وزارة التعليم العالي والبحث العلمي

Ministry of Higher Education and Scientific Research

BADJI MOKHTAR-ANNABA UNIVERSITY

UNIVERSITE BADJI MOKHTAR ANNABA

جامعة باجي مختار ـ عنابة ـ

Faculty of Sciences
Department of Mathematics

Year : 2024/2025

THESIS

Presented with a view to obtaining the doctorate degree

Estimation in Frailty Models

Stream
Applied Mathematics

Speciality
Probability and Statistics

By
HAMAMI Loubna

SUPERVISOR: GOUAL Hafida **M.C.A** **U.B.M. Annaba**

CO-SUPERVISOR: TALHI Hamida **M.C.A** **U.B.M. Annaba**

In front of the jury

PRESIDENT: ZEGHDOUDI Halim **Prof.** **U.B.M. Annaba**

EXAMINER: LEULMI Sarra **M.C.A** **U. Constantine 1**

EXAMINER: MEZHOUD Kenza Assia **M.C.A** **U. Constantine 1**

وزارة التعليم العالي والبحث العلمي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

BADJI MOKHTAR-ANNABA

UNIVERSITY

UNIVERSITE BADJI MOKHTAR

ANNABA

جامعة باجي مختار

ـعنابة

Faculté des Sciences

Département de Mathématiques

Année : 2024/2025

THÈSE

Présentée en vue de l'obtention du diplôme de Doctorat

Estimation Dans Des Modèles De Fragilité

Filière
Mathématiques Appliquées

Spécialité
Probabilités et Statistique

Par
HAMAMI Loubna

DIRECTEUR DE THÈSE: **GOUAL Hafida** **M.C.A** **U.B.M. Annaba**

CO-DIRECTEUR DE THÈSE: **TALHI Hamida** **M.C.A** **U.B.M. Annaba**

Devant le jury

PRESIDENT:	**ZEGHDOUDI Halim**	**Prof. U.B.M. Annaba**
EXAMINATEUR :	**LEULMI Sarra**	**M.C.A U. Constantine 1**
EXAMINATEUR :	**MEZHOUD Kenza Assia**	**M.C.A U. Constantine 1**

التقدير في نماذج الهشاشة

ملخّص

تُعد نماذج الهشاشة ضرورية في تحليل بيانات البقاء، لأنها تعالج مسألة التباين غير المرصود، والذي قد ينشأ من عوامل متعددة مثل الاستعداد الوراثي، أو التأثيرات البيئية، أو أنماط الحياة. في هذه الدراسة، نقترح نموذجين جديدين للهشاشة: نموذج شبه إكس-جاما (QXg-F) ونموذج جاما-أسي المختلط ،(MGEF) بهدف التعامل مع هذا التباين غير المرصود في بيانات البقاء الأحادية المتغير. يتم إدراج الهشاشة بشكل مضاعف في دالة الخطر الأساسية.

استخدمنا اختبار جودة المطابقة لـ Nikulin-Rao-Robson لتقييم مدى ملاءمة النماذج. ومن خلال دراسات محاكاة، أظهرنا كيف أن نموذجَي QXg-F وMGEF قادران على التقاط التباين وتحسين مطابقة النموذج، كما قمنا بتقييم أدائهما تحت معدلات رقابة مختلفة من النوع الأيمن، واختبرنا قدرة اختبار نسبة الإمكان على اكتشاف التباين غير المرصود حسب حجم العينة. كما قمنا بتطبيق هذه النماذج على بيانات حقيقية، بما في ذلك مجموعة بيانات جديدة تم جمعها من مستشفى طوارئ في الجزائر. تشير نتائجنا إلى أن نموذجي QXg-F وMGEF يشكلان بديلين صالحين لتوزيعات نماذج الهشاشة الحالية، ولديهما القدرة على تحسين دقة تحليلات البقاء في مختلف المجالات، بما في ذلك رعاية الطوارئ. بالإضافة إلى ذلك، قمنا بدراسة فعالية وأهمية نموذج QXg-F في قطاع التأمين من خلال المحاكاة والتطبيقات على بيانات التأمين.

كلمات مفتاحية: تحليل البقاء;البيانات المقطوعة;نموذج الهشاشة;التباين غير المتجانس;الاختبارات الإحصائية

Estimation in fragility models

Abstract

Fragility models are essential for survival data as they address the issue of unobserved heterogeneity, which can arise from various factors such as genetic predisposition, environmental influences, or lifestyle choices.

In this study, we propose two new fragility models: the quasi Xgamma model (QXg-F) and the Mixed Gamma-Exponential model (MGEF) to account for this unobserved heterogeneity in univariate survival data. Fragility is incorporated multiplicatively into the baseline hazard function.

We derive the unconditional survival and hazard functions using the Laplace transform of the fragility distribution, considering Weibull and Gompertz hazard functions as bases, and we estimate the parameters using the maximum likelihood method.

We employ the Nikulin-Rao-Robson goodness-of-fit test to assess model adequacy. Through simulation studies, we demonstrate how the QXg-F and MGEF models capture heterogeneity and enhance model fit, evaluating their performance under various right censoring rates and testing the likelihood ratio's ability to detect unobserved heterogeneity based on sample size. We also apply these models to real data, including a new dataset collected from an emergency hospital in Algeria. Our findings suggest that the QXg-F and MGEF models are viable alternatives to existing fragility modeling distributions and have the potential to improve the accuracy of survival analyses across various fields, including emergency care. Additionally, we examine the effectiveness and relevance of the QXg-F model in the insurance sector through simulations and applications to insurance data.

Keywords: Censored data, Frailty model, Heterogeneity, Laplace transformation, Maximum likelihood, New Mixed Gamma Exponential, Quasi Xgamma, Regression Models, Risk assessment, Statistical Testing, Survival Analysis .

Estimation dans des modèles de fragilité

Résumé

Les modèles de fragilité jouent un rôle crucial dans l'analyse des données de survie en abordant l'hétérogénéité non observée, qui peut résulter de facteurs tels que la génétique, l'environnement ou les choix de vie. Dans cette étude, nous introduisons deux nouveaux modèles de fragilité : le modèle quasi Xgamma (QXg-F) et le modèle Gamma-exponentiel mixte (MGEF) pour traiter cette hétérogénéité dans les données de survie univariées. La fragilité est intégrée de manière multiplicative dans la fonction de risque de base. Nous obtenons les fonctions de survie et de risque inconditionnelles via la transformation de Laplace de la distribution de fragilité, en utilisant des fonctions de risque de base de Weibull et Gompertz, et en estimant les paramètres par la méthode du maximum de vraisemblance. Nous appliquons le test d'ajustement de Nikulin-Rao-Robson pour évaluer la pertinence des modèles. À travers des simulations, nous démontrons comment les modèles QXg-F et MGEF capturent l'hétérogénéité et améliorent l'ajustement, en examinant divers taux de censure à droite et la capacité du rapport de vraisemblance à détecter l'hétérogénéité selon la taille de l'échantillon, ainsi qu'en utilisant des données réelles, incluant un nouvel ensemble de données collectées dans un hôpital d'urgence en Algérie. Nos résultats indiquent que les modèles QXg-F et MGEF constituent des alternatives prometteuses aux distributions de modélisation de fragilité existantes, pouvant améliorer la précision des analyses de survie dans divers domaines, y compris les soins d'urgence. De plus, nous analysons l'efficacité et la pertinence du modèle QXg-F dans le secteur de l'assurance à travers des simulations et des applications sur des données d'assurance.

Mots-clés : Analyse de survie ;Données censurées ; Évaluation des risques ; Hétérogénéité ; Maximum de vraisemblance ; Modèle de fragilité ; Modèles de régression ; Modèle Quasi Xgamma ; Nouveau modèle mixte Gamma-Exponentiel ; Transformation de Laplace ; Tests statistiques.

To Allah.

The Glorified and Exalted,
The only One worshipped,
The greatest Cherisher and Sustainer.

For his satisfaction, I have read and written.

He is the Best Disposer of affairs.
And, Him alone is sufficient for us.
May Him accept us as
his righteous slaves and submitters,
and pious worshippers.

All the praises to Allah.
My Almighty, Allah.
...

Aknowledgements

First and foremost, I would like to express my deepest gratitude to **Allah**, the Most Merciful, for His constant guidance and countless blessings that have illuminated my path throughout this work.

As I conclude this research endeavor, I am profoundly aware that doctoral work is far from a solitary pursuit. The successful completion of this thesis would not have been possible without the invaluable support of numerous individuals whose generosity, encouragement, and genuine interest in my research have guided me through this formative phase of my scholarly journey.

Before all else, I extend my gratitude to my thesis supervisor, Professor **Goual Hafida** of Badji Mokhtar Annaba University. She spared no effort to ensure this work was conducted under optimal conditions. I would like to convey my profound appreciation for guiding my first steps in scientific research. I am equally thankful for the considerable time she devoted to me, her remarkable pedagogical and scientific qualities, her frankness, and her kindness. Working alongside her has been an invaluable learning experience, for which I am truly grateful. Her support proved instrumental in completing this work. Were it not for her supervision and encouragement, I would not have succeeded in finalizing this thesis. I deeply admire her exceptional scientific expertise and personal qualities.

I would like to express my sincere gratitude to my thesis co-supervisor, Dr. **Talhi Hamida**, Associate Professor at Badji Mokhtar Annaba University, for his invaluable guidance and constant encouragement throughout this research.

I wish to express my respect and gratitude to the honorable members of my dissertation committee:

I wish to extend my sincere gratitude to Professor **Zeghdoudi Halim**, Director of the Laboratory, for both the warm welcome he extended to me within his research team since my Master's studies and for graciously accepting to chair my defense committee.

My sincere thanks to Dr. **Leulmi Sarra** , Associate Professor at , and to Dr. **Mezhoud Kenza Assia** , for honoring me by evaluating and examining this work.

I also extend my profound gratitude to Dr. **Meribout Khaoula kaouter**, for her encouragement, invaluable assistance with the statistical analysis, steadfast support, and expert guidance throughout this research.

I am equally grateful to all professors and lecturers in the Mathematics Department for their generous help, unwavering support, and kind encouragement during my research journey.

I extend my deepest thanks to my colleagues and dear friends: **Nadia, Nesrine, Nour**, From the bottom of my heart, I thank you for your unwavering support, your encouragement, and your kindness. Your presence throughout this journey has been invaluable.

I extend my sincere appreciation to all those not mentioned individually here, whose contributions were no less valuable to this work.

I would like to express my gratitude to my cousins **Laghouiter Sabra, Laghouiter Narimen** ,for their unwavering support and encouragement throughout my academic journey.

My final and deepest gratitude is reserved for my family, whose unwavering support has been my foundation throughout this journey. I could never forget to thank: **My Mother, My Father**, my sisters **Kater el Nada, Ghozllan**, and my brother **Abd El Rahman** . Thank you for supporting me when I needed it most, for always believing in me, and for living through every high and low of this thesis journey with me. My gratitude to you extends far beyond what these words can express.

Thank you very much.

Chapter 1

Introduction

Survival analysis is a fundamental branch of statistics concerned with modeling time-to-event data that measure the duration until the occurrence of a particular event of interest. Common examples include the time between the diagnosis of a disease and death, equipment failure in reliability studies, or unemployment duration in labor economics. Owing to its broad applicability, survival analysis is extensively employed across diverse disciplines such as medicine, biology, public health, engineering, economics, and the social sciences.

A foundational assumption in traditional survival models is that the time-to-event observations are independent and identically distributed (i.i.d.). However, this assumption is often violated in real world applications due to the presence of unobserved heterogeneity individual level variability that cannot be explained by the observed covariates. This latent heterogeneity, also referred to as frailty, may significantly influence an individual's risk of experiencing the event of interest, leading to biased estimates if not properly accounted for.

Another key challenge in survival analysis is handling censored data, which occurs when the event of interest is not observed for some individuals within the study period. Together, censoring and unobserved heterogeneity necessitate more sophisticated modeling approaches than those provided by standard parametric distributions such as the exponential, Weibull, or Gompertz models. While these classical distributions are mathematically tractable and widely used, they often possess limited flexibility in representing the complex hazard rate behaviors observed in empirical data. Specifically, their inability to accommodate non-monotonic hazard functions or capture extreme values limits their effectiveness in many applied settings.

To address these limitations, the field has seen the development of more flexible and generalizable modeling models. Among these, frailty models have gained prominence for their ability to incorporate random effects termed frailties that account for individual level variability not captured by observable covariates. The introduction of frailty into survival models allows the hazard function to vary across individuals, thereby improving the model's robustness and interpretability. Seminal works by Vaupel et al. (1979)[92] and Aalen (1988, 1992)[1, 2] laid the groundwork for this approach, highlighting the importance of incorporating stochastic heterogeneity into time-to-event modeling.

Frailty models play a pivotal role in survival analysis by accounting for unobserved heterogeneity arising from latent individual specific factors such as genetics, environmental influences, or lifestyle. These random effects, which cannot be captured by observed covariates alone, significantly impact survival outcomes and risk estimation. In critical domains such as emergency care where patient outcomes are shaped by a complex interplay of measured and unmeasured variables frailty models offer a robust model for understanding survival patterns, quantifying risk, and guiding clinical decision making (Teghri et al., 2024)[91]. Classical frailty models typically rely on the assumption that the frailty term follows a gamma or inverse Gaussian distribution. While these choices offer analytical tractability and, in many cases, closed form expressions for survival and hazard functions, they may inadequately reflect the true nature of latent heterogeneity, particularly in the presence of skewed or heavy-tailed effects.

To address this limitation, researchers have introduced more flexible frailty distributions capable of adapting to a broader range of survival scenarios. Notable examples include the generalized gamma frailty model (Balakrishnan and Peng, 2006)[16] and the weighted Lindley frailty model (Mota et al., 2021)[67], both of which incorporate additional shape parameters to enhance model adaptability. More recently, the quasi-Xgamma frailty model proposed by Loubna et al. (2024)[58] demonstrated improved performance under both censored and uncensored data through rigorous validation using the NIK-RR test. This model effectively captures complex forms of unobserved heterogeneity and offers superior goodness-of-fit compared to traditional approaches. As highlighted by Salem et al. (2023)[82], the development and validation of goodness-of-fit tests are critical for ensuring the reliability of survival models in practice. Furthermore, recent contributions such as the right-skewed actuarial distribution proposed by Hamedani et al. (2023)[40], the reciprocal Weibull extension by Yousof et al. (2023)[103], and the extended Gompertz model by Alizadeh et al. (2024)[8] underscore the continued innovation in modeling survival data across applications in healthcare, reliability, and finance. Both Bayesian and frequentist methodologies have been instrumental in enhancing the interpretability and precision of these models, as demonstrated by Ibrahim et al. (2023)[51], Khedr et al. (2023)[53], and Teghri et al. (2024)[91]. Taken together, these advancements affirm the centrality of frailty models in modern survival analysis and highlight the ongoing need to refine these models for increasingly complex and high-dimensional datasets.

The analysis of survival data is significantly enhanced through the application of frailty models, which serve as powerful tools for capturing unobserved heterogeneity among individuals or groups. These models not only improve the estimation of survival probabilities but also facilitate the identification of covariates associated with increased or decreased risk of failure. Over the years, various statistical distributions have been proposed in the literature to characterize the frailty term. Among the most prominent are the gamma distribution (Clayton, 1978; Vaupel et al., 1979)[19][92], the compound Poisson distribution (Aalen, 1988; 1992)[1][2], the log-normal distribution (McGilchrist and Aisbett, 1991)[65], and more recently, the weighted Lindley distribution (Mota et al., 2021)[67]. These distributions offer analytical tractability and have been widely adopted in reliability and biomedical research. However, their capacity to

fully capture the underlying variability in survival data remains constrained by structural limitations inherent to their forms. In particular, these classical distributions may fail to adequately model the skewness, kurtosis, or tail behavior observed in real-world datasets, thereby motivating the development of more flexible and robust alternatives.

A significant contribution of this thesis is the introduction and comprehensive study of the Quasi-Xgamma Frailty Model, a new parametric frailty structure embedded within the Cox proportional hazards model. This model is motivated by the need to overcome the restrictive assumptions inherent in classical frailty distributions, such as the gamma, inverse Gaussian, and log-normal, which often lack the flexibility to accommodate skewed or heavy-tailed latent effects observed in complex survival data. The Quasi-Xgamma distribution derived as a generalization of the Xgamma distribution introduces an additional shape parameter that enhances its ability to model a wider range of unobserved heterogeneity. This flexibility proves crucial in applications where traditional models underperform due to their inability to adequately capture extreme frailty behaviors or asymmetric variability among individuals.

Chapter 2 provides a detailed exposition of this model. It revisits the standard Cox frailty model and its limitations in capturing diverse latent structures. Then, we formally introduce the Quasi-Xgamma Frailty Model, deriving its density function, Laplace transform, and key properties relevant to survival analysis, such as the marginal survival and hazard functions. The model's identifiability conditions, moments, and asymptotic behavior are also examined to ensure theoretical soundness. Furthermore, the quasi-Xgamma frailty term is integrated into the Cox model through a shared frailty structure, allowing for inter-individual dependence within clustered survival data. Estimation procedures are developed using the maximum likelihood method, and the model's performance is evaluated under both censored and uncensored schemes using extensive simulation studies. These simulations assess bias, mean squared error, and convergence properties under varying levels of censoring and sample size, thereby validating the model's robustness and practical applicability.

By offering improved goodness-of-fit and better adaptability to real-world data compared to existing frailty models, the Quasi-Xgamma Frailty Model represents a novel and practically valuable addition to the frailty modeling literature. Its potential for application in fields such as medical prognosis, reliability engineering, and actuarial risk analysis is underscored by its strong empirical performance and theoretical flexibility.

In the context of this thesis, the application of the Nikulin Rao and Robson (NIK-RR) test and the Bagdonavicius-Nikulin (B-NIK) test plays a crucial role in validating both the Quasi-Xgamma Frailty Model (QXg-F) and the Mixed Gamma-Exponential Frailty Model (MGEF). These models represent key advancements in frailty modeling, and the robust goodness-of-fit tests ensure their theoretical integrity and empirical applicability in survival analysis.

The Quasi-Xgamma Frailty Model (QXg-F) is designed to address the limitations of classical frailty models by offering enhanced flexibility in capturing latent heterogeneity, particularly when frailty distributions are skewed or heavy-tailed. It extends

traditional frailty models by incorporating the quasi-Xgamma distribution, which provides a more accurate fit for complex survival data, making it suitable for real-world applications where individual-level variability is high. This model is pivotal in overcoming the constraints imposed by standard frailty models, such as the gamma or inverse Gaussian distributions, which often fail to represent the true variability in survival data.

Chapter 3 introduces the Mixed Gamma-Exponential Frailty Model (MGEF), a new statistical model designed to capture unobserved heterogeneity in survival data more flexibly than traditional frailty models. Classical choices for the frailty term—such as the gamma, inverse Gaussian, or log-normal distributions—offer analytical simplicity but may inadequately represent the full range of latent variability present in real-world survival data. In response to this limitation, the MGEF model blends the gamma and exponential distributions, creating a more adaptable frailty structure capable of accommodating both light- and heavy-tailed behaviors. This enhanced flexibility allows the model to better represent latent effects in heterogeneous populations, especially in clustered survival settings common in medicine, engineering, and finance.

The model is formulated by constructing the frailty term as a mixture of gamma and exponential components, with derivations provided for its probability density function, Laplace transform, and marginal survival and hazard functions. Identifiability, moment expressions, and asymptotic properties are analytically derived to ensure the model's theoretical rigor. Within the Cox proportional hazards model, the MGEF model is integrated as a shared frailty component, allowing it to account for interdependence within clusters—such as patients within hospitals or policyholders within insurance portfolios.

Estimation of model parameters is conducted through maximum likelihood methods, addressing both censored and uncensored data. A comprehensive simulation study evaluates the model's performance under varying sample sizes and censoring levels. Simulation results demonstrate robust parameter recovery, low bias, and consistent convergence properties, even under moderate to heavy censoring. These results align with the broader statistical literature on distributional extensions and frailty models, where improved flexibility often translates to better empirical performance (Goual et al., 2019; Ibrahim et al., 2019; Yadav et al., 2020)[37] [50] [95].

Similarly, the Mixed Gamma-Exponential Frailty (MGEF) Model represents a significant advancement in the modeling of unobserved heterogeneity within survival analysis. By synergistically integrating the characteristics of the gamma and exponential distributions, the MGEF model offers enhanced flexibility for capturing diverse risk profiles across individuals or clusters. This structure is particularly effective in scenarios where survival outcomes are influenced by both multiplicative and additive latent effects, which are common in medical, insurance, and engineering applications. The capacity of the MGEF model to account for such dual influences enables a more nuanced and accurate representation of frailty, thereby improving model fit and inference in datasets with complex and asymmetric risk patterns.

To ensure rigorous validation of both the Quasi-Xgamma Frailty (QXg-F) and MGEF models, this thesis employs a combination of robust goodness-of-fit testing procedures tailored to different data structures. The Nikulin-Rao-Robson (NIK-RR) test, due to its nonparametric nature, is particularly well-suited for model validation in

complete data settings. It does not rely on strict distributional assumptions, making it a powerful tool for assessing the adequacy of flexible frailty models. By comparing the empirical cumulative distribution function (ECDF) of the observed data to the theoretical distribution under the model, the NIK-RR test is sensitive to deviations such as skewness or heavy tails—features often observed in real-world survival data. Its effectiveness in small samples and robustness to non-normality further reinforce its utility in evaluating the empirical performance of both the QXg-F and MGEF models.

In contrast, the Bagdonavicius-Nikulin (B-NIK) test provides a parametric model specifically designed to accommodate right-censored data, which frequently arises in survival analysis. Assuming normality in the frailty component, the B-NIK test compares sample-based statistics (such as the mean and variance) to their theoretical expectations under the hypothesized model. This test is particularly valuable when the normality assumption is plausible and enables rigorous assessment of model fit in partially observed datasets. It plays a crucial role in confirming the validity of the QXg-F and MGEF models under censored survival schemes.

Taken together, the NIK-RR and B-NIK tests offer a comprehensive and complementary validation strategy. The former provides flexibility and robustness under minimal assumptions, while the latter offers specificity and rigor in the presence of censoring and parametric assumptions. Their combined application in this thesis ensures a thorough evaluation of both proposed frailty models, establishing their statistical soundness and practical relevance for real-world survival data characterized by latent heterogeneity and incomplete observations.

The practical relevance of the Quasi-Xgamma Frailty (QXg-F) and Mixed Gamma-Exponential Frailty (MGEF) models is compellingly demonstrated through their application to real-world survival data. In the field of emergency healthcare, these models were applied to patient survival data obtained from an Algerian hospital, where outcomes are heavily influenced by latent and unmeasured variables such as underlying comorbidities, severity at admission, and treatment variability. Traditional frailty models often fall short in capturing the full extent of such heterogeneity. The QXg-F model, by introducing additional shape flexibility, and the MGEF model, by combining gamma and exponential components to reflect both multiplicative and additive frailty effects, provide more accurate marginal survival and hazard estimates. These improvements translate into enhanced clinical decision-making tools, capable of identifying high-risk subgroups and tailoring medical interventions accordingly.

Beyond healthcare, the models are also evaluated within the insurance sector, where accurate modeling of lifetime risk is essential for actuarial forecasting and solvency assessments. Using real data from life insurance portfolios, the MGEF model in particular exhibited superior goodness-of-fit when compared to classical frailty models, due to its ability to accommodate skewed and heavy-tailed frailty distributions common in actuarial data. Its flexibility proved valuable in better capturing inter-policyholder variability and aligning predicted outcomes with observed claims. These applications underscore the versatility and robustness of the proposed models in handling complex survival data structures across disciplines. Together, the QXg-F and MGEF models not only enrich the theoretical landscape of frailty modeling but also offer significant practical benefits in high-stakes, real-world environments where accurate risk assessment is critical.

