. In the proposed class of hazard-based regression models, a covariate's effect is identified as having two distinct components, namely a relative hazard ratio and a time-scale change on hazard progression. The new approach is more adaptive to modelling lifetime data and could give more accurate survival forecasts. The nested structure that includes the AFT, AH, and PH models in the general hazard model may offer a numerical tool for identifying which of them is most appropriate for a certain dataset. In this study, we propose a method for applying these various parametric hazard-based regression models that is based on a tractable parametric distribution for the baseline hazard, known as the generalized log-logistic (GLL) distribution. This distribution is closed under all the PH, AH, and AFT frameworks and can incorporate all of the basic hazard rate shapes of interest in practice, such as decreasing, constant, increasing, V-shaped, unimodal, and J-shaped hazard rates. The Bayesian and frequentist approaches were used to estimate the model parameters. Comprehensive simulation studies were used to evaluate the performance of the proposed model's estimators and its nested structure. A right-censored cancer dataset is used to illustrate the application of the proposed approach. The proposed model performs well on both real and simulation datasets, demonstrating the importance of developing a flexible parametric general class of hazard-based regression models with both time-independent and time-dependent covariates for evaluating the hazard function and hazard ratio over time.

Keywords: survival analysis; proportional hazard model; oncology data; accelerated hazard model; generalized log-logistic distribution; Bayesian approach; accelerated failure time model; general hazard model; maximum likelihood estimation; censored data

MSC: 62N01; 62N02; 62F15; 65C60; 62P10

APPENDIX G
Paper 7

The seventh published manuscript

axioms

MDPI

Article

Amoud Class for Hazard-Based and Odds-Based Regression Models: Application to Oncology Studies

Abdisalam Hassan Muse [1,2,*], Samuel Mwalili [3], Oscar Ngesa [4], Christophe Chesneau [5,*], Huda M. Alshanbari [6] and Abdal-Aziz H. El-Bagoury [7]

1 Institute for Basic Sciences, Pan African University, Technology and Innovation (PAUSTI), Nairobi 62000-00200, Kenya
2 Faculty of Science and Humanities, School of Postgraduate Studies and Research, Amoud University, Borama 25263, Somalia
3 Department of Statistics and Actuarial Science, Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi 62000-00200, Kenya
4 Department of Mathematics and Physical Sciences, Taita Taveta University, Voi 635-80300, Kenya
5 Department of Mathematics, LMNO, CNRS-Université de Caen, Campus II, Science 3, 14032 Caen, France
6 Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
7 Basic Science Department, Higher Institute of Engineering and Technology, El-Mahala El-Kobra 6734723, Egypt
* Correspondence: abdisalam.hassan@amoud.edu.so (A.H.M.); christophe.chesneau@unicaen.fr (C.C.)

Abstract: The purpose of this study is to propose a novel, general, tractable, fully parametric class for hazard-based and odds-based models of survival regression for the analysis of censored lifetime data, named as the "Amoud class (AM)" of models. This generality was attained using a structure resembling the general class of hazard-based regression models, with the addition that the baseline odds function is multiplied by a link function. The class is broad enough to cover a number of widely used models, including the proportional hazard model, the general hazard model, the proportional odds model, the general odds model, the accelerated hazards model, the accelerated odds model, and the accelerated failure time model, as well as combinations of these. The proposed class incorporates the analysis of crossing survival curves. Based on a versatile parametric distribution (generalized log-logistic) for the baseline hazard, we introduced a technique for applying these various hazard-based and odds-based regression models. This distribution allows us to cover the most common hazard rate shapes in practice (decreasing, constant, increasing, unimodal, and reversible unimodal), and various common survival distributions (Weibull, Burr-XII, log-logistic, exponential) are its special cases. The proposed model has good inferential features, and it performs well when different information criteria and likelihood ratio tests are used to select hazard-based and odds-based regression models. The proposed model's utility is demonstrated by an application to a right-censored lifetime dataset with crossing survival curves.

Keywords: Bayesian analysis; survival models; general odds model; general hazard model; accelerated odds model; accelerated failure time model; Amoud class; proportional odds model; accelerated hazard model; survival analysis; generalized log-logistic distribution; proportional hazard model; censored data

MSC: 62N01; 62N02; 62F15; 65C60; 62P10

Citation: Muse, A.H.; Mwalili, S.; Ngesa, O.; Chesneau, C.; Alshanbari, H.M.; El-Bagoury, A.-A.H. Amoud Class for Hazard-Based and Odds-Based Regression Models: Application to Oncology Studies. *Axioms* **2022**, *11*, 606. https://doi.org/10.3390/axioms11110606

Academic Editors: Jiajuan Liang and Kaitai Fang

Received: 5 October 2022
Accepted: 27 October 2022
Published: 1 November 2022

APPENDIX H
R Package 1

The first published R Package

Package 'AmoudSurv'

September 8, 2022

Type Package

Title Tractable Parametric Odds-Based Regression Models

Version 0.1.0

Maintainer Abdisalam Hassan Muse <abdisalam.h.muse@gmail.com>

Description Fits tractable fully parametric odds-based regression models for survival data, including proportional odds (PO), accelerated failure time (AFT), accelerated odds (AO), and General Odds (GO) models in overall survival frameworks. Given at least an R function specifying the survivor, hazard rate and cumulative distribution functions, any user-defined parametric distribution can be fitted. We applied and evaluated a minimum of seventeen (17) various baseline distributions that can handle different failure rate shapes for each of the four different proposed odds-based regression models. For more information see Bennet et al., (1983) <doi:10.1002/sim.4780020223>, and Muse et al., (2022) <doi:10.1016/j.aej.2022.01.033>.

License GPL-3

Encoding UTF-8

LazyData true

Imports AHSurv, flexsurv, pracma, stats, stats4

Depends R (>= 2.10)

RoxygenNote 7.2.1

NeedsCompilation no

Author Abdisalam Hassan Muse [aut, cre]
(<https://orcid.org/0000-0003-4905-0044>),
Samuel Mwalili [aut, ctb],
Oscar Ngesa [aut, ctb],
Christophe Chesneau [aut, ctb]

Repository CRAN

Date/Publication 2022-09-08 09:12:56 UTC

APPENDIX I
R Package 2

The second published R Package

Package 'AHSurv'

June 2, 2022

Type Package

Title Flexible Parametric Accelerated Hazards Models

Version 0.1.0

Author Authors@R. c(person(``Abdisalam", ``Hassan",``Muse"
email=``abdisalam.h.muse@gmail.com",
role=c(``aut", ``ctb",``cre")),
person(``Samuel", ``Mwalili",
role=c(``ctb")),
person(``Oscar", ``Ngesa",
role=c(``ctb")),
person(``Mutua", ``Kilai",
role = c(``ctb"))
)

Maintainer Abdisalam Hassan Muse <abdisalam.h.muse@gmail.com>

Description Flexible parametric Accelerated Hazards (AH) regression models in overall and relative survival frameworks with 13 distinct Baseline Distributions. The AH Model can also be applied to lifetime data with crossed survival curves. Any user-defined parametric distribution can be fitted, given at least an R function defining the cumulative hazard and hazard rate functions. See Chen and Wang (2000) <doi:10.1080/01621459.2000.10474236>, and Lee (2015) <doi:10.1007/s10985-015-9349-5> for more details.

License GPL-3

Encoding UTF-8

LazyData true

Imports flexsurv, rootSolve, stats, stats4

Depends R (>= 2.10)

RoxygenNote 7.1.2

NeedsCompilation no

Repository CRAN

Date/Publication 2022-06-02 11:20:05 UTC