In summary, the development of the quasi-Xgamma frailty model and the mixed gamma-exponential frailty model represents a significant advancement in survival analysis. Their practical utility is highlighted through applications to real-world datasets, particularly in emergency care and insurance, where they address the challenges posed by unobserved heterogeneity. The validation of these models against actual data underscores their robustness and relevance. As we progress through this thesis, we will delve into the theoretical foundations of these models, present the methodologies employed, and discuss the implications of our findings for future research and practice in the field of survival analysis.

Chapter 2

Theoretical Foundations of Survival Analysis

Survival analysis is an essential statistical field that studies the time until an event of interest occurs, such as product failure or an individual's death. This chapter aims to lay the theoretical foundations of this discipline by exploring the distributions commonly used in survival analysis. We will begin with an overview of parametric models, which provide structured approaches for modeling survival data.

Next, we will discuss the Laplace transformation, a powerful mathematical tool that facilitates the analysis of survival times. Following this, we will explore regression modeling, focusing on the Cox proportional hazards model, which remains one of the most popular methods due to its flexibility and ability to handle covariates.

We will also examine the accelerated failure time model, which analyzes how covariates influence the time until the event occurs. Finally, we will introduce frailty models, defining their types and discussing their relevance for modeling unobserved heterogeneity in survival data.

Thus, this chapter will establish a solid theoretical framework for subsequent analyses, providing the necessary tools and concepts to understand the complexities of survival data and the models that describe them.

2.1 Background on Distributions in Survival Analysis

Let T be a continuous nonnegative random variable representing the time from a well defined specific starting point until the occurrence of an event. The distribution of the random variable T can be specified through mathematically related functions, where the knowledge of one of them is sufficient to derive the others. These functions are called probability density function, survival and hazard functions; are particularly useful in survival applications.

2.1.1 Probability Density Function

A continuous random variable's distribution can be described using the probability density function (PDF),a fundamental idea in probability theory. This function enables the determination of the probability that a random variable falls inside a specified range. The probability density function (PDF), known by $f(t)$, is the highest chance of something failing in the small range $[t, t+\Delta t]$, scaled by the width Δt. Another way to say this is that the PDF is the chance of failing in an infinitesimal range per unit of time. The PDF is defined as a :

$$f(t) = \lim_{\Delta t \to 0} \frac{P[t < T < t + \Delta t]}{\Delta t} \tag{2.1}$$

where Δt is an infinitely small interval, $f(t) \geq 0$ and $\int_0^\infty f(t)dt = 1$.

The foundational work of 17th-century mathematicians, including Devlin (2010)[26], has significantly shaped the development of probability theory.

The probability density function (PDF) finds utility across a wide array of fields. In economics, for instance, it serves as a crucial tool for assessing the returns on financial assets. In the realm of physics, it helps clarify phenomena such as the dispersal of particles. In biostatistics, comprehending time-to-event data is vital, and the PDF is instrumental for conducting survival analysis and clinical research. Furthermore, in inferential statistics, the PDF facilitates the formulation of conclusions based on sample data. Thus, it is evident that the probability density function holds substantial importance in both theoretical frameworks and practical applications within research.

2.1.2 Cumulative Distribution Function

The cumulative distribution function (CDF) is a crucial concept in probability theory that defines the probability of an event occurring before time t Collett (1994)[20]. A random variable provides a complete characterisation of its distribution, whether discrete or continuous.

the CDF is denoted as $F(t)$ Wienke (2010)[94],defined as :

$$F(t) = P(T \leq t), \quad t > 0 \tag{2.2}$$

The cumulative distribution function is widely utilized across multiple disciplines, including economics, engineering, biology, and social sciences, for various analyses such as reliability, survivability, and risk management. It is pertinent as it can quantify the probability linked to various events, therefore enabling informed decision-making.

2.1.3 Survival Function

In survival analysis, The survival function is a mathematical tool used to measure the probability that an individual would survive beyond the specified time t. This function is the proportion of individuals in a population who, up to that particular point in time, have not encountered the event of interest (e.g., mortality or product malfunction). Denoted as $S(t)$, it is defined as follows:

$$S(t) = P(T > t), \quad t \geq 0 \tag{2.3}$$

The survival function is utilized across multiple disciplines, including medicine, engineering, social sciences, and economics. It is utilized to assess patient survival duration, evaluate product reliability, durability and analyze timing of economic events, such as bankruptcy or employee turnover, to get important insights into the labor market.

2.1.4 Hazard Function

The hazard function gives the instantaneous failure rate at t given that the individual has survived up to time t. The hazard function, denoted by $\lambda(t)$, is mathematically defined as:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}, \quad t > 0 \tag{2.4}$$

where, The numerator is the conditional probability that the event occurs in the small time interval $[t, t + \Delta t]$, given that it has not yet occurred at time t,

The origins of the hazard function are closely associated with the development of survival analysis in the early $20th$ century. The evolved hazard function has become a crucial concept in reliability theory and survival analysis.

Thus, the hazard function has emerged as a crucial component in the fields of reliability theory and survival analysis. The failure rate function, or instantaneous risk function, quantifies the frequency of an event, such as mechanical failure, mortality or breakdown, for a specific topic that has endured until a specified period t.

2.1.5 Cumulative Risk Function

The cumulative risk function is a key concept in statistics, particularly in survival analysis and reliability engineering. It assesses, at a specific time, the total risk of an occurrence such as mortality or failure. This capability elucidates the likelihood of various outcomes and enables scholars and professionals to comprehend the accumulation of risk over time. The cumulative risk function, denoted as $\Lambda(t)$, is formally represented as the integral of the hazard function $\lambda(t)$:

$$\Lambda(t) = \int_0^t \lambda(u)\, du, \quad t > 0 \tag{2.5}$$

The cumulative risk function serves various important functions across numerous disciplines. Medical research widely utilizes it for survival analysis, assessing patients' survival durations to facilitate the evaluation of treatment effectiveness and patient prognosis. In engineering, the function facilitates the assessment of system and component reliability, thereby enhancing failure prediction and maintenance planning. Furthermore, finance and insurance utilize it in risk management to assess and mitigate policy and investment-related risks.

The relationship of $S(t)$ and $\lambda(t)$ is given below

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d \log S(t)}{dt}. \tag{2.6}$$

Likewise,

$$S(t) = \exp\left[-\int_0^t \lambda(u) du\right] = \exp[-\lambda(t)], \qquad t \geq 0 \tag{2.7}$$

and the cumulative hazard function which can be obtained from the survival function since

$$\Lambda(t) = -\log S(t). \tag{2.8}$$

while the probability density function can be given by

$$f(t) = \lambda(t) \exp\left[-\int_0^t \lambda(u)du\right], \qquad t \geq 0 \tag{2.9}$$

All these functions give a mathematical equivalent specification of the distributions of the survival time T. If one of them is known, then the other two can be determined.

2.2 Overview of Some Parametric Models

2.2.1 Exponential Distribution

The exponential distribution is a continuous probability distribution primarily used to model the waiting time before a random event occurs, such as the lifetime of a component, the time between two arrivals in a Poisson process, or the time before a system failure. A random variable t follows an exponential distribution with parameter $\lambda > 0$ if its probability density function (PDF) is given by:

$$f(t) = \lambda e^{-\lambda t}, t \geq 0 \tag{2.10}$$

where, λ is the rate parameter, representing the occurrence rate of the event (e.g., the average number of events per unit time) The cumulative distribution function $F(t)$ gives the probability that the random variable T is less than or equal to a given value x:

$$F(t) = P(T \leq t) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda t}. \quad t \geq 0 \tag{2.11}$$

The survival function $S(t)$ represents the probability that an event has not yet occurred after a given time t. It is defined as:

$$S(t) = P(T > t) = 1 - F(t) = e^{-\lambda t}, \quad t \geq 0 \tag{2.12}$$

This function decreases exponentially, indicating that the probability of survival decreases over time. The instantaneous hazard function $\lambda(t)$, also called the failure rate function, is defined as:

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda, \tag{2.13}$$

The hazard function of the exponential distribution is therefore constant ($\lambda(t) = \lambda$), which means that the instantaneous probability of the event occurring remains the same regardless of the elapsed time.

The cumulative hazard function $\Lambda(t)$, which represents the integral of the hazard function up to x, is given by:

$$\Lambda(t) = \int_0^x \lambda(t)dt = \lambda t, \qquad t \geq 0 \tag{2.14}$$

This function is linear, reflecting the accumulation of risk over time.

A key characteristic of an exponential random variable $T \to \exp(\lambda)$ is its memorylessness.

This attribute corresponds to the following relationship, applicable for all $s, t > 0$:

$$Pr(T > s + t|T > t) = Pr(T > s),$$

The conditional probability of T exceeding $s+t$, given that T exceeds t, is equivalent to the chance of T exceeding s.

The expectation (mean value) of a random variable T following an exponential distribution with parameter λ is given by:

$$\mu = \frac{1}{\lambda}. \tag{2.15}$$

This result implies that the larger the rate parameter λ, the shorter the expected duration before the event occurs. In other words, when events happen more frequently (higher λ), the average waiting time between occurrences is reduced.

The variance of T is given by:

$$\sigma^2 = \frac{1}{\lambda^2}. \tag{2.16}$$

This indicates that the spread of the distribution decreases as λ increases, meaning that for higher event rates, the variability in waiting times is lower.

The distribution's memoryless property makes it essential for modeling electronic component lifetimes, medical survival data, and actuarial risk. However, its limitations in capturing complex failure dynamics have spurred modern enhancements, including machine learning for industrial failure prediction (e.g and Bayesian methods for survival analysis. This evolution underscores its enduring relevance while addressing its constraints in real-world applications.)

2.2.2 Weibull Distribution

The Weibull distribution is a widely used probability distribution in reliability analysis and product lifetime evaluation. It is particularly effective for modeling the time until an event occurs, such as the failure of a product. Its ability to model different shapes of distribution makes it very flexible for various applications.

The probability density function of the Weibull distribution is given by:

$$f(t;\nu,\varphi) = \begin{cases} \frac{\varphi}{\nu}\left(\frac{t}{\nu}\right)^{\varphi-1} e^{-\left(\frac{t}{\nu}\right)^{\varphi}}, & t \geq 0, \\ 0, & t < 0, \end{cases} \tag{2.17}$$

where , t is the random variable,$\nu > 0$ is the scale parameter,$\varphi > 0$ is the shape parameter.

The cumulative distribution function is given by:

$$F(t;\nu,\varphi) = 1 - \exp -(\frac{t}{\nu}), \qquad t \geq 0 \tag{2.18}$$

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

$$S(t;\nu,\varphi) = 1 - F(t;\nu,\varphi) = \exp^{-\left(\frac{t}{\nu}\right)^{\varphi}}. \tag{2.19}$$

The hazard function $\nu(t)$ is given by:

$$\lambda(t;\nu,\varphi) = \frac{f(t;\nu,\varphi)}{S(t;\nu,\varphi)} = \frac{\varphi}{\nu}\left(\frac{t}{\nu}\right)^{\varphi-1}. \tag{2.20}$$

The cumulative hazard function $\Lambda(x)$ is expressed as:

$$\Lambda(x;\lambda,\varphi) = \int_0^x \lambda(t;\lambda,\varphi)\,dt = \left(\frac{x}{\lambda}\right)^{\varphi} \tag{2.21}$$

For the Weibull distribution with scale parameter ν and shape parameter φ:
The mean μ is given by:

$$\mu = \nu\Gamma\left(1+\frac{1}{\varphi}\right). \tag{2.22}$$

where Γ is the gamma function. The variance σ^2 is given by:

$$\sigma^2 = \nu^2\left\{\Gamma\left(1+\frac{2}{\varphi}\right) - \left[\Gamma\left(1+\frac{1}{\varphi}\right)\right]^2\right\}. \tag{2.23}$$

The first person to use the Weibull distribution was the Swedish statistician Wallodi Weibull (1951)[93]. Originally concentrating on material strength assessments, the Weibull distribution soon found applications in many disciplines, including business, engineering, and biology.

2.2.3 Gompertz Distribution

The gompertz distribution is a continuous probability distribution named after Benjamin Gompertz(1825)[34] . It is widely used to describe the distribution of adult lifetimes in demography and actuarial science, and it has applications in various fields such as biology, gerontology, and even computer science for modeling software failure rates. Benjamin Gompertz(1825)[34]introduced his function in a paper titled "On the Nature of the Function Expressive of the Law of Human Mortality." His work was fundamental to the development of statistical science, particularly in the field of life insurance and annuities. The Gompertz function was designed to model human mortality, showing that mortality rates increase exponentially with age. The probability density function of the Gompertz distribution is given by:

$$f(t;\xi,\kappa) = \kappa\exp(\xi t_i)\exp^{-\frac{\kappa}{\xi}[\exp(\xi t_i)-1]}.,\xi>0,\kappa>0,t\geq 0 \tag{2.24}$$

where ξ is the scale parameter and κ is the shape parameter.
The cumulative distribution function is expressed as:

$$F(t;\xi,\kappa) = 1-\exp^{-\frac{\kappa}{\xi}[\exp(\xi t_i)-1]}. \tag{2.25}$$

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

$$S(t;\xi,\kappa) = 1-F(t;\xi,\kappa) = \exp^{-\frac{\kappa}{\xi}[\exp(\xi t_i)-1]}. \tag{2.26}$$

The hazard function $\lambda(t)$ is given by:

$$\lambda(t;\xi,\kappa) = \frac{f(t;\xi,\kappa)}{S(t;\xi,\kappa)} = \kappa e^{\xi t}. \tag{2.27}$$

The cumulative hazard function $\Lambda(t)$ is expressed as:

$$\Lambda(t;\xi,\kappa) = \frac{\kappa}{\xi}\left[\exp(\xi t_i) - 1\right]. \tag{2.28}$$

The mean μ of the Gompertz distribution is given by:

$$\mu = \frac{1}{\kappa}\ln\left(\frac{\xi}{\kappa}\right). \tag{2.29}$$

The variance σ^2 is given by:

$$\sigma^2 = \frac{1}{\kappa^2}\frac{1}{\kappa^2} - \frac{1}{\kappa^2}\ln^2\left(\frac{\xi}{\kappa}\right). \tag{2.30}$$

2.2.4 Gamma Distribution

The gamma distribution is a continuous probability distribution that is widely used in statistics, especially in queuing models, reliability analysis, and Bayesian statistics. It is defined by two parameters: shape η and scale ζ.

The PDF of the gamma distribution is given by:

$$f(t;\eta,\zeta) = \frac{t^{\eta-1}\exp(-\frac{t}{\zeta})}{\zeta^{\eta}\Gamma(\eta)},\ \eta,\zeta > 0,\ t \geq 0 \tag{2.31}$$

where, η is the shape parameter (also known as the alpha parameter), ζ is the scale parameter (also known as the beta parameter).

$\Gamma(\eta)$ is the gamma function, which generalizes the factorial function.

The cumulative distribution function of the gamma distribution is :

$$F(t;\eta,\zeta) = \frac{1}{\Gamma(\eta)}\int_0^x \frac{t^{\eta-1}\exp(-\frac{t}{\zeta})}{\zeta^{\eta}}dt,\ t \geq 0 \tag{2.32}$$

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

$$S(t;\eta,\zeta) = 1 - \frac{\gamma(\eta,\frac{t}{\zeta})}{\Gamma(\eta)},\ t \geq 0 \tag{2.33}$$

The hazard function $\lambda(t)$ is given by:

$$\lambda(t;\eta,\zeta) = \frac{f(t;\eta,\zeta)}{S(t;\eta,\zeta)} = \frac{t^{\eta-1}\exp(-\frac{t}{\zeta})}{\zeta^{\eta}\left[\Gamma(\eta) - \gamma(\eta,\frac{t}{\zeta})\right]},\ t \geq 0 \tag{2.34}$$

The cumulative hazard function $\Lambda(t)$ is expressed as:

$$\Lambda(t;\eta,\zeta) = -\ln[S(t;\eta,\zeta)] = -\ln\left[1 - \frac{\gamma(\eta,\frac{t}{\zeta})}{\Gamma(\eta)}\right]. \tag{2.35}$$

The mean μ of the Gompertz distribution is given by:

$$\mu = \eta\zeta. \tag{2.36}$$

The variance σ^2 is given by:

$$\sigma^2 = \eta\zeta^2. \tag{2.37}$$

Euler first explored the Gamma function in the 18th century, while Laplace and Pearson later formalized the Gamma distribution in the early 20th century (Pearson, 1922)[71]. Pearson introduced the gamma distribution as an extension of the exponential and chi-squared distributions for asymmetric data, and it became essential in Bayesian statistics for modeling exponential and Poisson likelihoods. The distribution has since been applied in reliability engineering (e.g., failure duration modeling), medical statistics (biased survival data analysis), and meteorology (precipitation modeling). Its additive properties make it useful for modeling aggregated events (e.g., total insurance claims), and its exponential family connection ensures optimal maximum likelihood estimation properties.

2.3 Laplace Transformation

The Laplace transform is a mathematical technique that converts time-domain functions into complex frequency-domain functions. This transformation simplifies the solution of linear ordinary differential equations and the analysis of system behavior.

Definition

The Laplace transform of a function $f(t)$, defined for $t \geq 0$, is expressed as:

$$\mathcal{L}\{f(t)\} = F(S) = \int_0^\infty e^{-St} f(t)\, dt, \tag{2.38}$$

where:

- $F(S)$ is the Laplace transform,
- $S = \sigma + j\omega$ is a complex variable,
- e^{-St} is the exponential decay factor.

2.3.1 Sufficient Conditions For Existence Of Laplace Transforms

The Laplace transform exists if $f(t)$ is piecewise continuous on $[0, \infty)$ and does not grow faster than an exponential function. More precisely, there must exist constants M and a such that:

$$|f(t)| \leq Me^{at} \quad \text{for all } t \geq 0 \tag{2.39}$$

2.3.2 Some Important Propreties Of Laplace Transforms

The Laplace transform has several important properties:

Linearity

The Laplace transform is a linear operator:

$$\mathcal{L}\{af(t) + bg(t)\} = aF(s) + bG(s)$$

for all constants a and b.

Time Shifting

If $f(t)$ is delayed by a:

$$\mathcal{L}\{f(t-a)u(t-a)\} = e^{-as}F(s)$$

where $u(t)$ is the unit step function.

Frequency Shifting

Multiplying $f(t)$ by an exponential function results in frequency shifting:

$$\mathcal{L}\{e^{at}f(t)\} = F(s-a)$$

Differentiation

The transform of the derivative of $f(t)$ is given by:

$$\mathcal{L}\{f'(t)\} = sF(s) - f(0)$$

For the n-th derivative:

$$\mathcal{L}\{f^{(n)}(t)\} = s^nF(s) - s^{n-1}f(0) - s^{n-2}f'(0) - \cdots - f^{(n-1)}(0)$$

Integration

The Laplace transform of the integral of $f(t)$ is:

$$\mathcal{L}\left\{\int_0^t f(\tau)\,d\tau\right\} = \frac{F(s)}{s}$$

2.3.3 Inverse Laplace Transform

To return from the frequency domain to the time domain, the inverse Laplace transform is defined as:

$$\mathcal{L}^{-1}\{F(s)\} = f(t) \tag{2.40}$$

This can be computed using contour integration or by referring to transform tables.

2.3.4 Common Laplace Transforms

Here are some common functions and their Laplace transforms:

1. **Unit step function**:
$$\mathcal{L}\{u(t)\} = \frac{1}{s}$$

2. **Exponential function**:
$$\mathcal{L}\{e^{at}\} = \frac{1}{s-a} \quad (\text{Re}(s) > a)$$

3. **Sine function**:
$$\mathcal{L}\{\sin(\omega t)\} = \frac{\omega}{s^2+\omega^2}$$

4. **Cosine function**:
$$\mathcal{L}\{\cos(\omega t)\} = \frac{s}{s^2+\omega^2}$$

5. **Polynomial function**:
$$\mathcal{L}\{t^n\} = \frac{n!}{s^{n+1}} \quad (n \in \mathbb{N})$$

The Laplace transform is a powerful mathematical tool that converts time-domain functions to the complex frequency domain, extending Fourier analysis to non-periodic signals. Its key properties include linearity, time/frequency shifting capabilities, and the ability to transform differential operations into algebraic expressions. Requiring only exponential decay or controlled polynomial growth for convergence, it provides crucial insights into system stability and causality. The transform finds wide application in electrical circuit design, control systems, and communication engineering. It also serves as a fundamental technique for solving differential equations in physics, economics, and signal processing. By converting complex differential equations into simpler algebraic forms, it bridges theoretical mathematics with practical engineering solutions. This versatility makes the Laplace transform indispensable across applied mathematics, engineering, and physical sciences.

2.4 Regression Modeling

2.4.1 Simple and Multiple Linear Regression

Linear regression has its roots in the pioneering work of Legendre and Gauss(2002), who developed the least squares method in the early 19th century, with later contributions from Galton, Pearson (1922) [71]and Fisher (1999)[31] formalizing its theoretical foundations. The method relies on key assumptions including linearity, normality, independence, and homoscedasticity of residuals. Ordinary least squares (OLS) estimation provides optimal parameter estimates under these conditions, with inference conducted via t-tests and F-tests. This continuous evolution, from its classical formulation to contemporary adaptations, ensures linear regression remains a cornerstone of statistical modeling and machine learning.

Simple Linear Regression

Simple linear regression establishes a linear relationship between a quantitative dependent (response) variable Y and an independent (explanatory) variable X.

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \tag{2.41}$$

where:

- Y_i: Observed value of the response variable for individual i
- X_i: Value of the predictor for individual i
- β_0: Intercept(value of Y when $X = 0$)
- β_1: Slope coefficient (effect of a one-unit increase in X on Y)
- ϵ_i:Error term(residual), assumed $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$

The estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are obtained by minimizing the sum of squared residuals(SSR), which measures the discrepancy between observed values Y_i and predicted values $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$. Mathematically, this is expressed as:

$$\text{SSR} = \sum_{i=1}^{n} \left(Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i) \right)^2. \tag{2.42}$$

The analytical solutions to this minimization yield the ordinary least squares (OLS) estimators:

$$\hat{\beta}_1 = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}, \tag{2.43}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}. \tag{2.44}$$

These estimators possess several important properties. First, they are unbiased:

$$E[\hat{\beta}_0] = \beta_0 \quad \text{and} \quad E[\hat{\beta}_1] = \beta_1, \tag{2.45}$$

guaranteeing that on average, they converge to the true parameter values. Their variances, which measure their precision, are given by:

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}, \tag{2.46}$$

$$\text{Var}(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right). \tag{2.47}$$

From a geometric perspective, the regression line $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$ represents the orthogonal projection of the observations Y onto the vector space spanned by the explanatory variable X and a constant vector. This projection minimizes the Euclidean distance between the observed points and the estimated line, illustrating the underlying optimization of the least squares method. This approach provides powerful visual intuition for how the model fits the data.

Multiple Linear Regression

Multiple linear regression extends simple regression by incorporating several explanatory variables to predict a continuous response variable.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2) \tag{2.48}$$

Where

- Y_i: Response variable for observation i (e.g., annual income)
- $X_{1i}, X_{2i}, \ldots, X_{pi}$: Predictor variables (e.g., age, education level, work experience)
- β_0: Intercept (value of Y when all $X = 0$)
- $\beta_1, \beta_2, \ldots, \beta_p$: Regression coefficients (marginal effect of each X_j on Y)
- ϵ_i: Error term (unexplained variability)

Matrix Representation (for n observations and p predictors):

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{2.49}$$

where: $\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$: Response vector , $\mathbf{X} = \begin{bmatrix} 1 & X_{11} & \ldots & X_{p1} \\ 1 & X_{12} & \ldots & X_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & \ldots & X_{pn} \end{bmatrix}$: Design matrix , $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$: Coefficient vector , $\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$: Error vector.

The OLS method estimates the regression coefficients $\boldsymbol{\beta}$ by minimizing the residual sum of squares (RSS), which measures the discrepancy between observed (Y_i) and predicted values. Formally, the estimator $\hat{\boldsymbol{\beta}}$ is derived by solving:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left[Y_i - (\beta_0 + \beta_1 X_{1i} + \cdots + \beta_p X_{pi}) \right]^2, \tag{2.50}$$

yielding the analytical solution $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$, where $\mathbf{X}$ is the design matrix and $\mathbf{Y}$ the response vector.

Linear regression requires four critical assumptions for valid inference :

- Linearity between the dependent variable Y and each X_j must be linear
- Absence of multicollinearity, with Variance Inflation Factors (VIF < 5).
- Homoscedasticity of residuals (constant variance);
- Approximately normal residual distribution, particularly important for small samples (n < 30). Violations of these assumptions may require remedial measures like transformations or robust regression techniques .

Logistic Regression

Logistic regression is a statistical model designed to predict the probability of a binary outcome (e.g., success/failure, presence/absence). Unlike linear regression, it models the log-odds(logit) of the probability rather than the response variable itself. Developed by David Cox (1958)[24], logistic regression evolved from early probability models (19th century) to become the standard for binary outcome modeling, with modern extensions like LASSO enhancing its utility.

Binary Logistic Model The binary logistic regression model is a statistical method used when the dependent variable (Y) is dichotomous(e.g., Yes/No, Success/Failure, 1/0). Unlike linear regression, which predicts continuous outcomes, logistic regression predicts the probability that Y belongs to a particular category. The general form is given by:

$$\log\left[\frac{P(Y=1)}{1-P(Y=1)}\right] = \beta_0 + \beta X, \tag{2.51}$$

where:

- $P(Y=1)$ is the probability of the event occurring
- β_0 is the intercept
- β is the coefficient vector for predictors X

The inverse transformation (sigmoid function) gives the probability:

$$P(Y=1) = \frac{e^{\beta_0+\beta X}}{1+e^{\beta_0+\beta X}} \tag{2.52}$$

This ensures $P(Y=1)$ remains bounded between 0 and 1.

Multinomial Logistic Regression Multinomial logistic regression (MLR) is an extension of binary logistic regression used when the dependent variable Y is nominal categorical with $K > 2$ unordered categories(e.g., disease types, vehicle choices: car/bus/train). Unlike ordinal logistic regression, MLR does not assume any natural ordering among categories.

For a dependent variable Y with K classes, MLR models the probability of each class relative to a reference category (typically K) using the softmax function. For each class k (where $k = 1, 2, \ldots, K-1$):

$$\log\left[\frac{P(Y=k)}{P(Y=K)}\right] = \beta_{0k} + \beta_{1k}X_1 + \beta_{2k}X_2 + \cdots + \beta_{pk}X_p \tag{2.53}$$

- $P(Y=k)$: Probability of observation belonging to class k
- $P(Y=K)$: Probability of the reference class K
- β_{pk}: Coefficients for predictor X_p and class k

The probabilities for all classes are derived using the **softmax function**:

$$P(Y=k) = \frac{e^{\beta_{0k}+\beta_k X}}{\sum_{j=1}^{K} e^{\beta_{0j}+\beta_j X}} \tag{2.54}$$

Multinomial logistic regression is a versatile tool for multiclass classification, though its assumptions (IIA) and computational demands warrant careful validation.

The coefficients β in logistic regression are estimated by maximizing the log-likelihood function:

$$\ell(\beta) = \sum_{i=1}^{n} [Y_i \log(P_i) + (1 - Y_i) \log(1 - P_i)] \tag{2.55}$$

where $P_i = \frac{e^{\beta^T X_i}}{1+e^{\beta^T X_i}}$ represents the predicted probability for observation i. These estimators possess optimal statistical properties (consistency, efficiency) when the model assumptions are satisfied. In practice, implementation typically relies on statistical software that integrates these numerical optimization methods while providing convergence diagnostics.

2.4.2 Cox Proportional Hazards Model

The models discussed thus far are predicated on the assumption of independent and identically distributed (i.i.d.) variables, which inherently presumes population homogeneity. Nevertheless, in empirical research settings, populations are typically heterogeneous, characterized by variability across multiple demographic, clinical, and lifestyle factors. these variables are formally designated as covariates. The Cox proportional hazards model (Cox, 1972)[23] constitutes a regression framework in which event time serves as the response variable. This model facilitates the straightforward incorporation of observed covariates into survival analysis, rendering it the most extensively utilized methodological approach in the field. Let $\lambda(t|X)$ denote the hazard of an individual at time t with covariate vector $X' = (X_1, \ldots, X_k)$. Here X' denotes the transpose of the column vector X. The proportional hazards model specifies that

$$\lambda(t|X) = \lambda_0(t)\lambda(X), \tag{2.56}$$

where: $\lambda_0(t)$ = baseline hazard (unspecified, nonparametric) and $h(\cdot)$ some positive function.

The model assumes a baseline hazard that all individuals in the study population have in common. The parameters of primary interest are contained in $\lambda(X) = \lambda(\beta, X)$, often

$$\lambda(X) = e^{\beta' X}, \tag{2.57}$$

with $\beta = (\beta_1, \beta_2, \ldots, \beta_k)$ denoting the vector of regression parameters.
The Cox proportional hazards model incorporates covariates through a multiplicative effect on the baseline hazard function, thereby modifying an individual's risk profile according to their prognostic characteristics. This formulation yields a mathematically tractable model with clinically interpretable parameters. The model's key theoretical foundation lies in its decomposition of temporal effects (captured by the baseline hazard function) and covariate effects (modeled through an exponential component), providing both analytical flexibility and clinical relevance. The survival function given the covariates $\mathbf{X}$ is

$$S(t|\mathbf{X}) = S_0(t)^{e^{\beta X}}, \tag{2.58}$$

where $S_0(t) = e^{-\int_0^t \lambda_0(s)\,ds}$ denotes the baseline survival function and the components of the vector β are unknown regression parameters. That means the survival function of an individual with covariate vector $\mathbf{X}$ is a power of the baseline survival function. The parametric approach to this modeling framework requires specification of the baseline hazard function from among established parametric survival distributions. This specification enables more precise estimation when sample sizes are limited, though at the cost of distributional assumptions.

2.4.3 Accelerated Failure Time Model

One of the important regression model in survival analysis is the accelerated failure time (AFT) model. The Cox model and its various generalizations are mainly used in the medical and biostatistical field, while the AFT model is primarily applied in reliability theory and industrial experiments and as an alternative if the proportional hazards assumption does not hold. The hazard of the AFT regression model can be written in the form:

$$\lambda(t|X) = \lambda_0(te^{\beta X})e^{\beta X}, \tag{2.59}$$

with β is a vector of coefficients, $\lambda_0(\cdot)$ as baseline hazard,The term $e^{\beta X}$ adjusts the survival time, accelerating or delaying the time until the event occurs based on the covariates. Due to computational difficulties AFT models are mainly used based on parametric approaches . In contrast to the proportional hazards model, the AFT model can best characterized and interpreted in terms of the survival function. Assuming a model with only one single binary covariate X (for example, indicating treatment $(X = 1)$ and control group $(X = 0)$ in a randomized clinical trial) for the survival function, the relation

$$S(t|X = 1) = S(e^{\beta t}|X = 0). \tag{2.60}$$

holds. The hazard function (2.59) of an individual in the control group is then the baseline hazard $\lambda(t|0) = \lambda_0(t)$. The factor e^{β} is called acceleration factor, meaning that the probability to survive time point t in the treatment group is similar to the probability to survive time point te^{β} in the control group. The AFT model will not be treated in detail here in this thesis, but we would like to mention that these models offer a number of advantages. In particular, they provide a wide variety of shapes of baseline hazard functions since the family includes unimodal hazards such as the one in the log-normal distribution.

2.5 Frailty Models: Definition and Types

The frailty model proposed by Vaupel et al. (1979)[92] offers a way to relax the proportional hazards assumption inherent in the Cox model. As an extension of the Cox framework, it accounts for unobserved heterogeneity in survival data. Frailty provides a flexible approach to modeling random effects, latent variability, or potential dependencies within survival models. In its basic form, frailty represents an unobserved random effect that multiplicatively shifts the hazard function for an individual or a cluster of correlated observations. This effect is modeled using a random variable with a specified probability distribution. The parameters of this distribution—alongside the model's other parameters—are estimated during the fitting process.

Many models, each with their own specific frailty structure, have been proposed since Vaupel et al. (1979)[92], and these models allow for different modeling approaches. In this section, we describe some models frequently used in the literature.

2.5.1 Univariate Frailty Models

This section explores univariate survival data, focusing on event times for independent individuals. Traditional survival models assume independence and identical distribution, indicating population homogeneity based on observed covariates. However, significant individual differences, particularly in treatment responses, highlight the importance of unobserved heterogeneity, a critical factor in medical and biological research. Frailty models, which have gained prominence, address this by assigning individual vulnerability levels, where frailer individuals experience earlier events, leading to selection bias favoring more resilient survivors. Mortality rates typically exhibit a unimodal pattern, initially rising, peaking, and then declining or stabilizing. This current section addresses the integration of covariates in survival analysis, primarily via proportional hazards models. However, some risk factors may be unrecognized or their relevance not fully understood, while others may be too costly or time-consuming to measure. Consequently, it is essential to differentiate between two sources of variation in duration data: (1) variability explained by observed covariates, which is theoretically predictable, and (2) heterogeneity from unobserved factors, which remains inherently unpredictable. This section primarily focuses on the latter, noting that recognizing unobserved heterogeneity can elucidate unexpected results, such as non-proportional or declining hazard functions.

In survival analysis, these latent risks are represented by a mixture variable known as frailty, a random variable assumed to follow a specific probability distribution. The relationship between individual and population-level hazard patterns depends on the distribution of frailty among individuals. Different distributions can be selected to model these unobserved covariates (frailty). A critical factor is how the Laplace transform of the frailty distribution influences the survival and hazard functions. The variance of frailty indicates the degree of unobserved heterogeneity and serves as a marker for significant risk factors.

Concept of Univariate Frailty

Before studying specific frailty distributions in detail, we collect a few results that hold for all these distributions in general. Special focus is on the Laplace transform of the frailty, because unconditional survival and hazard functions can be easily expressed using it. Hence, likelihood functions can also be expressed by means of the Laplace transform, making distributions with simple Laplace transforms especially popular.

Frailty modeling began with Beard (1959)[17], Vaupel et al. (1979)[92], and Lancaster (1979)[55]. The Cox proportional hazards (Cox-PH) model (Cox, 1972)[23] accounts for observed covariates, but unobserved heterogeneity may remain. The univariate frailty model addresses this by incorporating latent variables to capture unexplained variation in failure rates across independent subjects, under the assumption that frailties are mutually independent. Let $Z > 0$ denote an unobserved random variable representing the frailty of a subject. Consequently, the hazard rate function for

the ith subject can be expressed as:

$$\lambda\left(t_i|z_i, x_i\right) = z_i\lambda_0(t_i)\exp(x_i^T\beta), \quad i = 1, 2, ..., n, \tag{2.61}$$

Let $\lambda_0(\,\cdot\,)$ denote the baseline hazard function, and let represent the vector of unknown regression coefficients, where $p < n$ (Ibrahim et al., 2001)[45]. Each subject i is associated with a distinct frailty term z_i, an unobserved non-negative random variable. If $z_i > 1$, the frailty amplifies the risk of the event of interest, whereas $z_i < 1$ diminishes it. Notably, the standard Cox proportional hazards (Cox-PH) model emerges as a special case when $z_i = 1$ for all subjects. As in Cox's proportional hazards model (2.56), the baseline hazard function $\lambda_0(\cdot)$ may be specified either nonparametrically or parametrically. In the nonparametric case, common estimators for the cumulative hazard function include the Breslow and Nelson-Aalen (1980)[18]estimators along with their modified variants. This specification results in a semiparametric frailty model, where a parametric frailty distribution is assumed while leaving the baseline hazard unspecified .

For parametric specifications, standard baseline hazard functions from distributions such as exponential, lognormal, Gompertz, and Weibull are frequently employed Wienke (2010)[94]. Each parametric choice offers distinct advantages:

- Exponential: Constant hazard rate
- Weibull: Monotonic (increasing/decreasing) hazards
- Gompertz: Flexible mortality patterns
- Lognormal: Non-monotonic hazard shapes

Given this formulation, the conditional survival function for the i^{th} subject as follows:

$$S(t_i|z_i, x_i) = \exp\left[-z_i\Lambda_0(t_i)\exp(x_i^T\beta)\right] \ , i = 1, ..., n \tag{2.62}$$

where the cumulative baseline hazard-rate function is $\Lambda_0(t_i) = \int_0^{t_i}\lambda_0(s)ds$. The conditional survival function (2.62) thus indicates the likelihood that the ith subject will live until time t_i given $Z = z_i$. To obtain unconditional survival and hazard functions independent of unobserved heterogeneity we integrate the conditional survival function (2.62) over the frailty distribution. This integration is mathematically equivalent to evaluating the Laplace transform of the frailty variable Z.

Following Elbers (1982) [29], identifiability of the resulting model requires the frailty distribution to have a unit mean, i.e., $\mathbb{E}[Z] = 1$. Under this constraint, the unconditional survival function $S(t \mid \mathbf{x})$ is derived by marginalizing $S(t \mid z, \mathbf{x})$ over Z, yielding:

$$S(t \mid \mathbf{x}) = \int_0^\infty S(t \mid z, \mathbf{x})\, f_Z(z)\, dz, \tag{2.63}$$

where $f_Z(z)$ is the probability density function of Z.
given that $Z = z_i$. the marginal Survival function :

$$S(t_i|x_i) = \int_0^\infty \exp\left[-z_i\Lambda_0(t_i)\exp(x_i^T\beta)\right] f(z_i)dz_i = \mathcal{L}_f\left[\Lambda_0(t_i)\exp(x_i^T\beta)\right] \tag{2.64}$$

where $L_f(\cdot)$ denotes the Laplace transform of the frailty distribution, and the corresponding marginal probability density function (MPDF) is given by:

$$f(t_i|x_i) = -\lambda_0(t_i)\exp(x_i^T\beta)L_f'\left[\Lambda_0(t_i)\exp(x_i^T\beta)\right], i = 1, ..., n.$$

It is noteworthy that the Laplace transform admits a closed-form expression for each of the aforementioned distributions. Consequently, the marginal hazard (Mar-H) function can be derived from Equation (2.64) as follows:

$$\lambda(t_i|x_i) = -\frac{\lambda_0(t_i)\exp(x_i^T\beta)\mathcal{L}_f'\left[\Lambda_0(t_i)\exp(x_i^T\beta)\right]}{\mathcal{L}_f\left[\Lambda_0(t_i)\exp(x_i^T\beta)\right]}, i = 1, 2, ..., n. \tag{2.65}$$

where $\mathcal{L}_f'(t) = \frac{\partial}{\partial t}\mathcal{L}_f(t)$. For a randomly selected individual from the study population, the associated risk and survival probabilities can be derived using the hazard and marginal survival (Mar-S) functions presented above (Wienke, 2010)[94]. As previously noted, employing a frailty distribution with a closed-form Laplace transform significantly facilitates parameter estimation and is essential for computing the Mar-S and hazard functions. In cases where the frailty distribution does not possess a closed-form Laplace transform, numerical integration techniques or Markov Chain Monte Carlo (MCMC) methods become necessary (Balakrishnan (2006)[16]; Hougaard (2012)[43]; Robert (2013)[78]). The selection of an appropriate frailty distribution requires careful consideration of computational tractability in both univariate and multivariate survival analysis frameworks (Pickles (1995); Wienke (2010))[74, 94]. This investigation develops novel hazard and Mar-S functions for.

2.5.2 Multivariate Frailty Models

A second key application of frailty models lies in the analysis of multivariate survival data. Alternatively—and from an entirely distinct perspective—the frailty term can be employed to model dependencies between event times, an approach originally introduced by Clayton (1978)[19]. Such kind of data occurs,for example, if lifetimes (or times of onset of a disease) of relatives (twins, parent-child) or recurrent events like infections in the same individual are considered. In such cases, independence between the clustered survival times cannot be assumed. Multivariate models are able to account for the presence of dependence between these event times. A commonly used and very general approach is to specify independence among observed data items conditional on a set of unobserved or latent variables. The dependence structure in the multivariate model develops from a latent variable in the conditional models for multiple observed survival times. Frailty models for multivariate survival data are derived under conditional independence assumption by specifying latent variables that actmultiplicatively on the baseline hazard funtion. We begin with an example of a shared frailty model, then develop a correlated multivariate frailty model.

Concept of Shared Frailty

We explore here multivariate event-time models with dependent hazards, extending the classical univariate frailty model to account for the interdependence of clustered event times. Such survival models are particularly useful for analyzing complex processes related to aging, disease, disability, and mortality, where dependencies may arise among related individuals or recurrent events.

A prominent method in this field is the shared frailty concept, where the hazard model for each individual reflects the conventional univariate frailty model, but with frailty being a measure of relative risk shared among group members. This approach assumes that all failure times within a cluster are conditionally independent given the frailty, which is constant over time and shared among individuals, leading to positively dependent event times.

Originally introduced by David Clayton (1978)[19], who studied the bivariate case, the shared frailty model has become a cornerstone in the literature on multivariate frailty models. We will provide a concise overview of this model to clarify the foundational concepts of multivariate frailty models.

The fundamental definition of a shared frailty model in survival analysis is as follows. Suppose there are n clusters, and cluster i has n_i observations associated with the unobserved frailty Z_i (where $1 \leq i \leq n$). The vector X_{ij} (where $1 \leq i \leq n, 1 \leq j \leq n_i$) contains the covariate information for the event time T_{ij} of the j-th observation in the i-th cluster. Given the frailty term Z_i, the survival times in cluster i are assumed to be independent, with hazard functions defined as:

$$\lambda(t|X_{ij}, Z_i) = Z_i \lambda_0(t) e^{\beta X_{ij}},$$

where $\lambda_0(t)$ denotes the baseline hazard function, and β is a vector of fixed-effect parameters to be estimated. The frailties Z_i (for $i = 1, \ldots, n$) are considered to be independent and identically distributed random variables with a specified density function $f(z)$. The frailty density relies on unknown parameters that must be estimated. Like the univariate case, a semiparametric shared frailty model includes a nonparametric baseline hazard $\lambda_0(t)$.

The conditional multivariate survival function for individuals in the i-th cluster. Conditional on the shared frailty Z_i, it holds that:

$$\begin{aligned} S(t_{i1}, \ldots, t_{in_i}|X_i, Z_i) &= S(t_{i1}|X_{i1}, Z_i) \cdots S(t_{in_i}|X_{in_i}, Z_i) \\ &= \exp\left[-Z_i \sum_{j=1}^{n_i} \Lambda_0(t_{ij}) e^{\beta X_{ij}}\right], \end{aligned} \tag{2.66}$$

where $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$ denotes the cumulative baseline hazard function, and $X_i = (X_{i1}, \ldots, X_{in_i})$ is the covariate matrix of individuals in cluster i. This expression serves as the starting point for deriving the unconditional joint survival function. Averaging expression with respect to the frailty Z_i yields the marginal survival function:

$$\begin{aligned} S(t_{i1}, \ldots, t_{in_i}|X_i) &= E[S(t_{i1}, \ldots, t_{in_i}|X_i, Z_i)] \\ &= E\left\{\exp\left[-Z_i \sum_{j=1}^{n_i} \Lambda_0(t_{ij}) e^{\beta X_{ij}}\right]\right\} \\ &= \mathcal{L}\left[\sum_{j=1}^{n_i} \Lambda_0(t_{ij}) e^{\beta X_{ij}}\right], \end{aligned} \tag{2.67}$$

where $\mathcal{L}$ denotes the Laplace transform of the frailty variable. Thus, the multivariate survival function is expressed as the Laplace transform of the frailty distribution, evaluated at the cumulative baseline hazard. The joint survival function for all

event-time data is now the product of the survival functions of all clusters, due to the assumption of independence between clusters:

$$S(t_{11},\ldots,t_{nnn}|X_1,\ldots,X_n)=\prod_{i=1}^{n}\mathcal{L}\left[\sum_{j=1}^{n_i}\Lambda_0(t_{ij})e^{\beta X_{ij}}\right]. \tag{2.68}$$

The univariate unconditional survival functions can also be expressed using the Laplace transform:

$$\begin{aligned} S(t_{ij}|X_{ij}) &= E[S(t_{ij}|X_{ij},Z_i)] \\ &= E\left\{\exp\left[-Z_i\Lambda_0(t_{ij})e^{\beta X_{ij}}\right]\right\} \\ &= \mathcal{L}\left[\Lambda_0(t_{ij})e^{\beta X_{ij}}\right]. \end{aligned} \tag{2.69}$$

A more detailed presentation of shared frailty models can be found in the excellent book by Duchateau and Janssen (2008)[27].

Concept of Correlated Frailty

The correlated frailty model represents a key concept in multivariate frailty models, bridging the shared frailty approach and the univariate frailty model. In this model, the frailties of individuals within a cluster exhibit correlations without being uniformly shared. This flexibility allows for the incorporation of additional correlation parameters, facilitating the analysis of relationships between event times. In the bivariate context (absent observed covariates), the conditional survival function is defined as follows:

$$S(t_1,t_2|Z_1,Z_2)=S_1(t_1|Z_1)S_2(t_2|Z_2)=e^{-Z_1\Lambda_{01}(t_1)}e^{-Z_2\Lambda_{02}(t_2)}, \tag{2.70}$$

where Z_1 and Z_2 are two correlated frailties. The distribution of the random vector (Z_1,Z_2) must be specified, as it determines the association structure of the event times in the model.

Consider various bivariate event times, such as the lifetimes of twins, or the times to blindness in each eye, as well as the failure times of patients' left and right kidneys. In the bivariate correlated frailty model, each individual's frailty within a pair serves as a measure of relative risk, akin to the definition used in univariate models. However, unlike the shared frailty model, the frailties of the two individuals in a pair do not have to be identical. Thus, the hazard function for the individual j (where $j=1,2$) in pair i (where $i=1,\ldots,n$) is expressed as:

$$\lambda(t|X_{ij},Z_{ij})=Z_{ij}\lambda_{0j}(t)e^{\beta X_{ij}}, \tag{2.71}$$

where t denotes age or time, X_{ij} is a vector of observed covariates, β is a vector of regression parameters describing the effect of the covariates X_{ij}, $\lambda_{0j}(\cdot)$ are the baseline hazard functions, and Z_{ij} are the frailties. Bivariate correlated frailty models are characterized by the joint distribution of a two-dimensional vector of frailties (Z_{i1},Z_{i2}). If the two frailties are independent, the resulting lifetimes will also be independent, indicating the absence of clustering within the model. Conversely, if the two frailties are identical, the shared frailty model emerges as a specific instance of the correlated frailty model, characterized by a correlation of one between the frailties.

The conditional survival function of the j-th individual in the i-th pair is:

$$S(t|X_{ij}, Z_{ij}) = e^{-Z_{ij}\Lambda_{0j}(t)e^{\beta X_{ij}}}, \tag{2.72}$$

where $\Lambda_{0j}(t)$ denotes the cumulative baseline hazard function.

All these formulas can be readily adapted to the multivariate scenario; however, they necessitate a specification of the correlation structure among individuals within a cluster via a multivariate density function, which complicates the analysis.

Chapter 3

The Quasi-Xgamma Frailty Model

The main motivation of this chapter is to:

- Present a new flexible frailty model called the QXg-F model for the survival analysis.
- Employ the QXg-F model in the survival analysis under a newly collected data called emergency care data.
- Using the MaxLE method is used for estimating the QXg-F model's parameters of Weibul Baseline Hazard Function and in case of Gompertz Baseline Hazard Function.
- Propose an alternative frailty model which overcomes the weak point of the gamma frailty model, compound Poisson frailty model, log-normal frailty model and weighted Lindley frailty model.
- Test the goodness-of-fit validation using the NIK-RR test statistic in case of complete data.
- Test the goodness-of-fit validation using the B-NIK test statistic in case of censored data.
- Test the ability of the new QXg-F model in risk analysis by studying a set of commonly used financial indicators such as the V-R, TLV-R, T-VC, TM-V, M-EXL function under different estimation methods like the MaxLE method, OrLSE method, the WLSE method, and the AnDE method.

3.1 Introduction to the Quasi-Xgamma Model

The analysis of time-to-event data presents fundamental challenges across multiple applied disciplines, including medical research, reliability engineering, and actuarial science. While numerous statistical models exist for such data, the efficacy of parametric methods critically depends on selecting an appropriate underlying distribution. This reliance becomes particularly problematic when real-world data exhibit patterns that deviate substantially from conventional probability models, potentially compromising the validity of statistical inferences and predictions.This chapter presents the quasi-Xgamma distribution, a two-parameter extension of the Xgamma distribution

originally developed by Sen et al. (2017)[84]. The Xgamma distribution, defined for a continuous random variable x through its probability density function (PDF) of the form :

$$f(x) = \frac{\vartheta^2}{1+\vartheta}\left(1+\frac{\vartheta}{2}x^2\right)\exp\left(-\vartheta x\right) \ , x>0,\, \vartheta>0 \tag{3.1}$$

and is denoted by $x \to Xgamma\,(\vartheta)$.
The cumulative density function(CDF) of x is given by

$$F(x) = 1 - \frac{\left(1+\vartheta+\vartheta X+\frac{\vartheta^2x^2}{2}\right)}{1+\vartheta}\exp\left(-\vartheta x\right) \ . \ x>0,\, \vartheta>0 \tag{3.2}$$

, has proven valuable for lifetime data modeling. This advancement addresses critical limitations in current lifetime data modeling approaches while opening new research directions. The quasi-Xgamma distribution's implementation in statistical software packages could significantly enhance its accessibility to both researchers and practitioners. As demonstrated through our analysis, this model represents a versatile analytical tool that successfully bridges theoretical sophistication with practical applicability across multiple domains. Following Sen and Chandra (2017)[84], A continuous

random variable x, with support $[0,\infty[$, is said to follow a quasi Xgamma distribution with shape parameters α and scale parameter ϑ

The cumulative distribution function (CDF) of QXg is given by

$$F_{\alpha,\vartheta}(x) = 1 - \frac{1}{1+\alpha}\exp\left(-\vartheta x\right)\left(1+\alpha+\vartheta x+\frac{1}{2}\vartheta^2x^2\right)|\alpha,\vartheta>0, \tag{3.3}$$

where $x>0$ and $\alpha,\vartheta>0$.

The probability density function (PDF) is obtained as

$$f_\phi(x) = \frac{\vartheta}{1+\alpha}\exp\left(-\vartheta x\right)\left(\alpha+\frac{1}{2}\vartheta^2x^2\right)|\alpha,\vartheta>0, \tag{3.4}$$

where $\phi=(\alpha,\vartheta)$. and is denoted by $x \to QXg(\alpha,\vartheta)$.

The Momoents of The probability density function (PDF) corresponding to (3.4) :

$$E\left(x\right) = \frac{3+\alpha}{\vartheta\left(1+\alpha\right)}.$$

$$E\left(x^2\right) = \frac{2\left(6+\alpha\right)}{\vartheta^2\left(1+\alpha\right)}.$$

The variance of x can be deduced as

$$\mathrm{Var}\left(x\right) = \frac{\alpha^2+8\alpha+3}{\varphi^2(1+\alpha)^2}.$$

It is to be noted that the quasi Xgamma distribution such obtained includes Xgamma distribution as a special case. For particular values of α, from (3.4) we obtain the following special cases:

(1) When α= 0, we have gamma distribution with shape parameter 3 and scale parameter ϑ, i.e., $x \to G(\vartheta,3)$

(2) When $\alpha = 1$, we obtain a new class of distributions with PDF

$$f(x) = \frac{\vartheta}{2}\left(1 + \frac{\vartheta^2}{2}x^2\right)\exp(-\vartheta x)\ ;\ x > 0 \text{and } \vartheta > 0 \tag{3.5}$$

(3) When $\alpha = \vartheta$, we have the Xgamma distribution with PDF given in (3.1).

Note: The quasi Xgamma distribution characterized by the PDF given in (3.4) is a special mixture of exponential distribution with mean $\frac{1}{\vartheta}$and gamma distribution with shape parameter 3 and scale parameter ϑ, with mixing proportions $\frac{\alpha}{1+\alpha}$ and $\frac{1}{1+\alpha}$ respectively. The Xgamma distribution is also synthesized mixing exponential distribution with mean $\frac{1}{\vartheta}$ and gamma distribution with shape parameter 3 and scale parameter ϑ, with mixing proportions $\frac{\vartheta}{1+\vartheta}$ and $\frac{1}{1+\vartheta}$ respectively. The parameter α regulates the shape of the distribution for a fixed value of ϑ.

3.2 Quasi Xgamma Frailty Model

Consider the frailty model presented in (2.61), where Z represents our frailty variable, which follows a QXg distribution with a mean of one. This assumption is essential for identifying the parameters of the derived model (refer to Elbers and Ridder, 1982)[29]. Therefore, utilizing the alternative parameterization of the QXg model proposed by sen et al. (2017)[84] in terms of the mean, the probability density function (PDF) of the QXg-F model can be expressed as follows:

$$f(Z) = \frac{3+\alpha}{(1+\alpha)^2} \exp\left(-\frac{3+\alpha}{1+\alpha} Z\right) \left[\alpha + \frac{1}{2}\left(\frac{3+\alpha}{1+\alpha}\right)^2 Z^2\right] |\alpha, Z > 0, \tag{3.6}$$

where $\alpha > 0$ represents the shape parameter. The variance of the frailty distribution serves as a key measure for quantifying the degree of unobserved heterogeneity within a study sample. Assuming that the probability density function (PDF) given in equation (3.6) represents our frailty model, the variance can be mathematically expressed as follows:

$$\sigma^2 = \frac{1}{(3+\alpha)^2}\left(\alpha^2 + 8\alpha + 3\right).$$

Let $\mathcal{L}_f(S)$ denote the Laplace transform of the frailty probability density function $f(Z)$ specified in (3.6), parameterized by its variance σ^2 and a scaling factor $S \in \mathbb{R}$, has the following form:

$$\begin{aligned} \mathcal{L}_f(S) &= \left(1+\sqrt{13-12\sigma^2}\right) \\ &\times \left\{ \begin{array}{c} \left(4-3\sigma^2+\sqrt{13-12\sigma^2}\right) \\ \left[S\left(3-2\sigma^2+\sqrt{13-12\sigma^2}\right)+1+\sqrt{13-12\sigma^2}\right]^2 + \left(1+\sqrt{13-12\sigma^2}\right)^2(\sigma^2-1) \end{array} \right\} \\ &\times \left\{\left(3-2\sigma^2+\sqrt{13-12\sigma^2}\right)\left[S\left(3-2\sigma^2+\sqrt{13-12\sigma^2}\right)+1+\sqrt{13-12\sigma^2}\right]^3\right\}^{-1}, \sigma^2 \leq \frac{13}{12}. \end{aligned} \tag{3.7}$$

To simplify the analysis, we evaluate (3.8) at $S = \Lambda_0(t_i)\exp(\mathbf{x}_i^\top \boldsymbol{\beta})$. Under the QXg-F model, the marginal survival function given in (2.64) takes the following form:

$$\begin{aligned} S(t_i|x_i) &= \left(1+\sqrt{13-12\sigma^2}\right) \\ &\times \left\{ \begin{array}{c} \left(4-3\sigma^2+\sqrt{13-12\sigma^2}\right) \\ \left[\begin{array}{c} \Lambda_0(t_i)\exp(x_i^T\boldsymbol{\beta})\left(3-2\sigma^2+\sqrt{13-12\sigma^2}\right) \\ +1+\sqrt{13-12\sigma^2} \end{array} \right]^2 \\ +\left(1+\sqrt{13-12\sigma^2}\right)^2(\sigma^2-1) \end{array} \right\} \\ &\times \left\{ \begin{array}{c} \left(3-2\sigma^2+\sqrt{13-12\sigma^2}\right) \\ \left[\begin{array}{c} \Lambda_0(t_i)\exp(x_i^T\beta) \\ \times\left(3-2\sigma^2+\sqrt{13-12\sigma^2}\right)+1+\sqrt{\mathbb{C}\left(\sigma^2\right)} \end{array} \right]^3 \end{array} \right\}^{-1}. \end{aligned} \tag{3.8}$$

The resulting Marginal Hazard function (2.65) is as follows:

$$\begin{aligned}\lambda(t_i|x_i) &= \left[\lambda_0(t_i)\exp(x_i^T\boldsymbol{\beta})\right] \\ &\times \begin{pmatrix} \left[3-2\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right] \\ \{\Lambda_0(t_i) \\ exp(x_i^T\beta)\left[3-2\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right]+1+\sqrt{\mathbb{C}\left(\sigma^2\right)}^{-1} \end{pmatrix} \\ \times &\left[\begin{pmatrix} \left[4-3\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right]\left\{ \begin{matrix} \Lambda_0(t_i)\exp(x_i^T\beta)\left[3-2\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right] \\ +1+\sqrt{\mathbb{C}\left(\sigma^2\right)} \end{matrix} \right\}^2 \\ +3(\sigma^2-1)\left[1+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right]^2 \end{pmatrix} \right. \\ &\left. \begin{pmatrix} \left[4-3\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right]\left\{ \begin{matrix} \Lambda_0(t_i)\exp(x_i^T\beta)\left[3-2\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right] \\ +1+\sqrt{\mathbb{C}\left(\sigma^2\right)} \end{matrix} \right\}^2 \\ +(\sigma^2-1)\left[1+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right]^2 \end{pmatrix}^{-1} \right],\end{aligned} \tag{3.9}$$

where

$$\mathbb{C}\left(\sigma^2\right) = 13-12\sigma^2, \sigma^2 \geq \frac{13}{12}.$$

In the following, we refer to the model with unconditional survival and hazard functions as the QXg frailty model. As discussed in Chapter1, the baseline hazard function can be specified either parametrically or non-parametrically. In this study, we adopt a parametric approach by considering Weibull and Gompertz distributions as baseline hazard functions. These two-parameter models are particularly flexible, capable of representing constant, increasing, and decreasing hazard rates.

The Weibull distribution is widely employed in survival analysis and reliability engineering, while the Gompertz distribution finds predominant application in actuarial science and demographic studies [94]. Notably, the Gompertz distribution as a baseline hazard function in survival analysis presents an important characteristic: it may exhibit a defective property. This property implies the existence of a subpopulation that is not susceptible to the event of interest, effectively representing cured individuals or long-term survivors .

3.2.1 QXg-F Model Under the Weibull Baseline Hazard Function

The Weibull distribution provides parametric baseline hazard and cumulative hazard functions defined as follows:

$$\lambda_0(t_i) = \frac{\nu}{\varphi}\left(\frac{t_i}{\varphi}\right)^{\nu-1} \quad \text{and} \quad \Lambda_0(t_i) = \left(\frac{t_i}{\varphi}\right)^{\nu} \qquad t_i > 0, \tag{3.10}$$

where $\nu > 0$ represents the shape parameter and $\varphi > 0$ represents the scale parameter.
The hazard-rate function of the Weibull distribution drops monotonously for $\nu < 1$; it is constant over time for $\nu = 1$ (exponential distribution); and it grows monotony when $\nu > 1$ (Wienke (2010)[94]). As a result of substituting (3.10)into (3.9) and (3.10), the Quasi Xgamma Frailty model's Marginal Survival and hazard functions with Weibull Baseline Hazard Function are, respectively,

$$\begin{aligned} S(t_i|x_i) &= \left[1+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right] \\ &\times \left(\begin{array}{c} \left[4-3\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right] \\ \left\{ \begin{array}{c} \left(\frac{t_i}{\varphi}\right)^{\nu} \exp(x_i^T\boldsymbol{\beta}) \\ \left[3-2\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right]+1+\sqrt{\mathbb{C}\left(\sigma^2\right)} \end{array} \right\}^2 \\ +\left[1+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right]^2\left(\sigma^2-1\right) \end{array} \right) \\ &\times \left(\begin{array}{c} \left[3-2\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right] \\ \left\{ \begin{array}{c} \left(\frac{t_i}{\varphi}\right)^{\nu} \exp(x_i^T\beta) \\ \left[3-2\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right]+1+\sqrt{\mathbb{C}\left(\sigma^2\right)} \end{array} \right\}^3 \end{array} \right)^{-1}, \end{aligned} \tag{3.11}$$

The marginal hazard functions with Weibull Baseline Hazard Function is :

$$\begin{aligned}\lambda(t_i|x_i) &= \left[\frac{\nu}{\varphi}\left(\frac{t_i}{\varphi}\right)^{\nu-1}\exp(x_i^T\beta)\right] \\ &\times\left(\begin{array}{c}\left[3-2\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right] \\ \left\{\left(\frac{t_i}{\varphi}\right)^{\nu}\exp(x_i^T\beta)\left[3-2\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right]+1+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right\}^{-1}\end{array}\right) \\ &\times\left[\begin{array}{c}\left(\begin{array}{c}\left[4-3\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right]\left\{\begin{array}{c}\left(\frac{t_i}{\varphi}\right)^{\nu}\exp(x_i^T\beta)\left[3-2\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right] \\ +1+\sqrt{\mathbb{C}\left(\sigma^2\right)}\end{array}\right\}^2 \\ +3(\sigma^2-1)\left[1+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right]^2\end{array}\right) \\ \left(\begin{array}{c}\left[4-3\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right]\left\{\begin{array}{c}\left(\frac{t_i}{\varphi}\right)^{\nu}\exp(x_i^T\beta)\left[3-2\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right] \\ +1+\sqrt{\mathbb{C}\left(\sigma^2\right)}\end{array}\right\}^2 \\ +(\sigma^2-1)\left[1+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right]^2\end{array}\right)^{-1}\end{array}\right],\end{aligned} \tag{3.12}$$

where

$$\mathbb{C}\left(\sigma^2\right)=13-12\sigma^2.$$

Figure 3.1,3.2 illustrates various configurations of the survival and hazard functions for the QXg frailty model with Weibull baseline hazard, using selected parameter values. The graphical analysis reveals two key properties:

1. The survival function maintains proper behavior, satisfying:

$$\begin{aligned}S(0)&=1\\ \lim_{t\to\infty}S(t)&=0\end{aligned}$$

2. The hazard function exhibits both monotonically decreasing and unimodal shapes

Notably, the unimodal hazard pattern represents a distinctive feature not achievable in the classical Weibull model. This expanded flexibility in hazard shape specification enhances the model's capability to capture complex survival patterns observed in empirical data.

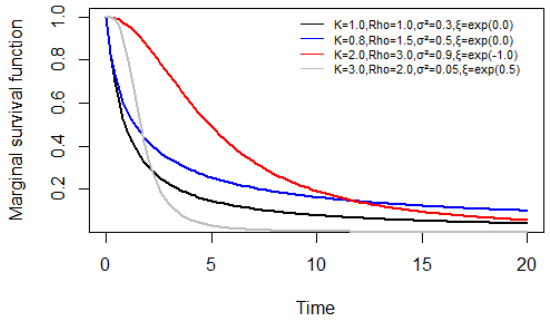

Figure 3.1: Plots of the Marginal Survival Function of the QXg-F Distribution whit Weibull Baseline Hazard Functions.

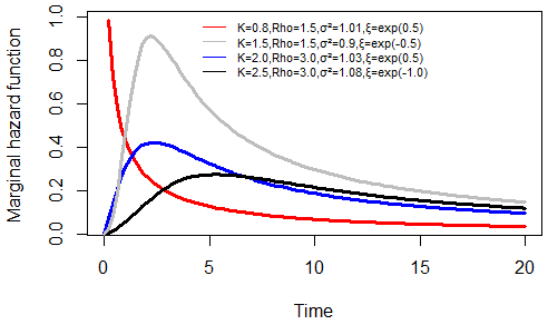

Figure 3.2: Plots of the Marginal Hazard Function of the QXg-F Distribution whit Weibull Baseline Hazard Functions.

3.2.2 QXg-F Model Under the Gompertz Baseline Hazard Function

The Gompertz distribution provides parametric baseline hazard and cumulative hazard functions given by:

$$\lambda_0(t_i) = \kappa \exp(\xi t_i) \quad \text{and} \quad \Lambda_0(t_i) = \frac{\kappa}{\xi}\left[\exp(\xi t_i) - 1\right] \quad , t_i > 0 \tag{3.13}$$

Let $\xi > 0$ and $\kappa > 0$ denote the shape and scale parameters, respectively, of the Gompertz distribution. When $\xi < 0$, the distribution exhibits a defective property characterized by:

$$\lim_{t\to\infty} H_0(t) = -\frac{\kappa}{\xi}$$

$$p_0 = \exp\left(\frac{\kappa}{\xi}\right)$$

where p_0 represents the proportion of cured individuals or long-term survivors in the population. The Gompertz distribution reduces to the exponential distribution when $\xi = 0$.

The hazard rate function demonstrates three distinct regimes:

- $\xi < 0$: Decreasing hazard (cure fraction present)
- $\xi = 0$: Constant hazard (exponential case)
- $\xi > 0$: Increasing hazard (standard Gompertz)

Substituting the Gompertz baseline hazard (3.13)into equations(3.9) and (3.10) yields the marginal survival and hazard functions for the QXg-F model with Gompertz baseline hazard frailty :

$$
\begin{aligned}
S(t_i|x_i) &= \left[1+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right] \qquad (3.14)\\
&\times \left[\begin{array}{c}
\left[4-3\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right] \\
\left(\begin{array}{c}
\left\{\frac{\kappa}{\xi}\left[\exp(\xi t_i)-1\right]\exp(x_i^T\beta)\right\} \\
\left[3-2\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right]+1+\sqrt{\mathbb{C}\left(\sigma^2\right)}
\end{array}\right)^2 \\
+\left[1+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right]^2\left(\sigma^2-1\right)
\end{array}\right] \\
&\times \left(\begin{array}{c}
\left[3-2\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right] \\
\left\{\begin{array}{c}
\frac{\kappa}{\xi}\left[\exp(\xi t_i)-1\right]\exp(x_i^T\beta)\left[3-2\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right] \\
+1+\sqrt{\mathbb{C}\left(\sigma^2\right)}
\end{array}\right\}^3
\end{array}\right)^{-1}.
\end{aligned}
$$

The hazard functions for the QXg-F model with Gompertz baseline hazard :

$$\begin{aligned}\lambda(t_i|x_i) &= \left[\kappa \exp(\xi t_i)\exp(x_i^T\beta)\right] \\ &\times \left(\begin{array}{c} \left[3-2\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right] \\ \left\{\frac{\kappa}{\xi}\left[\exp(\xi t_i)-1\right]\exp(x_i^T\beta)\left[3-2\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right]+1+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right\}^{-1} \end{array} \right) \\ &\times \left[\begin{array}{c} \left(\begin{array}{c} \left[4-3\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right]\left\{ \begin{array}{c} \frac{\kappa}{\xi}\left[\exp(\xi t_i)-1\right]\exp(x_i^T\beta) \\ \left[3-2\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right]+1+\sqrt{\mathbb{C}\left(\sigma^2\right)} \end{array} \right\}^2 \\ +3(\sigma^2-1)\left[1+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right]^2 \end{array} \right) \\ \left(\begin{array}{c} \left[4-3\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right]\left\{ \begin{array}{c} \frac{\kappa}{\xi}\left[\exp(\xi t_i)-1\right] \\ \exp(x_i^T\beta)\left[3-2\sigma^2+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right] \\ +1+\sqrt{\mathbb{C}\left(\sigma^2\right)} \end{array} \right\}^2 \\ +(\sigma^2-1)\left[1+\sqrt{\mathbb{C}\left(\sigma^2\right)}\right]^2 \end{array} \right)^{-1} \end{array} \right], \end{aligned} \tag{3.15}$$

where

$$\mathbb{C}\left(\sigma^2\right) = 13 - 12\sigma^2.$$

The graphical representations in Figures 3.3 3.4 demonstrate the behavior of marginal survival and hazard functions for the QXg frailty model employing a Gompertz baseline hazard, evaluated at specific parameter configurations. Consistent with earlier findings, the analysis confirms that the marginal survival function remains mathematically proper when $\kappa > 0$ but becomes invalid for $\kappa < 0$. Notably, the model's structure permits diverse hazard function profiles, encompassing unimodal patterns along with strictly increasing and decreasing trajectories. This comparative analysis reveals that the Gompertz-based QXg frailty model offers enhanced modeling versatility relative to its Weibull-based alternative.

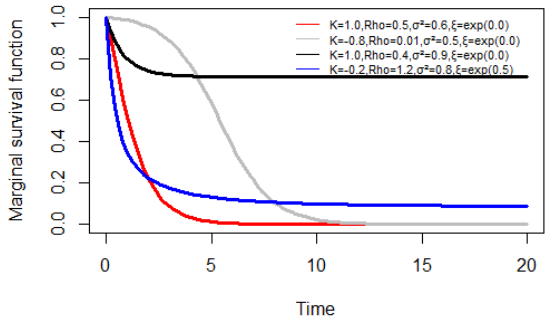

Figure 3.3: Plots of the Marginal Survival Function of the QXg-F Distribution whit Gompertz Baseline Hazard Functions.

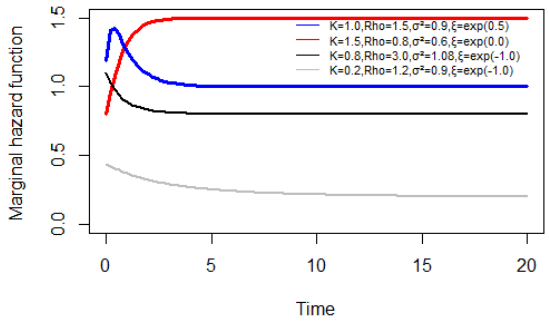

Figure 3.4: Plots of the Marginal Hazard Function of the QXg-F Distribution whit Gompertz Baseline Hazard Functions.

3.3 Estimating the Quasi Xgamma Frailty Model

This section presents the Maximum Likelihood Estimation (MaxLE) methodology for parameter estimation in the QXg-F model utilizing Weibull and Gompertz baseline hazard functions. The maximum likelihood estimators (MLEs) possess several desirable statistical properties under appropriate regularity conditions, including consistency, asymptotic efficiency, and asymptotic normality . A common practical consideration in survival analysis is the occurrence of incomplete lifetime observations. Specifically, certain study participants may provide right-censored data, where the exact failure time is unknown but confirmed to exceed the recorded observation time. Let T_i denote the lifetime and C_i the censoring time for the i^{th} subject in the study population, where $i = 1, \ldots, n$. We assume:

(A1) T_i and C_i are independent random variables

(A2) The censoring indicator is defined as $\delta_i = \mathbb{I}(T_i \leq C_i)$, where:

$$\delta_i = \begin{cases} 1 & \text{if the event time is observed (uncensored)} \\ 0 & \text{otherwise (censored)} \end{cases}$$

Consequently, the observed time for each subject is defined as $t_i = \min\{T_i, C_i\}$. Let $\mathbf{x}_i$ denote a $p \times 1$ covariate vector for the i^{th} subject. Under the assumption of non-informative censoring, the likelihood function for the parameter vector $\boldsymbol{\phi}$ based on a sample of n independent participants is given by:

$$\mathcal{L}(\boldsymbol{\phi}) = \prod_{i=1}^{n} \lambda(t_i \mid \mathbf{x}_i)^{\delta_i} S(t_i \mid \mathbf{x}_i) \tag{3.16}$$

where:

- $S(\cdot \mid \mathbf{x}_i)$ represents the survival function(3.12)
- $\lambda(\cdot \mid \mathbf{x}_i)$ denotes the hazard function (3.13)

3.3.1 Case of Weibull Baseline Hazard Function

The corresponding log-likelihood function is obtained by taking the natural logarithm of $\mathcal{L}(\boldsymbol{\phi})$. Assuming the Weibull baseline hazard function and using (3.12) and (3.13), For the parameter vector $\boldsymbol{\phi} = (\nu, \varphi, \sigma^2, \boldsymbol{\beta})$, this becomes:

$$\begin{aligned}
\log \mathcal{L}(\phi) &= \sum_{i=1}^{n} \left[\delta_i \log \lambda(t_i \mid \mathbf{x}_i) + \log S(t_i \mid \mathbf{x}_i)\right] \qquad (3.17)\\
&= r\log(\nu) + (\nu-1)\sum_{i=1}^{n}\delta_i \log(t_i) - Kr\log(\varphi) + \sum_{i=1}^{n}\delta_i x_i^T\beta \\
&\quad + \sum_{i=1}^{n}\delta_i \log\left[3 - 2\sigma^2 + \sqrt{13-12\sigma^2}\right] - \sum_{i=1}^{n}\delta_i \log\left(A_1\right) \\
&\quad + \sum_{i=1}^{n}\delta_i \log\left(B_1\right) - \sum_{i=1}^{n}\delta_i \log\left(C_1\right) + \sum_{i=1}^{n}\delta_i \log\left[1 + \sqrt{13-12\sigma^2}\right] \\
&\quad + \sum_{i=1}^{n}\delta_i \log\left(C_1\right) - \sum_{i=1}^{n}\delta_i \log\left[3 - 2\sigma^2 + \sqrt{13-12\sigma^2}\right] - \sum_{i=1}^{n}\delta_i 3\log\left(A_1\right);
\end{aligned}$$

where: $r = \sum_{i=1}^{n}\delta_i$ is the failure number,

$$A_1 = \left(\frac{t_i}{\varphi}\right)^{\nu} \exp(x_i^T\beta)\left(3 - 2\sigma^2 + \sqrt{13-12\sigma^2}\right) + 1 + \sqrt{13-12\sigma^2},$$

$$B_1 = \left\{ \begin{array}{c} \left(4 - 3\sigma^2 + \sqrt{13-12\sigma^2}\right) \\ \left[\begin{array}{c} \left(\frac{t_i}{\varphi}\right)^{\nu} \exp(x_i^T\beta)\left(3 - 2\sigma^2 + \sqrt{13-12\sigma^2}\right) \\ +1 + \sqrt{13-12\sigma^2} \end{array}\right]^2 \end{array} \right\} + 3\left(1 + \sqrt{13-12\sigma^2}\right)^2(\sigma^2 - 1),$$

and

$$C_1 = \left\{ \begin{array}{c} \left(4 - 3\sigma^2 + \sqrt{13-12\sigma^2}\right) \\ \times \left[\begin{array}{c} \left(\frac{t_i}{\varphi}\right)^{\nu} \exp(x_i^T\beta)\left(3 - 2\sigma^2 + \sqrt{13-12\sigma^2}\right) \\ +1 + \sqrt{13-12\sigma^2} \end{array}\right]^2 \end{array} \right\} + \left(1 + \sqrt{13-12\sigma^2}\right)^2(\sigma^2 - 1).$$

Setting the nonlinear system of the score equations $\mathbf{I}_{(\nu)} = 0, \mathbf{I}_{(\varphi)} = 0, \mathbf{I}_{(\sigma^2)} = 0$ and $\mathbf{I}_{(\beta)} = 0$ and solving them simultaneously yields the MaxLE $\widehat{\phi} = (\widehat{\nu}, \widehat{\varphi}, \widehat{\sigma^2}, \widehat{\beta})^{\intercal}$. It is usually more convenient to use nonlinear optimization methods to solve these equations; such as the quasi-Newton algorithm to numerically maximize $\log L(\phi)$.

3.3.2 Simulations: Case of Weibull Baseline Hazard Function

We examine the QXg-F model with Weibull baseline hazard function. The simulation study was conducted with the following specifications:

- Number of replications: $N = 13{,}000$
- True parameter values: $\nu = 0.8$, $\varphi = 0.9$, $\sigma^2 = 0.6$, $\beta_1 = 0.5$
- Sample sizes: $n \in \{20, 50, 300, 1000\}$

The simulations were implemented in R software using the Barzilai-Borwein (BB) optimization algorithm. For each scenario, we computed:

$$\text{Bias}(\widehat{\phi}) = \frac{1}{N}\sum_{i=1}^{N}(\widehat{\phi}_i - \phi) \tag{3.18}$$

$$\text{MSE}(\widehat{\phi}) = \frac{1}{N}\sum_{i=1}^{N}(\widehat{\phi}_i - \phi)^2 \tag{3.19}$$

where $\phi \in \{\nu, \varphi, \sigma^2, \beta_1\}$ represents the true parameter value and $\widehat{\phi}_i$ its estimate in the i^{th} replication.

Table 3.1 presents the average estimates and their corresponding MSEs across all replications. The results demonstrate that:

- The maximum likelihood estimators show convergence as sample size increases
- Both bias and MSE decrease with larger sample sizes
- The BB algorithm provides stable numerical solutions

These findings confirm the asymptotic consistency of the estimators for the QXg-F model with Weibull Baseline Hazard Function specification.

Table 3.1: Bias and MSE of the ML estimates for the simulated data of the QXg-F model with Weibull baseline hazard function.

n		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
		0%cens.		10%cens.		30%cens.		50%cens.	
20	φ	0.94368	0.04958	0.927084	0.04857	0.895344	0.03958	0.9535934	0.04977
	ν	0.83387	0.04758	0.800598	0.04932	0.818610	0.04109	0.764556	0.04067
	σ^2	0.64890	0.04768	0.5782824	0.04674	0.5872129	0.02999	0.5856559	0.04908
	β_1	0.56165	0.04927	0.4936702	0.04919	0.4890512	0.04164	0.521098	0.04916
50	φ	0.93728	0.04002	0.918837	0.04536	0.898274	0.03371	0.938401	0.04836
	ν	0.83100	0.03995	0.800416	0.04784	0.81178	0.02511	0.788897	0.04004
	σ^2	0.64201	0.04284	0.58813	0.04377	0.590121	0.02027	0.59604	0.04857
	β_1	0.55123	0.03879	0.495124	0.04509	0.496481	0.03366	0.520001	0.04806
300	φ	0.92418	0.03648	0.910006	0.03486	0.899901	0.02546	0.917234	0.04799
	ν	0.82431	0.02794	0.800372	0.02711	0.806742	0.02134	0.797896	0.04001
	σ^2	0.62745	0.02451	0.591302	0.01937	0.595342	0.00996	0.598732	0.04777
	β_1	0.51935	0.01867	0.497301	0.02232	0.498004	0.01021	0.514638	0.04738
1000	φ	0.91314	0.02374	0.900521	0.01658	0.90241	0.01340	0.905565	0.04500
	ν	0.80658	0.02006	0.800142	0.01745	0.80186	0.02001	0.801247	0.03844
	σ^2	0.61111	0.01865	0.61003	0.01008	0.602451	0.00618	0.599999	0.04627
	β_1	0.50041	0.01485	0.51008	0.00851	0.499989	0.00339	0.507430	0.03897

3.3.3 Case of Gompertz Baseline Hazard Function

For the QXg-F model with Gompertz Baseline Hazard Function, the log-likelihood function :$\phi = (\xi, \kappa, \sigma^2, \beta)$ is given as follows:

$$\begin{aligned}\log \mathcal{L}(\phi) &= r\log(\kappa) + \sum_{i=1}^{n}\delta_i(\xi t_i + x_i^T\beta) + \sum_{i=1}^{n}\delta_i \log\left[3 - 2\sigma^2 + \sqrt{\mathbb{C}\left(\sigma^2\right)}\right] \\ &\quad - \sum_{i=1}^{n}\delta_i\log(A_2) + \sum_{i=1}^{n}\delta_i\log(B_2) - \sum_{i=1}^{n}\delta_i\log\left(C_2\right) + \sum_{i=1}^{n}\log\left[1 + \sqrt{\mathbb{C}\left(\sigma^2\right)}\right] \\ &\quad + \sum_{i=1}^{n}\log(C_2) - \sum_{i=1}^{n}\log\left[3 - 2\sigma^2 + \sqrt{\mathbb{C}\left(\sigma^2\right)}\right] - \sum_{i=1}^{n}3\log(A_2),\end{aligned} \tag{3.20}$$

where

$$\mathbb{C}\left(\sigma^2\right) = 13 - 12\sigma^2,$$

$$A_2 = \left(\frac{\kappa}{\xi}\right)[\exp(\xi t_i) - 1]\exp(x_i^T\beta)\left[3 - 2\sigma^2 + \sqrt{\mathbb{C}\left(\sigma^2\right)}\right] + 1 + \sqrt{\mathbb{C}\left(\sigma^2\right)},$$

$$\begin{aligned}B_2 &= \left(\begin{array}{c}\left[4 - 3\sigma^2 + \sqrt{\mathbb{C}\left(\sigma^2\right)}\right] \\ \times\left\{\begin{array}{c}\left(\frac{\kappa}{\xi}\right)[\exp(\xi t_i) - 1]\exp(x_i^T\beta)\left[3 - 2\sigma^2 + \sqrt{\mathbb{C}\left(\sigma^2\right)}\right] \\ +1 + \sqrt{\mathbb{C}\left(\sigma^2\right)}\end{array}\right\}^2\end{array}\right) \\ &\quad + 3\left[1 + \sqrt{\mathbb{C}\left(\sigma^2\right)}\right]^2(\sigma^2 - 1),\end{aligned}$$

and

$$
\begin{aligned}
C_2 \;=\; & \left(\begin{array}{c} \left[4 - 3\sigma^2 + \sqrt{\mathbb{C}\left(\sigma^2\right)}\right] \\ \times \left\{ \begin{array}{c} \left(\frac{\kappa}{\xi}\right)\left[\exp(\xi t_i) - 1\right]\exp(x_i^T\beta)\left[3 - 2\sigma^2 + \sqrt{\mathbb{C}\left(\sigma^2\right)}\right] \\ +1 + \sqrt{\mathbb{C}\left(\sigma^2\right)} \end{array} \right\}^2 \end{array} \right) \\
& + \left[1 + \sqrt{\mathbb{C}\left(\sigma^2\right)}\right]^2 (\sigma^2 - 1).
\end{aligned}
$$

The maximum likelihood estimators $\widehat{\boldsymbol{\phi}}$ are derived through numerical optimization of the log-likelihood functions in Equations (3.18) and (3.21), as analytical solutions are unobtainable. We implement the Barzilai-Borwein algorithm via the `BBsolve` package in R for this nonlinear optimization problem. The iterative procedure converges to parameter estimates satisfying predetermined tolerance criteria while ensuring solution optimality. This methodology efficiently handles complex likelihood surfaces and facilitates statistical inference through numerical approximation of the Hessian matrix. Under standard regularity conditions, the resulting estimators demonstrate desirable asymptotic properties including consistency and normality.

3.3.4 Simulations: Case of Gompertz Baseline Hazard Function

We examine the QXg-F model with Gompertz baseline hazard function through comprehensive simulation studies. The experimental design incorporates:

- $N = 13{,}000$ Monte Carlo replications
- True parameter values: $\kappa = 1.07$, $\xi = 0.5$, $\sigma^2 = 0.7$, $\beta_1 = 0.3$
- Sample sizes: $n \in \{20, 50, 300, 1000\}$

Implementation was performed in R using the Barzilai-Borwein (BB) optimization algorithm . For each scenario, we compute:

$$\text{MSE}(\widehat{\phi}) = \frac{1}{N}\sum_{i=1}^{N}(\widehat{\phi}_i - \phi)^2 \tag{3.21}$$

where $\phi \in \{\kappa, \xi, \sigma^2, \beta_1\}$ denotes the true parameter value. Table 3.2 presents the empirical means and MSEs of the maximum likelihood estimators across all replications. The results demonstrate the consistent convergence of the estimators as sample size increases, with both bias and MSE exhibiting the expected asymptotic behavior.

Table 3.2: Bias and MSE of the ML estimates for the simulated data of the QXg-F model with Gompertz baseline hazard function.

n		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
		0% cens.		10% cens.		30% cens.		50% cens.	
20	κ	0.52772	0.04752	0.4812452	0.03258	0.493951	0.03696	0.4934327	0.04985
	ξ	1.06838	0.02935	1.0691513	0.03564	1.0684768	0.04812	1.0695826	0.04721
	σ^2	0.720627	0.01847	0.7326296	0.04446	0.7476699	0.02999	0.7235148	0.03954
	β_1	0.32221	0.04907	0.334851	0.04685	0.2834354	0.04770	0.2841911	0.04798
50	κ	0.513308	0.03251	0.487938	0.01357	0.496684	0.03496	0.4971112	0.04405
	ξ	1.069354	0.01524	1.0693819	0.01985	1.0687684	0.03914	1.0697468	0.03216
	σ^2	0.7162405	0.01847	0.717224	0.04021	0.7111120	0.01734	0.710743	0.01196
	β_1	0.317044	0.00965	0.328514	0.04612	0.293000	0.04382	0.290011	0.04380
300	κ	0.508674	0.01358	0.489064	0.01003	0.4980094	0.02685	0.498946	0.03945
	ξ	1.072641	0.00758	1.0696649	0.01876	1.068888	0.01493	1.0698437	0.02999
	σ^2	0.708657	0.00302	0.7063219	0.03681	0.706110	0.01112	0.6944002	0.01007
	β_1	0.309869	0.00063	0.312045	0.44378	0.298777	0.04128	0.290515	0.04268
1000	κ	0.501102	0.00648	0.495876	0.00882	0.5120436	0.02506	0.499997	0.03945
	ξ	1.070043	0.00012	1.069994	0.00586	1.069889	0.01002	1.0699254	0.00177
	σ^2	0.710030	0.00100	0.699589	0.02699	0.7003225	0.00659	0.700033	0.01001
	β_1	0.305201	0.00038	0.304061	0.44000	0.301066	0.04100	0.299999	0.04160

3.4 Validation Testing of the QXg-F Model

3.4.1 Validating the QXg-Frailty Model Using the NIK-RR Test

The NIK-RR test statistic serves as a comprehensive measure of goodness-of-fit for statistical models across multiple domains. As a generalized testing framework, it evaluates model adequacy for various specifications including survival analysis, regression models, and time series applications. The test's principal utilities encompass three key analytical functions: (1) model selection among competing specifications, (2) goodness-of-fit assessment for chosen models, and (3) diagnostic identification of model deficiencies. This versatility establishes the NIK-RR statistic as an essential tool for rigorous statistical evaluation. The methodological foundations and theoretical properties of the test are thoroughly established in the seminal works of Nikulin (1973a,b,c)[68, 69, 70] and Rao and Robson (1974)[76].

The NIK-RR test statistic provides superior detection of distributional deviations compared to conventional goodness-of-fit tests, particularly for non-normal data patterns. Its robust design maintains reliability even with extreme outliers, unlike many standard tests that are sensitive to anomalous values. These properties make the NIK-RR especially valuable for financial applications where accurately modeling tail events is critical, including market crashes and extreme price movements. The test's dual capabilities - sensitivity to subtle distributional changes while remaining resistant to outlier effects - offer unique advantages for risk assessment. Researchers and analysts frequently employ the NIK-RR when evaluating heavy-tailed distributions or datasets prone to extreme observations. Its effectiveness in these challenging scenarios has established the NIK-RR as an important tool for modern statistical analysis, particularly in econometrics and risk management applications.

The following are some uses and significance of the NIK-RR test statistic:

- The NIK-RR test statistic is useful for comparing the fit of various statistical models to the same dataset. This aids in model selection by identifying which model offers the best fit to the data.
- The NIK-RR test statistic serves to evaluate the goodness of fit of a statistical model to the data. A small NIK-RR test statistic suggests a good fit between the model and the data, while a large statistic indicates a poor fit.
- The NIK-RR test statistic can be employed to detect outliers in the data. Outliers are data points that deviate from the general pattern and can significantly affect the model's fit. The NIK-RR test helps identify these outliers, contributing to a better model fit.
- The NIK-RR test statistic can be utilized to diagnose issues within a statistical model. A large NIK-RR test statistic may indicate that the model is unspecified or that there are problems with its assumptions.

Under the NIK-RR statistic, we need to test the following null hypothesis

$$H_0 : \Pr\{z_i \leq z\} = F_\phi(z), \;\; z \in \mathbb{R}, \;\; \phi = (\phi_1, \phi_2, \cdots, \phi_s)^T, \tag{3.22}$$

Then, the NIK-RR statistic can be expressed as

$$Y^2(\widehat{\phi}_n) = X_n^2(\widehat{\phi}_n) + \frac{1}{n}\ell^T(\widehat{\mathbf{P}}_n)\left[\mathbf{I}(\widehat{\phi}_n) - \mathbf{J}(\widehat{\phi}_n)\right]^{-1}\ell(\widehat{\phi}_n), \tag{3.23}$$

where

$$X_n^2(\phi) = \left\{[np_1(\phi)]^{-\frac{1}{2}}\left[-np_1(\phi)+\phi_1\right], \cdots, [np_b(\phi)]^{-\frac{1}{2}}\left[-np_b(\phi)+\phi_b\right]\right\}^T$$

and

$$\mathbf{J}(\phi) = B(\phi)^T B(\phi), \tag{3.24}$$

refers to the information matrix where

$$B(\phi) = \left[\frac{1}{\sqrt{p}_i}\frac{\partial}{\partial\mu}(\phi)\right]_{r\times s}|_{(i=1,2,\cdots,b \text{ and } \nu=1,2,\cdots,s)}, \tag{3.25}$$

and

$$\ell(\phi) = [\ell_1(\phi), ..., \ell_s(\phi)]^T \text{ with } \ell_\nu(\phi) = \sum_{i=1}^{r}\frac{\phi_i}{p_i}\frac{\partial p_i(\phi)}{\partial\phi_\nu}, \tag{3.26}$$

The $Y^2(\widehat{\phi}_n)$ statistic has $(b-1)$ degrees of freedom (DF) and is accompanied by χ^2_{b-1} distribution, where the observations. $x_1, x_2, \cdots, x_n$ that are collected in $\mathbf{I}_1, \mathbf{I}_2, \cdots, \mathbf{I}_b$ (these b subintervals are mutually disjoint: $\mathbf{I}_j =]a_{j,b}-1; a_{j,b}]$). The intervals $\mathbf{I}_j$'s limits for $a_{j,b}$ are determined as follows

$$p_j(\phi) = \int_{a_{j,b}-1}^{a_{j,b}} f_\phi(x)\,dx|_{(\ j=1,2,\cdots,b)}, \tag{3.27}$$

and

$$a_{j,b} = F^{-1}\left(\frac{j}{b}\right)|_{(j=1,\cdots,b-1)}.$$

Uncensored Asessing for the NIK-RR Statistic

In many cases, the objective of a goodness-of-fit test is not only to assess whether a specific distribution fits the data but also to estimate its parameter values. Simulation studies can offer insights into the accuracy and precision of these parameter estimates under various scenarios, aiding in the decision-making process regarding which distribution to use for subsequent analyses. Overall, uncensored simulation studies under the NIK-RR statistics are a valuable tool for evaluating and comparing different probability distributions in a controlled environment. These studies can offer valuable insights into the performance of the NIK-RR tests across various scenarios and help inform decisions about which distribution to use for subsequent analyses. Through numerical simulation, we conducted a thorough analysis to validate the claims presented in this work.To verify the null hypothesis H_0, we calculated the NIK-RR statistics for the QXg-F model, confirming that the sample size is $13,000$ using simulated samples with sizes $n = 25, 50, 150, 350, 550$, and 1000. We examined various theoretical levels $(\epsilon = 0.01, 0.02, 0.05, 0.1)$ to compute the average number of non-rejections of the null hypothesis, where $Y^2 \leq \chi^2_\epsilon(b-1)$. The results, presented in Table 3.3, indicate a strong correlation between the calculated empirical level and its theoretical counterpart. Consequently, we conclude that the proposed test demonstrates considerable efficacy for the QXg-F distribution.

Table 3.3: Uncensored assessing for the NIK-RR statistic for $\epsilon = 0.01, 0.02, 0.05, 0.1$ and $N = 13000$.

$n \downarrow \epsilon \longrightarrow$	$\epsilon = 0.01$	$\epsilon = 0.02$	$\epsilon = 0.05$	$\epsilon = 0.1$
$n = 25$	0.9929	0.9825	0.9531	0.9029
$n = 50$	0.9927	0.9820	0.9527	0.9021
$n = 150$	0.9920	0.9813	0.9522	0.9012
$n = 350$	0.9911	0.9809	0.9510	0.9008
$n = 550$	0.9906	0.9805	0.9506	0.9004
$n = 1000$	0.9903	0.9804	0.9505	0.9002

3.4.2 Validating the QXg-Frailty Model Using the B-NIK Test

As established by Bagdonavicius and Nikulin (2011)[14] and Bagdonavicius et al. (2013)[15], the adequacy of the QXg-F model can be assessed under conditions of unknown parameters and censored data. Within this model, the null hypothesis may be expressed as follows:

$$H_0 : F(x) \in F_0 = \left\{ F_0(x, \phi), x \in R^1,\ \phi \in \phi \subset R^s \right\}, \tag{3.28}$$

Let's divide the limited amount of time $[0, \tau]$ into $\nu | \nu = 1, 2, \cdots, s$ shorter time periods. Where is the maximum runtime of the research and $\mathbf{I}_j = (a_{j-1}, a_{j,b}]$; $0 =< a_{0,b} < a_{1,b} ... < a_{\nu-1,b} < a_{\nu,b} = +\infty$. The anticipated worth of $\widehat{a_{j,b}}$ can be said the following if $x_{(i)}$ is the i^{th} element in the ordered statistics $(x_{(1)}, , , x_{(n)})$ and if Λ^{-1} refers to the cumulative hazard-rate function and

$$\widehat{a_{j,b}} = \Lambda^{-1} \left\{ \left[E_{j,X} - \sum_{l=1}^{i-1} \Lambda(x_{(l)}, \widehat{\phi}) \right] / (n - i + 1), \widehat{\phi} \right\}, \quad \widehat{a_\nu} = x_{(n)}|_{(j=1,...,\nu)}, \tag{3.29}$$

where

$$e_{j,Z} = E_\nu / \nu \text{ for every } j.$$

$$\Lambda(x, \phi) = -\ln \left[\begin{array}{c} \left[1 + \sqrt{\mathbb{C}\left(\sigma^2\right)}\right] \\ \left[4 - 3\sigma^2 + \sqrt{\mathbb{C}\left(\sigma^2\right)}\right] \\ \times \left(\times \left\{ \begin{array}{c} \Lambda_0(t_i) \exp(x_i^T \beta) \\ \left[3 - 2\sigma^2 + \sqrt{\mathbb{C}\left(\sigma^2\right)}\right] + 1 + \sqrt{\mathbb{C}\left(\sigma^2\right)} \end{array} \right\}^2 \right) \\ + \left[1 + \sqrt{\mathbb{C}\left(\sigma^2\right)}\right]^2 (\sigma^2 - 1) \\ \times \left(\left[3 - 2\sigma^2 + \sqrt{\mathbb{C}\left(\sigma^2\right)}\right] \left\{ \begin{array}{c} \Lambda_0(t_i) \exp(x_i^T \beta) \left[3 - 2\sigma^2 + \sqrt{\mathbb{C}\left(\sigma^2\right)}\right] \\ +1 + \sqrt{\mathbb{C}\left(\sigma^2\right)} \end{array} \right\}^3 \right)^{-1} \end{array} \right], \tag{3.30}$$

and

$$E_{j,Z} = (n-i+1)\Lambda(\widehat{a_{j,b}}, \widehat{\phi}) + \sum_{l=1}^{i-1} \Lambda(x_{(l)}, \widehat{\phi}) = \sum_{i:z_i > a_{j,b}} (\Lambda(a_{j,b} \wedge z_i, \widehat{\phi}) - \Lambda(a_{j-1}, \widehat{\phi}), E_\nu = \sum_{i=1}^{n} \Lambda(z_i, \widehat{\phi}). \tag{3.31}$$

The $a_{j,b}$ functions for random data, and the $e_{j,Z}$ For the ν selected periods, anticipated failure rates are equal. Statistical data $Y_n^2 = \mathbf{Z}^T \widehat{\mathbf{S}}^{-1} \mathbf{Z}$, where $\mathbf{Z} = (Z_1, ..., Z_\nu)^x$, $Z_j = \frac{1}{\sqrt{n}}(\mathbf{W}_{j,Z} - e_{j,Z})|_{(\ j=1,2,...,\nu)}$ and $\mathbf{W}_{j,Z}$ can be used to test a hypothesis since it reflects the total number of failures that have been recorded throughout these time-shared. The elements of the B-NIK test statistic

$$Y_n^2 = \sum_{j=1}^{\nu} \frac{1}{\mathbf{W}_{j,Z}} (\mathbf{W}_{j,Z} - e_{j,Z})^2 + \mathbf{D}_{W,G} \tag{3.32}$$

where

$$\begin{aligned}
\mathbf{D}_{W,G} &= \widehat{\mathbf{V}}^T \widehat{\mathbf{G}}^{-1} \widehat{\mathbf{V}}, \widehat{\mathbf{S}}^{-1} = \widehat{\mathbf{B}}^{-1} + \widehat{\mathbf{M}}^{-1} \widehat{\mathbf{B}}^T \widehat{\mathbf{G}}^{-1} \widehat{\mathbf{M}} \widehat{\mathbf{B}}^{-1}, \\
\widehat{\mathbf{G}} &= [\widehat{g}_{ll'}]_{s \times s} = \widehat{i} - \widehat{\mathbf{M}} \widehat{\mathbf{B}}^{-1} \widehat{\mathbf{M}}^x, \\
\widehat{\mathbf{M}}_{lj} &= \frac{1}{n} \sum_{i:z_i \in \mathbf{I}_j} \varphi_i \frac{\partial}{\partial \phi} \ln \left[\lambda_{i,\phi}(z_i)\right], \mathbf{W}_{j,Z} = \sum_{i:z_i \in \mathbf{I}_j} \varphi_i, \ \widehat{\mathbf{B}}_j = n^{-1} \mathbf{W}_{j,Z}, \\
\widehat{\mathbf{V}}_l &= \sum_{j=1}^{\nu} \widehat{\mathbf{M}}_{lj} \widehat{\mathbf{B}}_j^{-1} \mathbf{Z}_j, \qquad l, l' = 1, ..., s, \\
\widehat{i}_{ll'} &= n^{-1} \sum_{i=1}^{n} \varphi_i \frac{\partial}{\partial \phi_l} \ln \left[\lambda_{i,\phi}(z_i)\right] \frac{\partial}{\partial \phi_{l'}} \ln \left[\lambda_{i,\widehat{\mathbf{P}}}(z_i)\right]
\end{aligned}$$

and

$$\widehat{g}_{ll'} = \widehat{i}_{ll'} - \sum_{j=1}^{\nu} \widehat{\mathbf{M}}_{lj} \widehat{\mathbf{M}}_{l'j} \hat{A}_j^{-1},$$

and

$$\widehat{\mathbf{M}}_{lj} = \frac{1}{n} \sum_{i:z_i \in \mathbf{I}_j} \varphi_i \frac{\partial}{\partial \phi} \ln \left[\lambda_{i,\widehat{\phi}}(z_i)\right].$$

Censored Assessing for the B-NIK Statistic

Censored simulation studies utilizing the Bagdonavičius-Nikulin (B-NIK) statistics constitute an essential methodological approach for evaluating and comparing the adequacy of probability distributions when analyzing censored data. These systematic investigations offer crucial insights into the behavior and performance of B-NIK tests under various censoring scenarios, including different censoring types (e.g., right, left, or interval censoring) and varying degrees of censoring intensity. By rigorously examining distributional fits through these simulation frameworks, researchers can make informed decisions regarding the most appropriate probability models for their specific censored data applications, particularly in fields such as survival analysis, reliability engineering, and medical statistics. The findings from such studies not only enhance our understanding of distributional robustness but also contribute to improved statistical practice when handling incomplete data.The sample produced ($N = 13000$) is

intended to be censored at 25%, with degrees of freedom $DF = 5$. To assess whether the sample aligns with the null hypothesis of the QXg-F model, grouping intervals will be employed. For various theoretical levels, we will calculate the average value of the non-rejection numbers of the null hypothesis ($\epsilon = 0.01, 0.02, 0.05, 0.1$), where $Y^2 \leq \chi^2_\epsilon(r-1)$. Table 3.4 compares the theoretical and empirical levels, illustrating how closely the calculated empirical level aligns with the corresponding theoretical level. Consequently, we conclude that the customized test is well-suited for the QXg-F model.

Table 3.4: Censored assessing for the B-NIK statistic for $\epsilon = 0.01, 0.02, 0.05, 0.1$ and $N = 13000$.

$n \downarrow \& \epsilon \longrightarrow$	$\epsilon = 0.01$	$\epsilon = 0.02$	$\epsilon = 0.05$	$\epsilon = 0.1$
$n = 25$	0.9925	0.9829	0.9532	0.9019
$n = 50$	0.9920	0.9819	0.9525	0.9011
$n = 150$	0.9914	0.9813	0.9519	0.9008
$n = 350$	0.9909	0.9808	0.9510	0.9006
$n = 550$	0.9905	0.9804	0.9505	0.9004
$n = 1000$	0.9903	0.9803	0.9504	0.9002

From these findings, we conclude that the empirical significance level of Y_n^2 corresponds to the theoretical level of the chi-square distribution with degrees of freedom, indicating the statistical level at which it is significant. Thus, based on this evidence, the censored data obtained from the QXg-F distribution can be effectively fitted using the proposed test.

3.5 The Emergency Care Data

In the field of medical research, frailty models are employed as a statistical method to investigate risk factors and prognostic indicators associated with different diseases. These models enable the examination of individual-level covariates such as age, sex, and genetic markers while accounting for unobserved heterogeneity that may influence disease progression or mortality. Frailty models are widely applied in epidemiological investigations, clinical trials, and cohort studies to assess the efficacy of therapeutic interventions and assess differential patient outcomes. The current study utilized real-world clinical data obtained from the emergency department of the public proximity health institution in Echatt, El Tarf, Algeria. The dataset comprised patient records collected during March 2023. This study aimed to investigate the association between clinical parameters and emergency department outcomes among patients presenting for medical care. Following institutional ethical approval, data were collected from 30 distinct patient cases. For each case, six clinical variables were recorded: (1) age (years), (2) minimum and maximum blood pressure (mmHg), (3) blood glucose level (mg/dL), (4) cardiac frequency (BPM), (5) oxygen saturation (SaO2%) were measured as six different variables. Rigorous quality control protocols were implemented throughout data collection to ensure data integrity. These measures included: (1) standardized

patient documentation procedures, (2) strict adherence to validated measurement protocols, and (3) systematic quality assurance checks to identify and rectify any data inconsistencies or omissions.The rigorous data collection process and diverse patient population render this dataset particularly valuable for investigating relationships between clinical factors and emergency department outcomes. Our analysis focuses on evaluating the QXg-F distribution's goodness-of-fit and its ability to accurately model observed patterns and variability in emergency care data. This examination enables assessment of the distribution's validity and potential applications. We provide point estimates for two model specifications: (1) a QXg-F model with Weibull baseline hazard function, and (2) a QXg-F model with Gompertz baseline hazard function. The well known modified chi-squared test (Bagdonavičius and Nikulin (2011)) is supplied to identify the best model among all fitted models to this data.

3.5.1 Validation of the QXg-F Model Under the Weibull Baseline Hazard-Rate Function

Under the assumption that the observed data follow a QXg-F distribution with Weibull baseline hazard function, we estimated the parameter vector $\boldsymbol{\phi} = (\nu, \varphi, \sigma^2, \beta)^\top$ via maximum likelihood estimation (MLE). The likelihood function takes the form:

$$\mathcal{L}(\boldsymbol{\phi}) = \prod_{i=1}^{n} \left[h(t_i|\mathbf{x}_i)^{\delta_i} S(t_i|\mathbf{x}_i) \right] \tag{3.33}$$

where δ_i denotes the censoring indicator and $S(t_i|\mathbf{x}_i)$ represents the survival function.

The numerical optimization was performed in R (version 4.2.0) using the `BB` package, which implements the Barzilai-Borwein spectral method for large-scale nonlinear optimization. The resulting maximum likelihood estimates $\widehat{\boldsymbol{\phi}}$ were obtained by solving:

$$\widehat{\boldsymbol{\phi}} =_{\phi} \log \mathcal{L}(\boldsymbol{\phi}) \tag{3.34}$$

The estimation procedure incorporated:

- Convergence tolerance of 10^{-6} for all parameters
- Analytical gradient calculations
- Box constraints to ensure parameter feasibility

$\widehat{\boldsymbol{\phi}}$ are obtained as

$$\begin{aligned} \widehat{\nu} &= 0.93359597, \widehat{\varphi} = 0.92339747, \widehat{\sigma^2} = 1.05033488, \\ \widehat{\beta_1} &= 0.09416582, \widehat{\beta_2} = 0.12605622, \widehat{\beta_3} = -0.10260262, \\ \beta_4 &= -0.43375125, \widehat{\beta_5} = 0.21897114, \widehat{\beta_6} = 0.46708022. \end{aligned}$$

Following the methodology of Bagdonavičius and Nikulin (2011) [14] for censored data analysis, we consider ($r = 5$) intervals as the number of classes. The elements of the estimated Fisher information matrix $\mathbf{I}(\widehat{\phi})$ are given by:

$$I\left(\widehat{\phi}\right) = \begin{pmatrix} 1.10224 & -2.85216 & 0.35487 & -3.2015 & 5.12410 & -2.6585 & 0.75482 & -3.0002 & 1.95243 \\ & 0.982451 & -6.3214 & 2.73618 & 0.36485 & 0.51247 & -1.6832 & -4.12574 & 1.55505 \\ & & 0.62541 & 2.93485 & -6.3298 & 0.66485 & -21.5348 & 2.87361 & -7.0164 \\ & & & 1.32048 & 0.00843 & 1.31405 & 0.90340 & 1.11619 & -9.3761 \\ & & & & 2.96511 & -4.0527 & 1.6233 & 0.61372 & 0.8391 \\ & & & & & 0.84755 & 2.0006 & 1.86254 & 0.37770 \\ & & & & & & 0.32651 & -8.58102 & 1.7438 \\ & & & & & & & 1.70015 & -10.3254 \\ & & & & & & & & 0.74006 \end{pmatrix}$$

The calculated value of the test statistic is $Y_n^2 = 8.076$. Comparing this with the critical value $\chi^2_{0.05}(4) = 9.488$, we observe that $Y_n^2 < \chi^2_{0.05}(4)$. This result indicates that the proposed QXg-F model with Weibull baseline hazard function provides an adequate fit to the data at the 5% significance level.

3.5.2 Validation of the QXg-F Model Under the Gompertz Baseline Hazard-Rate Function

Assuming that the data adhere to a QXg-F distribution with a Gompertz baseline hazard function, we estimate the parameter vector ϕ using maximum likelihood estimation. The numerical optimization was conducted in R statistical software (version 4.2.1) utilizing the `BB` package, which provides robust optimization algorithms for nonlinear estimation issues. The resulting maximum likelihood estimates $\widehat{\phi}$ are as follows:

$$\begin{aligned} \widehat{\kappa} &= 1.00593154, \widehat{\xi} = 0.74417604, \widehat{\sigma^2} = 1.07010909, \\ \widehat{\beta_1} &= 0.16228331, \widehat{\beta_2} = 0.17405692, \widehat{\beta_3} = 0.02678224, \\ \widehat{\beta_4} &= -0.56805721, \widehat{\beta_5} = -0.07338661, \widehat{\beta_6} = 0.44793647. \end{aligned}$$

We take $r = 5$ intervals and the estimated Fisher matrix expressed as

$$I\left(\widehat{\phi}\right) = \begin{pmatrix} 2.6352 & -3.6257 & 1.0654 & 0.8475 & -6.3258 & -3.1547 & 2.0001 & 0.9321 & 1.8547 \\ & 0.9658 & 0.8475 & 1.8654 & 1.3274 & -3.9547 & -8.1755 & -11.132 & 1.9584 \\ & & 1.3475 & 2.0084 & 0.6254 & 1.3648 & -2.3647 & 0.6845 & 0.9614 \\ & & & 0.6321 & -9.3020 & 3.0001 & 1.0854 & 0.6845 & -3.628 \\ & & & & 1.6847 & 1.0054 & 0.3754 & 1.0024 & 1.9045 \\ & & & & & 0.3617 & 1.6584 & 0.7845 & 1.0325 \\ & & & & & & 0.8647 & 0.1254 & 0.9658 \\ & & & & & & & 1.0954 & 1.8457 \\ & & & & & & & & 1.3643 \end{pmatrix},$$

Following the methodology of Bagdonavičius and Nikulin (2011)[14], we calculate the goodness-of-fit statistic $Y_n^2 = 7.43821106$. Comparing this with critical values at two significance levels : $\alpha = 5\%$ and $\alpha = 10\%$, we find $Y^2 < \chi^2_{0.05}(5-1) = 9.488$ and $Y^2 < \chi^2_{0.1}(5-1) = 7.779$ respectively. where the degrees of freedom (4) correspond to $r-1$ for $r = 5$ intervals. These results demonstrate that the emergency care data are consistent with our proposed QXg-F model specification using a Gompertz baseline hazard function at both significance levels.

3.6 Testing the Ability of the New QXg-Frailty in Risk Analysis

Frailty models are employed to predict the probability of events occurring, such as the risk of developing a specific disease or experiencing an adverse event, while also accounting for individual characteristics that may influence this risk. For example, a frailty model can be used to estimate the risk of hospitalization due to a particular medical condition.This approach would consider the variations in risk among individuals due to factors such as age, gender, and the presence of other medical conditions. To model risks and develop risk prediction models, frailty models have been used across various fields, including epidemiology, medical research, and actuarial science, among others. Employing frailty models can yield more precise risk estimates and enhance the understanding of the underlying mechanisms that influence the likelihood of an event occurring. Frailty models simultaneously address both unobserved heterogeneity and covariate effects, making them particularly valuable in actuarial science. These models account for variability in mortality rates attributable to observed factors (age, gender, health status) and latent risk factors (genetic predisposition, unmeasured lifestyle effects). By incorporating random effects, they provide more accurate risk stratification than traditional survival models. This dual capability is essential for pricing insurance products and reserving calculations. Their flexibility has established frailty models as a gold standard for mortality analysis in actuarial applications. Insurers can more effectively price their insurance products by utilizing frailty models, which also helps them manage the risks they encounter. In the realms of insurance and actuarial science, specific applications of frailty models include:

1. Frailty models can be used to estimate the likely future mortality of a group of individuals, which can then inform the pricing of life insurance premiums.
2. The present value of an annuity is the amount an individual needs to invest today to ensure a guaranteed income stream for their lifetime. To estimate this present value, frailty models can be employed, determining the necessary investment for acquiring such an income stream.
3. Frailty models can aid in identifying and managing the risks associated with a particular insurance product or portfolio. For example, an insurer might use a frailty model to pinpoint high-risk individuals who are more likely to submit claims.
4. The use of frailty models is a valuable tool that can improve the accuracy and efficiency of insurance pricing and risk management. As actuarial science evolves, frailty models are expected to see growing adoption within the insurance industry.

The use of frailty models in actuarial science and insurance offers several additional advantages:

1. Insurers can gain valuable insights into the risk of loss associated with specific individuals or groups by using frailty models. This understanding can result in more precise decisions related to pricing and underwriting.

2. By utilizing frailty models, insurance companies can develop solutions that better align with the needs of individual clients. This approach may enhance customer satisfaction and loyalty.

3. By using frailty models, insurers can automate a large part of the tasks involved in risk assessment and pricing. This can result in cost reductions and greater operational effectiveness.

3.6.1 Risk Indicators for the New QXg-F Model

No additional characterization of risk exposure is needed beyond what probability-based distributions can offer. Typically, a single value or a limited set of numbers describes the level of risk exposure. These risk exposure metrics are clearly dependent on a specific model.These are often called key risk indicators (KRIs), which stands for key risk indicator. Such indicators provide actuaries and risk managers with insights into the degree of a company's exposure to specific types of risk.A wide range of KRIs can be considered and analyzed, including V-R, TLV-R (also known as CVaR), T-VC, TM-V, and M-EXL, among others. Specifically, the Value Reconstruction Kernel (V-R) represents a quantile of the distribution of total losses. Actuaries and risk managers often focus on assessing the likelihood of negative outcomes, which can be conveyed using the V-R indicator at a defined probability or confidence level.

This indicator is often utilized to determine the capital required to manage potential negative scenarios. The V-R of the QXg-F model at the $100q\%$ level, denoted as V-R$(Z;\phi)$ or $\pi(q)$, represents the $100q\%$ quantile (or percentile). Thus, we can express it simply as:

$$\text{V-R}\left(Z;\phi\right)=\Pr\left(Z>Q\left(U\right)\right)=\begin{cases}1\%|_{q=99\%}\\ 5\%|_{q=95\%}\\ \vdots\end{cases}, \tag{3.35}$$

In this context, $Q(U)$ denotes the quantile function of the QXg-F model. For a one-year timeframe with $q = 99\%$, it indicates that there is only a minimal chance (1%) of the insurance company facing bankruptcy due to an adverse event within the next year. Generally, if gains (or losses) follow a normal distribution, the value V-R$(Z;\phi)$ is considered to meet all coherence criteria. However, insurance data, such as claims and reinsurance revenues, often exhibit skewness, either to the right or left. Consequently, utilizing the normal distribution to represent these revenues is not appropriate. The TLV-R$(Z;\phi)$ at the $100q\%$ confidence level represents the expected loss given that the loss exceeds the $100q\%$ threshold of the distribution of Z. Thus, the TLV-R of Z can be expressed as:

$$\text{TLV-R}\left(Z;\phi\right)=\mathbb{E}\left(X|X>\pi\left(q\right)\right)=\frac{1}{1-F_{\mathbf{V}}\left(\pi\left(q\right)\right)}\int\limits_{\pi(q)}^{\infty}z\,f_{\phi}\left(z\right)dz=\frac{1}{1-q}\int\limits_{\pi(q)}^{\infty}z\,f_{\phi}\left(z\right)dz. \tag{3.36}$$

,

The quantity TLV-R$(Z;\phi)$ offers additional insights into the tail of the QXg-F distribution and represents the average of all V-R values at the confidence level q. Furthermore, TLV-R(X) can be expressed as

$$\text{TLV-R}(Z;\phi)=e(Z;\phi)+\text{V-R}(Z;\phi),$$

where $e(Z;\phi)$ is the mean excess loss (M-EXL) function evaluated at the $100q\%$ quantile (see Acerbi and Tasche (2002)[6]). When $e(Z;\phi)$ approaches zero, it follows that

$$\text{TLV-R}(Z;\phi) = \text{V-R}(Z;\phi).$$

For very small values of $e(Z;\phi)$, the TLV-R will be almost equal to the V-R. The T-VC risk indicator, created by Furman and Landsman (2006)[32], assesses how much the loss deviates from the average along the tail. Furman and Landsman (2006)[32] also developed explicit formulas for the T-VC risk indicator under the multivariate normal distribution. The T-VC risk indicator $\text{T-VC}(Z;\phi)$ can thus be expressed as:

$$\text{T-VC}\left(Z;\phi\right) = \mathbb{E}\left(Z^2 | Z > \pi\left(q\right)\right) - \left[\text{TLV-R}(Z)\right]^2.$$

As a statistic for the best portfolio choice, Landeveloped the TMVK risk indicator, which is based on the T-VC risk indicator. Consequently, the TMVK risk indicator may be written as

$$\text{TMVK}(X) = \text{TLV-R}(Z) + \pi\text{T-VC}(Z)|_{0<\pi<1}.$$

Then, for any RV, $\text{TMVK}(Z;\phi)$ >$\text{T-VC}(Z;\phi)$ and, for $\pi = 1$, $\text{TMVK}(Z;\phi)$ =$\text{TLV-R}(Z;\phi)$.

3.6.2 Assessing Under Different Estimation Methods via Simulations

In this section, we evaluate the MaxLE, OLS, WLSE, and AnDE methods for calculating the Key Risk Indicators (KRIs). These quantities are estimated using $N = 1,000$ across different sample sizes ($n = 20, 50, 100$) and three confidence levels ($q = (50\%, 60\%, 70\%, 90\%, 99\%)$). The results are presented in Table 3.5–3.6 (KRIs for artificial data with $n = 20$), Table 3.7–3.8 (KRIs for $n = 50$), and Table 3.9**??** (KRIs for $n = 100$), from which we draw the following conclusions: $\text{V-R}(Z;\phi)$, $\text{TLV-R}(Z;\phi)$ and $\text{TMVK}(Z;\phi)$ increase when q increases for all estimation methods. $\text{V-R}(Z;\phi)_{\text{WLS}}$ <$\text{V-R}(Z;\phi)_{\text{AnDE}}$ <$\text{V-R}(Z;\phi)_{\text{MaxLE}}$ < $\text{V-R}(Z;\phi)_{\text{OrLSE}}$ for most q. $\text{TLV-R}(Z;\phi)_{\text{WLS}}$ <$\text{TLV-R}(Z;\phi)_{\text{AnDE}}$ <$\text{TLV-R}(Z;\phi)_{\text{MaxLE}}$ <$\text{TLV-R}(Z;\phi)_{\text{OrLSE}}$ for most q.

Table 3.5: A.Key Risk Indicators (KRIs) Generated from Synthetic Data (Sample Size: n=20).

Method	$\widehat{\alpha}$	$\widehat{\vartheta}$	V-R$(Z;\phi)$	TLV-R$(Z;\phi)$	T-VC$(Z;\phi)$	TMVK$(Z;\phi)$	MELS$(Z;\phi)$
MaxLE	0.14107	0.52017					
50%			4.6982252	7.9207537	8.1291089	11.9853081	3.2225285
60%			5.5534017	8.6216855	7.6896604	12.4665157	3.0682838
70%			6.5570515	9.4827166	7.2595049	13.112469	2.9256651
80%			7.8541408	10.6385741	6.8116389	14.0443935	2.7844333
90%			9.8829563	12.5095433	6.2864654	15.652776	2.626587
95%			11.769106	14.2920432	5.9283057	17.256196	2.5229372
99%			15.8459571	18.2225931	5.4063186	20.9257524	2.376636
OrLSE				0.25768,0.49028			
50%			4.6424595	8.1062567	9.3000559	12.7562847	3.4637972
60%			5.5685106	8.8587309	8.7761598	13.2468108	3.2902204
70%			6.6502786	9.7810246	8.2666565	13.9143528	3.1307459
80%			8.0427591	11.0166329	7.7395836	14.8864248	2.9738738
90%			10.2129492	13.0127723	7.1258076	16.5756761	2.7998232
95%			12.2254069	14.911675	6.7097639	18.2665569	2.6862681
99%			16.5664285	19.0933974	6.1069945	22.1468946	2.5269688

Table 3.6: B.Key Risk Indicators (KRIs) Generated from Synthetic Data (Sample Size: n=20).

Method	$\widehat{\alpha}$	$\widehat{\vartheta}$	V-R($Z;\phi$)	TLV-R($Z;\phi$)	T-VC($Z;\phi$)	TMVK($Z;\phi$)	MELS($Z;\phi$)
WLSE					0.27324,0.48524		
50%			4.6474756	8.1526927	9.5131474	12.9092664	3.5052172
60%			5.5853582	8.9140626	8.9747332	13.4014292	3.3287043
70%			6.6804225	9.8470259	8.4514401	14.072746	3.1666034
80%			8.0893887	11.096634	7.9104722	15.0518701	3.0072453
90%			10.2843424	13.1149134	7.2810078	16.7554173	2.830571
95%			12.3191384	15.0345255	6.8546307	18.4618409	2.7153871
99%			16.7072472	19.2611659	6.2373245	22.3798281	2.5539186
CVM					0.24569,0.49695		
50%			4.6131432	8.0262669	9.0380238	12.5452788	3.4131237
60%			5.5250504	8.7678136	8.5307887	13.033208	3.2427632
70%			6.5907158	9.676896	8.0372303	13.6955112	3.0861803
80%			7.9629605	10.8950332	7.526371	14.6582187	2.9320728
90%			10.102331	12.8633142	6.9310958	16.3288621	2.7609832
95%			12.0866897	14.7359852	6.5273691	17.9996698	2.6492955
99%			16.3679401	18.8604668	5.9421197	21.8315266	2.4925267

Table 3.7: A.Key Risk Indicators (KRIs) Generated from Synthetic Data (Sample Size: n=50).

Method	$\widehat{\alpha}$	$\widehat{\vartheta}$	V-R($Z;\phi$)	TLV-R($Z;\phi$)	T-VC($Z;\phi$)	TMVK($Z;\phi$)	MELS($Z;\phi$)
MaxLE					0.18714,0.50706		
50%			4.6845211	8.0084233	8.6125268	12.3146867	3.3239022
60%			5.5694373	8.7310154	8.1386873	12.800359	3.161578
70%			6.6058505	9.6178117	7.6762239	13.4559236	3.0119612
80%			7.9429961	10.8072504	7.1960754	14.4052881	2.8642543
90%			10.0313044	12.7310065	6.6347243	16.0483686	2.699702
95%			11.9707151	14.5626636	6.2528722	17.6890997	2.5919485
99%			16.1591777	18.5994275	5.6977361	21.4482955	2.4402497
OrLSE					0.21376,0.49717		
50%			4.7007731	8.1008547	8.9916608	12.5966851	3.4000817
60%			5.6075032	8.8397982	8.492301	13.0859487	3.2322951
70%			6.668335	9.7462022	8.0056485	13.7490265	3.0778672
80%			8.0357406	10.9613828	7.5011306	14.711948	2.9256421
90%			10.1695559	12.9258941	6.9122092	16.3819986	2.7563381
95%			12.1500806	14.7957162	6.5121632	18.0517978	2.6456356
99%			16.4253531	18.9153562	5.9313583	21.8810353	2.4900031

Table 3.8: B.Key Risk Indicators (KRIs) Generated from Synthetic Data (Sample Size: n=50).

Method	$\widehat{\alpha}$	$\widehat{\vartheta}$	V-R($Z;\phi$)	TLV-R($Z;\phi$)	T-VC($Z;\phi$)	TMVK($Z;\phi$)	MELS($Z;\phi$)
WLSE					0.22011,0.49643		
50%			4.689736	8.0972504	9.0263673	12.6104341	3.4075144
60%			5.5987983	8.8377616	8.5239999	13.0997615	3.2389633
70%			6.662104	9.7459829	8.034577	13.7632714	3.083879
80%			8.0324131	10.9634708	7.5273556	14.7271486	2.9310577
90%			10.1703532	12.9315076	6.9354919	16.3992535	2.7611544
95%			12.1544387	14.8045359	6.5335763	18.071324	2.6500971
99%			16.4369338	18.9309503	5.9502389	21.9060698	2.4940165
CVM					0.20859,0.49996		
50%			4.6892727	8.0684722	8.8853283	12.5111364	3.3791995
60%			5.5901462	8.8029166	8.3927509	12.999292	3.2127704
70%			6.6443362	9.7038884	7.9125764	13.6601766	3.0595522
80%			8.0034148	10.9118918	7.41464	14.6192118	2.908477
90%			10.1245679	12.8649676	6.8332241	16.2815797	2.7403997
95%			12.0935586	14.7240273	6.4381719	17.9431133	2.6304687
99%			16.3443116	18.8201913	5.8644697	21.7524262	2.4758797

Table 3.9: A.Key Risk Indicators (KRIs) Generated from Synthetic Data (Sample Size: n=100).

Method	$\widehat{\alpha}$	$\widehat{\vartheta}$	V-R($Z;\phi$)	TLV-R($Z;\phi$)	T-VC($Z;\phi$)	TMVK($Z;\phi$)	MELS($Z;\phi$)
MaxLE				0.1943,0.50444			
50%			4.6882397	8.0320981	8.7108906	12.3875434	3.3438585
60%			5.5788848	8.7589716	8.2304017	12.8741725	3.1800868
70%			6.6216992	9.650898	7.7616432	13.5317196	3.0291987
80%			7.9667667	10.8470672	7.2751553	14.4846448	2.8803005
90%			10.0669753	12.7814728	6.7066399	16.1347928	2.7144975
95%			12.0171294	14.6230983	6.3200621	17.7831293	2.6059689
99%			16.2282628	18.6815004	5.7582622	21.5606315	2.4532376
OrLSE				0.21376,0.49885			
50%			4.6849052	8.0735006	8.9310096	12.5390054	3.3885954
60%			5.5885719	8.8099477	8.4350185	13.027457	3.2213759
70%			6.6458198	9.7132897	7.9516488	13.6891141	3.0674699
80%			8.008606	10.924365	7.4505375	14.6496338	2.915759
90%			10.1352129	12.8822401	6.8655888	16.3150345	2.7470271
95%			12.1090473	14.7457459	6.4682413	17.9798666	2.6366986
99%			16.3698779	18.8514699	5.8913542	21.797147	2.481592

Table 3.10: B.Key Risk Indicators (KRIs) Generated from Synthetic Data (Sample Size: n=100).

Method	$\widehat{\alpha}$ $\widehat{\vartheta}$	V-R($Z;\phi$)	TLV-R($Z;\phi$)	T-VC($Z;\phi$)	TMVK($Z;\phi$)	MELS($Z;\phi$)
WLSE				0.21105,0.49985		
50%		4.6833079	8.0641711	8.8922752	12.5103087	3.3808631
60%		5.5847622	8.7989583	8.3988943	12.9984054	3.2141961
70%		6.6395315	9.7003094	7.9179995	13.6593092	3.0607779
80%		7.9992451	10.9087709	7.419382	14.6184619	2.9095258
90%		10.1212311	12.8625078	6.8372537	16.2811347	2.7412767
95%		12.0908908	14.7221388	6.4417673	17.9430224	2.6312479
99%		16.3429079	18.8194491	5.8675051	21.7532016	2.4765411
CVM				0.21109,0.50028		
50%		4.6791079	8.0570562	8.8769135	12.495513	3.3779483
60%		5.5797877	8.7912095	8.3843767	12.9833979	3.2114219
70%		6.6336488	9.6917823	7.904306	13.6439353	3.0581335
80%		7.9921894	10.8991994	7.4065397	14.6024693	2.90701
90%		10.1123418	12.8512459	6.8254163	16.263954	2.738904
95%		12.0802975	14.7092668	6.4306107	17.9245721	2.6289693
99%		16.3286324	18.8030275	5.8573383	21.7316967	2.4743951

3.6.3 Assessing Under Different Estimation Methods via Insurance Data

As a specific example, this section of the essay examines the insurance claims payment triangle for a United Kingdom Motor Non-Comprehensive account. We identify the years 2007 to 2013 as the most suitable origin period.The claims data is structured within the insurance claims payment data frame in a format typical for databases. The first column contains details about the origin year, which spans from 2007 to 2013, as well as the development year and incremental payments. Notably, the insurance claims data were initially examined using a probability-based distribution.This analysis was conducted at the outset of the process. The insurance company's capacity to manage such events is highly valued by actuaries, regulators, investors, and rating agencies. This study proposes several Key Risk Indicator (KRI) metrics for the left-skewed insurance claims data based on the QXg-F distribution. These metrics include VAR, TVAR, T-VC, and TM-V (refer to Artzner (1999)[**?**] for further details). Table 8 presents the KRIs derived from the insurance calculations data, categorized by each estimation method employed for the QXg-F model. Based on Table 8, the following results can be highlighted:

1. For all risk assessment methods:

$$\text{V-R}(Z;\widehat{\phi}|_{1-q=50\%}) < \text{V-R}(Z;\widehat{\phi}|_{1-q=40\%}) < ... < \text{V-R}(Z;\widehat{\phi}|_{1-q=5\%}) < \text{V-R}(Z;\widehat{\mathbf{P}}|_{1-q=1\%}).$$

2. For all risk assessment methods:

$$\text{TLV-R}(Z;\widehat{\phi}|_{1-q=50\%}) < \text{TLV-R}(Z;\widehat{\phi}|_{1-q=40\%}) < ... < \text{TLV-R}(Z;\widehat{\phi}|_{1-q=5\%}) < \text{TLV-R}(Z;\widehat{\phi}|_{1-q=1\%}).$$

3. For all risk assessment methods:

$$\text{T-VC}(Z;\widehat{\phi}|_{1-q=50\%}) > \text{T-VC}(Z;\widehat{\phi}|_{1-q=40\%}) > ... > \text{T-VC}(Z;\widehat{\phi}|_{1-q=5\%}) > \text{T-VC}(Z;\widehat{\phi}|_{1-q=1\%}).$$

4. For all risk assessment methods:

$$\text{TMVK}(Z;\widehat{\phi}|_{1-q=50\%}) > \text{TMVK}(Z;\widehat{\phi}|_{1-q=40\%}) > ... > \text{TMVK}(Z;\widehat{\phi}|_{1-q=5\%}) > \text{TMVK}(Z;\widehat{\phi}|_{1-q=1\%}).$$

5. For all risk assessment methods:

$$\text{MELS}(Z;\widehat{\phi}|_{1-q=50\%}) > \text{MELS}(Z;\widehat{\phi}|_{1-q=40\%}) > ... > \text{MELS}(Z;\widehat{\phi}|_{1-q=5\%}) > \text{MELS}(Z;\widehat{\phi}|_{1-q=1\%}).$$

6. Under the MaxLE method: The $\text{V-R}\left(Z;\widehat{\phi}\right)$ is monotonically increasing indicator, the $\text{TLV-R}\left(Z;\widehat{\phi}\right)$ in monotonically increasing indicator. However the $\text{T-VC}\left(Z;\widehat{\phi}\right)$, the $\text{TMVK}\left(Z;\widehat{\phi}\right)$ and the $\text{MELS}\left(Z;\widehat{\phi}\right)$ are monotonically decreasing.

7. Under the OrLSE method: The $\text{V-R}\left(Z;\widehat{\phi}\right)$ is monotonically increasing indicator, the $\text{TLV-R}\left(Z;\widehat{\phi}\right)$ in monotonically increasing indicator. However the $\text{T-VC}\left(Z;\widehat{\phi}\right)$, the $\text{TMVK}\left(Z;\widehat{\phi}\right)$ and the $\text{MELS}\left(Z;\widehat{\phi}\right)$ are monotonically decreasing.

8. Under the WLSE method: The $\text{V-R}\left(Z;\widehat{\phi}\right)$ is monotonically increasing indicator, the $\text{TLV-R}\left(Z;\widehat{\phi}\right)$ in monotonically increasing indicator. However the $\text{T-VC}\left(Z;\widehat{\phi}\right)$, the $\text{TMVK}\left(Z;\widehat{\phi}\right)$ and the $\text{MELS}\left(Z;\widehat{\phi}\right)$ are monotonically decreasing.

9. Under the AE method: The $\text{V-R}\left(Z;\widehat{\phi}\right)$ is monotonically increasing indicator, the $\text{TLV-R}\left(Z;\widehat{\phi}\right)$ in monotonically increasing indicator. However the $\text{T-VC}\left(Z;\widehat{\phi}\right)$, the $\text{TMVK}\left(Z;\widehat{\phi}\right)$ and the $\text{MELS}\left(Z;\widehat{\phi}\right)$ are monotonically decreasing.

10. For nearly all q values, the OrLSE method is preferred, as it offers the most reliable risk exposure analysis, followed by the MaxLE method as the second choice. Nevertheless, the other two methods also perform well.

Table 3.11: A.Key Risk Indicators (KRIs) Generated from Synthetic Data.

Method	$\widehat{\alpha}$ $\widehat{\vartheta}$	V-R$(Z;\phi)$	TLV-R$(Z;\phi)$	T-VC$(Z;\phi)$	TMVK$(Z;\mathbf{P})$	MELS$(Z;\phi)$
MaxLE				0.056648,0.00107		
50%		2408.285176	3957.286678	1894484.692667	951199.633011	1549.001501
60%		2816.579287	4294.591947	1795766.569883	902177.876889	1478.012661
70%		3297.89361	4709.755039	1698501.924311	853960.717195	1411.861429
80%		3922.142321	5268.031519	1596614.004881	803575.03396	1345.889198
90%		4901.547418	6173.031474	1478243.31119	745294.687069	1271.484056
95%		5814.002477	7036.477935	1395540.819773	704806.887822	1222.475459
99%		7789.512693	8942.216564	1278264.671458	648074.552293	1152.703872
OrLSE				0.272949,0.000899		
50%		2509.021497	4400.995807	2771619.894328	1390210.942971	1891.97431
60%		3015.24518	4811.953608	2614766.380651	1312195.143934	1796.708428
70%		3606.312937	5315.532936	2462318.334862	1236474.700367	1709.219999
80%		4366.817414	5990.028363	2304719.229173	1158349.64295	1623.210949
90%		5551.576191	7079.430874	2121334.380204	1067746.620976	1527.854683
95%		6649.89369	8115.579457	1997119.012415	1006675.085665	1465.685767
99%		9018.464769	10396.999002	1817273.115655	919033.556829	1378.534233

Table 3.12: B.Key Risk Indicators (KRIs) Generated from Synthetic Data.

Method	$\widehat{\alpha}$	$\widehat{\vartheta}$	V-R$(Z;\phi)$	TLV-R$(Z;\phi)$	T-VC$(Z;\phi)$	TMVK$(Z;\mathbf{P})$	MELS$(Z;\phi)$
WLSE					0.13893,0.001019		
50%			2402.375177	4047.542671	2119132.335979	1063613.710661	1645.167494
60%			2838.891506	4405.392352	2004673.486117	1006742.13541	1566.500846
70%			3351.247105	4844.998236	1892621.739739	951155.868105	1493.751131
80%			4013.456607	5435.154825	1775934.490963	893402.400306	1421.698218
90%			5049.313766	6390.469523	1639088.326576	825934.632811	1341.155757
95%			6012.378498	7300.638908	1545748.650037	780174.963926	1288.26041
99%			8094.095078	9307.560124	1412012.076223	715313.598236	1213.465046
CVM					0.215302,0.000933		
50%			2503.029113	4315.472068	2554671.581296	1281651.262716	1812.442955
60%			2986.414861	4709.366036	2412721.162258	1211069.947165	1722.951174
70%			3551.920461	5192.511163	2274394.097108	1142389.559717	1640.590702
80%			4280.816611	5840.228281	2130999.793612	1071340.125087	1559.411671
90%			5418.195151	6887.328172	1963634.040249	988704.348297	1469.133022
95%			6473.831185	7883.938659	1849952.187187	932860.032253	1410.107475
99%			8752.526684	10079.658969	1684916.244941	852537.78144	1327.132285

3.7 Discussion of Findings and Directions for Future Research

3.7.1 Discussion of Main Findings

This chapter introduced and evaluated the **quasi-Xgamma frailty model** (QXg-F), a new approach for modeling unobserved heterogeneity in survival data. Traditional frailty models, like the gamma or log-normal, are widely used to represent individual differences in risk that cannot be explained by observed covariates. However, these models can be restrictive in shape and tail behavior. The QXg-F model was proposed as a more flexible alternative, capable of capturing more complex forms of variation.

The model is based on the **quasi-Xgamma distribution**, which extends the classical Xgamma family by including an additional shape parameter. This gives it greater flexibility to model positively skewed data and a wider range of kurtosis levels. In the context of survival analysis, this means the QXg-F model can account for both mild and severe heterogeneity in risk, which is often present in medical, biological, and insurance-related datasets.

One of the key strengths of the QXg-F model is its **mathematical tractability**. The distribution has a closed-form Laplace transform, which makes it possible to derive explicit formulas for the marginal survival and hazard functions. These expressions are essential for building the likelihood function and estimating model parameters, especially in the presence of censored data.

To estimate parameters, the **maximum likelihood estimation** (MLE) method was used. Because the likelihood function for the QXg-F model does not have a closed-form solution, we used numerical optimization-specifically, the Barzilai–Borwein algorithm. Simulation studies showed that this method performs well in practice. Across multiple sample sizes and censoring levels, the estimates were accurate, stable, and showed decreasing bias and mean squared error as sample size increased. Even with 30–50% censoring, the estimators remained reliable.

To check how well the model fits real data, we used two statistical tests. The **Nikulin–Rao–Robson** (NIK-RR) test was applied to complete data, and the **Bagdonavicius–Nikulin** (B-NIK) test was used for censored data. Both tests confirmed that the QXg-F model matched the data well. Importantly, simulation studies confirmed that the observed significance levels of these tests closely matched theoretical expectations, which means the model does not overfit or underfit systematically.

A real dataset from an emergency department in Algeria was used to demonstrate the model's practical utility. This dataset included 30 patients and several clinical variables: age, blood pressure, glucose level, cardiac frequency, and oxygen saturation. We fitted two versions of the QXg-F model—one using a **Weibull** baseline hazard and the other using a **Gompertz** baseline. Both models gave a good fit to the data. In both cases, the estimated parameters were clinically plausible. For example, higher glucose levels and lower oxygen saturation were associated with higher hazard, consistent with medical knowledge.

Interestingly, the choice of baseline hazard affected some results. The Weibull version was better at modeling increasing hazard over time, while the Gompertz version was more responsive to early risks. This suggests that while the QXg-F frailty term captures unobserved variability, careful choice of the baseline hazard is still important for modeling different clinical situations.

Beyond clinical data, the QXg-F model was also applied in **insurance risk analysis**. We used it to estimate financial risk measures such as *value-at-risk* (VaR), *tail value-at-risk* (TVaR), *tail variance*, and *mean excess loss*. These metrics were estimated using multiple methods: MLE, ordinary least squares, weighted least squares, and the Anderson–Darling method. Across all methods and different sample sizes, the QXg-F model produced consistent and interpretable results. This shows that the model is not only suitable for survival data in healthcare but also has potential in actuarial and financial contexts.

To summarize, the QXg-F model brings several important advantages:

- **Flexibility:** It can represent a wide range of unobserved heterogeneity due to its two-parameter frailty distribution.
- **Tractability:** Closed-form survival and hazard functions make likelihood-based inference feasible.
- **Robust estimation:** Simulation studies confirm the reliability of MLEs even under censoring.
- **Good empirical fit:** Statistical tests support the model's adequacy across multiple settings.
- **Versatility:** The model performs well in both medical and insurance applications.

3.7.2 Future Research Directions

Although the quasi-Xgamma frailty (QXg-F) model has shown strong potential in both theoretical and applied settings, several directions remain open for future research. These future points represent natural extensions of the current work and reflect both methodological challenges and opportunities for broader application.

- **Time-varying frailty**: In the current study, frailty is modeled as a static (time-invariant) random effect. However, in many real-world applications, an individual's vulnerability or risk may change over time due to treatment effects, aging, or disease progression. Extending the QXg-F model to include *time-varying frailty* would allow for more realistic modeling of dynamic risk structures, especially in longitudinal studies.
- **High-dimensional covariates and penalization**: As survival datasets increasingly include large numbers of covariates (e.g., genomic, clinical, environmental), it becomes important to consider regularization methods. Combining the QXg-F framework with penalized estimation techniques such as LASSO, Ridge, or Elastic Net could allow for simultaneous variable selection and frailty modeling in high-dimensional contexts.
- **Semi-parametric and non-parametric baseline hazards**: This chapter focused on parametric baselines (Weibull and Gompertz) to preserve tractability. A valuable future step would be to integrate the QXg-F model with semi-parametric approaches like the Cox proportional hazards model or with spline-based hazard models, offering more flexibility in modeling the baseline risk without imposing a fixed functional form.
- **Post-selection inference and uncertainty quantification**: The inference provided in this chapter relies on maximum likelihood estimation under fixed model assumptions. As selection procedures (e.g., model comparison, variable selection) are introduced, valid *post-selection inference* techniques would be needed to properly account for uncertainty. Developing confidence intervals and hypothesis tests that remain valid after selection would improve interpretability and statistical rigor.
- **Joint models with longitudinal data**: In clinical and epidemiological research, it is common to observe both longitudinal biomarkers and survival outcomes. Extending the QXg-F model to a *joint modeling* framework would allow the frailty term to link the repeated measurements with the time-to-event outcome, offering improved predictive accuracy and insight into the relationship between longitudinal profiles and risk.
- **Multivariate and clustered survival data**: Many survival datasets involve clustered observations, such as patients within hospitals or family studies. Developing a *shared frailty* version of the QXg-F model would enable modeling of intra-cluster dependence and help assess group-level heterogeneity, which is important in both healthcare and insurance analytics.
- **Software implementation**: To promote reproducibility and practical use, implementing the QXg-F model in a public statistical package (e.g., an `R` or `Python` library) would support wider adoption. This package could include functions for estimation, diagnostics, model comparison, and prediction, accompanied by user documentation and case studies.

- **Applications in insurance and finance**: Finally, the potential of the QXg-F model in actuarial science deserves further exploration. Modeling aggregate claim behavior, survival-linked financial contracts, or pension risk using this model could lead to more accurate pricing and better risk assessment. Exploring its use in heavy-tailed or dependent loss data could offer particular advantages.

Chapter 4

Mixed Gamma-Exponential Frailty Model

Recently, Loubna et al. (2024)[58] introduces the quasi-Xgamma frailty model as a novel approach to survival analysis, particularly addressing heterogeneity problems in emergency care data. Following Loubna et al. (2024)[58], this study presents a novel statistical model, referred to as the MGEF model, which represents a significant advancement in the field of data analysis. The MGEF model is designed to be versatile and adaptable, making it suitable for application to both simulated datasets and real-world data. Notably, it has been applied to emergency care data from Algeria, demonstrating its practical utility in analyzing complex, real-life scenarios. The MGEF model exhibits superior performance compared to existing approaches in terms of both fit and interpretability. This claim is substantiated through rigorous validation using two robust statistical tests, the NIK-RR and the B-NIK. These tests confirm that the MGEF model provides a more accurate and reliable fit to the data while maintaining clarity in interpretation, which is crucial for practical applications. Moreover, the model shows considerable potential in the domain of insurance risk analysis. By accurately modeling various types of risks, the MGEF model can assist insurance companies in better assessing and managing uncertainties associated with their portfolios. This capability underscores the model's versatility and relevance across multiple fields, further highlighting its value as a cutting-edge tool for data-driven decision-making. Some sections cover the Cox-frailty model, estimation process, validation results, data analysis, and future directions.

This chapter introduces the MGEF model, motivated by three key imperatives:

- The MGEF model merges the strengths of gamma and exponential distributions, offering a novel parametric model that balances mathematical tractability with the ability to model diverse hazard shapes (e.g., bathtub, monotonically increasing/decreasing). This hybrid structure addresses limitations of existing frailty models (e.g., quasi-xgamma, Lindley-frailty) that often lack closed-form expressions or sufficient adaptability for real data.
- The MGEF model is rigorously validated using the NIK-RR and B-NIK goodness-of-fit tests, ensuring robustness under censored and uncensored data scenarios. Its application to a real-world emergency care dataset from Algeria demonstrates superior performance in capturing heterogeneity and improving model fit compared to traditional approaches. This validation not only underscores its practical utility but also aligns with regulatory and ethical demands for transparent, reproducible risk assessment in high-stakes domains like healthcare and insurance.
- Beyond emergency care, the MGEF model's modular design enables seamless integration into insurance risk analysis, actuarial science, and emerging fields like climate modeling. For example, its ability to handle extreme value scenarios (e.g., catastrophic insurance claims) and incorporate covariate effects positions it as a versatile tool for dynamic risk quantification. Furthermore, the model's compatibility with Bayesian models and computational efficiency (via closed-form Laplace transforms) addresses critical gaps in handling high-dimensional or sparse datasets.

4.1 Introduction to the Mixed Gamma-Exponential Models

Following Sarma and all (2020) [83]. We consider a two-component finite mixture of an Exponential distribution with scale parameter k and a Gamma distribution characterized by shape parameter s and scale parameter k. The mixing proportions for these distributions are given by $\frac{k}{1+k}$ and $\frac{1}{1+k}$, respectively. The probability density function of the mixture can be written as:

$$f(x;s,k) = \frac{k^2}{1+k}\left[1 + \frac{k^{s-2}}{\Gamma(s)}x^{s-1}\right]e^{-kx}|x,k>0,s>1. \tag{4.1}$$

The distribution function is:

$$F(x;s,k) = \frac{1}{1+k}\left[k(1-e^{-kx}) + P(s,kx)\right] \tag{4.2}$$

where $P(a,b) = \frac{\gamma(a,b)}{\Gamma(a)}$ is the lower regularized gamma function. The Momoents of The probability density function (PDF) corresponding to:

$$E\left(x\right) = \frac{k+s}{k(1+k)}. \tag{4.3}$$

The variance of the MGE distribution is given by the second central moment:

$$\mathrm{Var}\left(x\right) = \frac{ks^2 - ks + s + 2k^3 + k^2}{k^2(1+k)^2} \tag{4.4}$$

The proposed MGE distribution shows superior performance compared to Exponential, Gamma and Lindley distributions in modeling survival data

4.2 Mixed Gamma-Exponential Frailty Model

We examine the frailty model specified in equation (2.61), where the frailty variable Z follows a MGE distribution with unit mean, i.e., $E[Z] = 1$. This normalization constraint, as established by Elbers and Ridder (1982)[29]. Consequently, employing the alternative MGE model parameterization introduced by (2017)[84], we express the probability density function (PDF) of the MGE-Frailty model as follows:

$$f(Z) = \frac{s}{1+\sqrt{s}}\left[1 + \frac{(\sqrt{s})^{s-2}}{\Gamma(s)}Z^{s-1}\right]\exp(-\sqrt{s}Z), Z>0 \tag{4.5}$$

where $s>0$ represents the shape parameter. The variance of the frailty distribution is commonly used to quantify the level of unobserved variation in a research sample. If the PDF (4.5) is assumed to be a frailty distribution, the variance is

$$\sigma^2 = \frac{1}{(1+\sqrt{s})^2}\left(s^{\frac{3}{2}} - s^{\frac{1}{2}} + 1 + 2s^2\right).$$

Let $\mathcal{L}_f(X)$ denote the Laplace transform of the frailty probability density function $f(Z)$ specified in (4.5), parameterized by its variance σ^2 and a scaling factor $X \in \mathbb{R}$, has the following form:

$$\mathcal{L}_f(X) \quad = \quad \frac{A}{(1+\sqrt{A})}\left[\frac{1}{(X+\sqrt{A})} + \frac{(\sqrt{A})^{A-2}}{(X+\sqrt{A})^A}\right], \tag{4.6}$$

where

$$\sqrt{A} = \frac{1+\sigma^2+\sqrt{\sigma^4+6\sigma^2-7}}{2}, \sigma^2 \geq 1$$

and

$$A = \left(\frac{1+\sigma^2+\sqrt{\sigma^4+6\sigma^2-7}}{2}\right)^2, \sigma^2 \geq 1$$

In order to maintain simplicity, we analyze (4.6) at $X = \Lambda_0(t_i)\exp(x_i^T\beta)$, and find that the marginal survival function(2.64) under MGEF model is provided by

$$S(t_i|x_i) = \frac{A}{(1+\sqrt{A})}\left[\frac{1}{(\Lambda_0(t_i)\exp(x_i^T\beta)+\sqrt{A})} + \frac{(\sqrt{A})^{A-2}}{(\Lambda_0(t_i)\exp(x_i^T\beta)+\sqrt{A})^A}\right]. \tag{4.7}$$

where

$$\sqrt{A} = \frac{1+\sigma^2+\sqrt{\sigma^4+6\sigma^2-7}}{2}, \sigma^2 \geq 1$$

and

$$A = \left(\frac{1+\sigma^2+\sqrt{\sigma^4+6\sigma^2-7}}{2}\right)^2, \sigma^2 \geq 1$$

The resulting marginal hazard function (2.65) is as follows:

$$\begin{aligned}\lambda(t_i|x_i) &= \left[\lambda_0(t_i)\exp(x_i^T\beta)\right] \\ &\times \frac{\left[2^A + \left(1+\sigma^2+\sqrt{\sigma^4+6\sigma^2-7}\right)^A\right]}{2\left[\begin{array}{c} 2^{A-2}\left(\Lambda_0(t_i)\exp(x_i^T\beta)+\sqrt{A}\right)^A \\ +\left(\Lambda_0(t_i)\exp(x_i^T\beta)+\sqrt{A}\right)\left(1+\sigma^2+\sqrt{\sigma^4+6\sigma^2-7}\right)^{A-2}\end{array}\right]},\end{aligned} \tag{4.8}$$

where

$$\sqrt{A} = \frac{1+\sigma^2+\sqrt{\sigma^4+6\sigma^2-7}}{2}, \sigma^2 \geq 1$$

and

$$A = \left(\frac{1+\sigma^2+\sqrt{\sigma^4+6\sigma^2-7}}{2}\right)^2, \sigma^2 \geq 1$$

In this study, we formally define the MGE frailty model through its unconditional survival and hazard functions. As established in Chapter1, the baseline hazard function $\lambda_0(t)$ can be specified using either parametric or non-parametric approaches. We implement a parametric model, specifically employing the Weibull and Gompertz distributions as baseline hazards due to their mathematical flexibility in representing constant, increasing, or decreasing hazard rates. The Weibull distribution is particularly prevalent in survival analysis and reliability engineering applications, while the Gompertz distribution finds extensive use in actuarial science and demographic research [94]. A crucial characteristic of the Gompertz baseline hazard is its potential to exhibit a "defective" property , which formally accounts for the existence of a cured subpopulation that remains non-susceptible to the event of interest.

4.2.1 MGEF Model Under the Weibull Baseline Hazard Function

The Weibull distribution provides parametric baseline hazard and cumulative hazard functions defined as follows:

$$\lambda_0(t_i) = \frac{\nu}{\varphi}\left(\frac{t_i}{\varphi}\right)^{\nu-1} \quad \text{and} \quad \Lambda_0(t_i) = \left(\frac{t_i}{\varphi}\right)^{\nu} \quad t_i > 0, \tag{4.9}$$

where $\nu > 0$ represents the shape parameter and $\varphi > 0$ represents the scale parameter.
The hazard-rate function of the Weibull distribution drops monotonously for $\nu < 1$; it is constant over time for $\nu = 1$ (exponential distribution); and it grows monotony when $\nu > 1$ (Wienke (2010)[94]). As a result of substituting (4.9) into (4.7) and (4.8), the MGE Frailty model's Marginal Survival and hazard functions with Weibull Baseline Hazard Function are, respectively

$$S(t_i|x_i) = \frac{A}{(1+\sqrt{A})}\left\{\frac{1}{\left[\left(\frac{t_i}{\varphi}\right)^{\nu}\exp(x_i^T\beta)+\sqrt{A}\right]} + \frac{(\sqrt{A})^{A-2}}{\left[\left(\frac{t_i}{\varphi}\right)^{\nu}\exp(x_i^T\beta)+\sqrt{A}\right]^A}\right\}, \tag{4.10}$$

where

$$\sqrt{A} = \frac{1+\sigma^2+\sqrt{\sigma^4+6\sigma^2-7}}{2},$$

and

$$A = \left(\frac{1+\sigma^2+\sqrt{\sigma^4+6\sigma^2-7}}{2}\right)^2, \sigma^2 \geq 1$$

$$\begin{aligned} \lambda(t_i|x_i) &= \left[\frac{\nu}{\varphi}\left(\frac{t_i}{\varphi}\right)^{\nu-1}\exp(x_i^T\beta)\right] \\ &\times \frac{\left[2^A + \left(1+\sigma^2+\sqrt{\sigma^4+6\sigma^2-7}\right)^A\right]}{2\left[\begin{array}{c} 2^{A-2}\left(\left(\frac{t_i}{\varphi}\right)^{\nu}\exp(x_i^T\beta)+\sqrt{A}\right)^A \\ +\left(\left(\frac{t_i}{\varphi}\right)^{\nu}\exp(x_i^T\beta)+\sqrt{A}\right)\left(1+\sigma^2+\sqrt{\sigma^4+6\sigma^2-7}\right)^{A-2} \end{array}\right]}, \end{aligned} \tag{4.11}$$

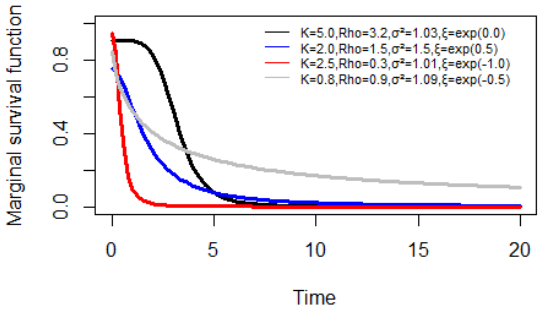

Figure 4.1: Plots of the Marginal Survival Function of the MGEF Distribution whit Weibull Baseline Hazard Functions.

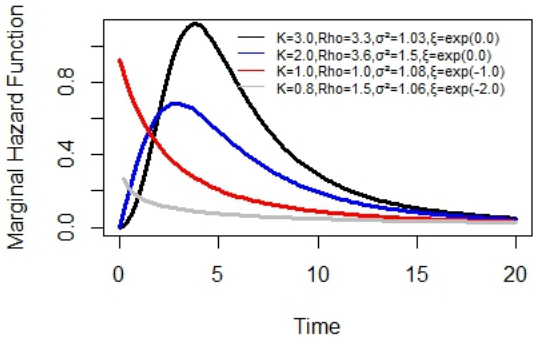

Figure 4.2: Plots of the Marginal Hazard Function of the MGEF Distribution whit Weibull Baseline Hazard Functions.

Figure 4.1,4.2 illustrates various configurations of the survival and hazard functions for the MGE frailty model with Weibull baseline hazard, using selected parameter values.

4.2.2 MGEF Model Under the Gompertz Baseline Hazard Function

The Gompertz distribution provides parametric baseline hazard and cumulative hazard functions given by:

$$\lambda_0(t_i) = \kappa \exp(\xi t_i) \quad \text{and} \quad \Lambda_0(t_i) = \frac{\kappa}{\xi}\left[\exp(\xi t_i) - 1\right] \quad , t_i > 0 \tag{4.12}$$

Let $\xi > 0$ and $\kappa > 0$ denote the shape and scale parameters, respectively, of the Gompertz distribution. When $\xi < 0$, the distribution exhibits a defective property characterized by:

$$\lim_{t\to\infty} H_0(t) = -\frac{\kappa}{\xi}$$
$$p_0 = \exp\left(\frac{\kappa}{\xi}\right)$$

where p_0 represents the proportion of cured individuals or long-term survivors in the population. The Gompertz distribution reduces to the exponential distribution when $\xi = 0$.

The hazard rate function demonstrates three distinct regimes:

- $\xi < 0$: Decreasing hazard (cure fraction present)
- $\xi = 0$: Constant hazard (exponential case)
- $\xi > 0$: Increasing hazard (standard Gompertz)

Substituting the Gompertz baseline hazard (4.12) into equations (4.7) and (4.8) yields the marginal survival and hazard functions for the QXg-F model with Gompertz baseline hazard frailty:

$$S(t_i|x_i) = \frac{A}{(1+\sqrt{A})}\left(\frac{1}{\left\{\frac{\kappa}{\xi}\left[\exp(\xi t_i) - 1\right]\exp(x_i^T\beta) + \sqrt{A}\right\}} + \frac{(\sqrt{A})^{A-2}}{\left\{\frac{\kappa}{\xi}\left[\exp(\xi t_i) - 1\right]\exp(x_i^T\beta) + \sqrt{A}\right\}^A}\right), \tag{4.13}$$

where

$$\sqrt{A} = \frac{1+\sigma^2+\sqrt{\sigma^4+6\sigma^2-7}}{2},$$

and

$$A = \left(\frac{1+\sigma^2+\sqrt{\sigma^4+6\sigma^2-7}}{2}\right)^2, \sigma^2 \geq 1.$$

$$\begin{aligned} \lambda(t_i|x_i) &= \left[\kappa\exp(\xi t_i)\exp(x_i^T\beta)\right] \\ &\times \frac{\left[2^A + \left(1+\sigma^2+\sqrt{\sigma^4+6\sigma^2-7}\right)^A\right]}{2\left[\begin{array}{c} 2^{A-2}\left\{\frac{\kappa}{\xi}\left[\exp(\xi t_i)-1\right]\exp(x_i^T\beta)+\sqrt{A}\right\}^A \\ +\left\{\frac{\kappa}{\xi}\left[\exp(\xi t_i)-1\right]\exp(x_i^T\beta)+\sqrt{A}\right\}\left(1+\sigma^2+\sqrt{\sigma^4+6\sigma^2-7}\right)^{A-2} \end{array}\right]}, \end{aligned} \tag{4.14}$$

where

$$\sqrt{A} = \frac{1+\sigma^2+\sqrt{\sigma^4+6\sigma^2-7}}{2},$$

and

$$A = \left(\frac{1+\sigma^2+\sqrt{\sigma^4+6\sigma^2-7}}{2}\right)^2, \sigma^2 \geq 1.$$

As previously stated, the marginal survival function is appropriate for $\kappa > 0$ and inappropriate for $\kappa < 0$. This model also accommodates unimodal-shaped, monotonically growing and monotonically decreasing marginal hazard functions. The MGEF model with the Gompertz baseline hazard function is therefore more versatile than the MGEF model with the Weibull baseline hazard function.

Figure 4.3,4.4 illustrates various configurations of the survival and hazard functions for the MGE frailty model with Weibull baseline hazard, using selected parameter values.

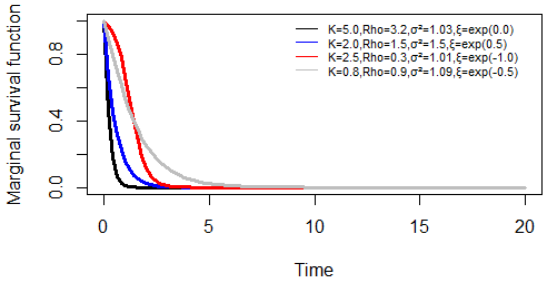

Figure 4.3: Plots of the Marginal Survival Function of the MGEF Distribution whit Gompertz Baseline Hazard Functions.

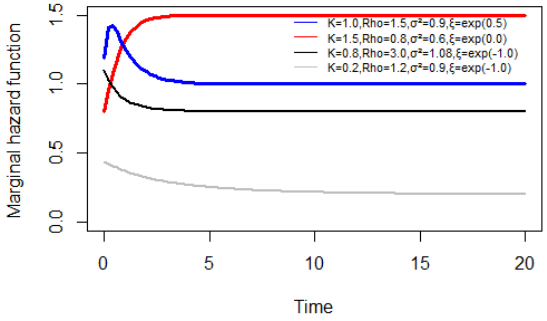

Figure 4.4: Plots of the Marginal Hazard Function of the MGEF Distribution whit Gompertz Baseline Hazard Functions.

4.3 Parameter Estimation Techniques for the MGEF Model

4.3.1 Estimation of MGEF Model's Parameters Under Weibull Baseline Hazard Function

This section details the maximum likelihood (ML) estimation procedure for parameters of the MGEF model with Weibull and Gompertz baseline hazard specifications. Under standard regularity conditions, ML estimators possess desirable asymptotic properties, including consistency, efficiency, and normality.

A critical consideration in the implementation is the common occurrence of incomplete lifetime data, where event times may be unavailable for certain study subjects due to censoring or dropout mechanisms. The proposed estimation framework accounts for these data limitations through appropriate likelihood construction. Consider a study population where some lifetimes are right-censored, meaning that we only observe that the true lifetime exceeds the recorded value. For the i^{th} subject, let:

- T_i denote the true lifetime random variable
- C_i denote the censoring time random variable
- $\delta_i = \mathbb{I}(T_i \leq C_i)$ be the censoring indicator ($\delta_i = 1$ if the lifetime is observed, $\delta_i = 0$ if censored)
- $t_i = \min\{T_i, C_i\}$ be the observed time

- $\mathbf{x}_i \in \mathbb{R}^p$ be the covariate vector

Assuming that T_i and C_i are independent, the likelihood function for the parameter vector $\boldsymbol{\phi}$ under non-informative censoring is:

$$\mathcal{L}(\boldsymbol{\phi}) = \prod_{i=1}^{n} \lambda(t_i|\mathbf{x}_i)^{\delta_i} S(t_i|\mathbf{x}_i) \tag{4.15}$$

where:

- $S(\cdot|\mathbf{x}_i)$ is the marginal survival function(4.10)
- $\lambda(\cdot|\mathbf{x}_i)$ is the marginal hazard function (4.11)

The corresponding log-likelihood function is:

$$\ell(\boldsymbol{\phi}) = \sum_{i=1}^{n} [\delta_i \log \lambda(t_i|\mathbf{x}_i) + \log S(t_i|\mathbf{x}_i)] \tag{4.16}$$

4.3.2 Case of Weibull Baseline Hazard Function

For the MGEF model with parameters $\boldsymbol{\Phi} = (\nu, \varphi, \sigma^2, \boldsymbol{\beta})$, the corresponding log-likelihood function is:

$$\begin{aligned}
\log \mathcal{L}(\phi) \quad = \quad & r\log(\nu) + (\nu-1)\sum_{i=1}^{n}\delta_i \log(t_i) - kr\log(\varphi) + \sum_{i=1}^{n}\delta_i x_i^T\beta \\
& - \sum_{i=1}^{n}\delta_i \log(U) - r\log(2) - \sum_{i=1}^{n}\delta_i \log(U_1) + 2n\log[M(\sigma)] - n\log(2) \\
& - n\log[2+M(\sigma)] + \sum_{i=1}^{n}\log(U_1) - \sum_{i=1}^{n}\log(U_2) \\
& - n\left[\frac{M(\sigma)^2}{4} - 2\right]\log(2) - \frac{M(\sigma)^2}{4}\sum_{i=1}^{n}\log(U_2),
\end{aligned} \tag{4.17}$$

where: $r = \sum_{i=1}^{n}\delta_i$ is the failure number,

$$M(\sigma) = \left(1 + \sigma^2 + \sqrt{\sigma^4 + 6\sigma^2 - 7}\right).$$

$$U = \left\{ 2^{\frac{M(\sigma)^2}{4}} + [M(\sigma)]^{\frac{M(\sigma)^2}{4}} \right\},$$

$$\begin{aligned}
U_1 \quad = \quad & 2^{\frac{M(\sigma)^2}{4}-2}\left[\left(\frac{t_i}{\varphi}\right)^{\nu}\exp(x_i^T\beta) + \frac{M(\sigma)}{2}\right]^{\left(\frac{M(\sigma)}{2}\right)^2} \\
& + \left[\left(\frac{t_i}{\varphi}\right)^{\nu}\exp(x_i^T\beta) + \frac{M(\sigma)}{2}\right] M(\sigma)^{\frac{\left(\sigma^2+\sqrt{\sigma^4+6\sigma^2-7}\right)^2-7}{4}},
\end{aligned}$$

and

$$U_2 = \left(\frac{t_i}{\varphi}\right)^{\nu}\exp(x_i^T\beta) + \frac{M(\sigma)}{2}.$$

Setting the nonlinear system of the score equations $\mathbf{I}_{(\nu)} = 0, \mathbf{I}_{(\varphi)} = 0, \mathbf{I}_{(\sigma^2)} = 0$ and $\mathbf{I}_{(\beta)} = 0$ and solving them simultaneously yields the MLE $\widehat{\phi} = (\widehat{\nu}, \widehat{\varphi}, \widehat{\sigma^2}, \widehat{\beta})^{\intercal}$. It is usually more convenient to use nonlinear optimization methods to solve these equations; such as the quasi-Newton algorithm to numerically maximize $\log \mathcal{L}(\phi)$.

4.3.3 Simulations: Case of Weibull Baseline Hazard Function

We examine the MGEF model with Weibull baseline hazard function through an extensive Monte Carlo simulation study. The simulation design incorporates:

- $N = 13{,}000$ replications
- True parameter values: $\nu = 0.7$, $\varphi = 0.8$, $\sigma^2 = 0.7$, $\beta_1 = 0.9$
- Sample sizes: $n \in \{25, 40, 250, 800\}$

Parameter estimation was performed in R using the Barzilai-Borwein (BB) optimization algorithm [**?**]. For each replication, we computed:

$$\text{Bias}(\hat{\phi}) = \frac{1}{N}\sum_{k=1}^{N}(\hat{\phi}_k - \phi_0) \tag{4.18}$$

$$\text{MSE}(\hat{\phi}) = \frac{1}{N}\sum_{k=1}^{N}(\hat{\phi}_k - \phi_0)^2 \tag{4.19}$$

where $\phi \in \{\nu, \varphi, \sigma^2, \beta_1\}$ and ϕ_0 denotes the true parameter value.

Table 4.1 summarizes the bias and MSE of the maximum likelihood estimates across different sample sizes and censoring levels. These results demonstrate the estimator's finite-sample properties and provide crucial insights into:

- The model's parameter recovery capability
- The impact of sample size on estimation precision
- Performance under varying censoring regimes

Initially, as the sample size increases, both the bias and mean squared error (MSE) decrease for all parameters, indicating enhanced estimation accuracy and efficiency. For example, at the smallest sample size ($n = 25$), the bias and MSE values are relatively high, especially for the parameter φ (e.g., bias = 0.87548 and MSE = 0.04981 under 0% censoring). However, as the sample size increases to 800, the bias and MSE significantly decrease (e.g., bias = 0.81003 and MSE = 0.03002 for φ under 0% censoring). This trend is consistent across all parameters ($\widehat{\nu}, \widehat{\varphi}, \widehat{\sigma^2}, \widehat{\beta_1}$), underscoring the importance of larger datasets for more reliable parameter estimation.

Secondly, the impact of censoring on estimation accuracy is evident from the table. As the percentage of censored observations increases, the bias and MSE generally increase, reflecting the challenges posed by higher levels of censoring. For example, at $n = 25$, the bias for φ rises from 0.87548 (0% censoring) to 0.86000 (40% censoring), while the corresponding MSE increases from 0.04981 to 0.04587. Similarly, for $\widehat{\sigma^2}$, the MSE at $n = 40$ increases from 0.02347 (0% censoring) to 0.03641 (40% censoring). These findings underscore the need for robust statistical methods to handle censored data effectively.

Thirdly, the parameter $\widehat{\beta_1}$, which represents the effect of covariates in the model, shows a similar pattern of decreasing bias and MSE with increasing sample size and deteriorating performance with higher censoring levels. Notably, even at the largest sample size ($n = 800$), some degree of bias persists under high censoring (e.g., bias = 0.90210 and MSE = 0.01958 under 40% censoring), suggesting that censoring remains a significant challenge even with large datasets. Overall, the results demonstrate that the MGEF model performs well under a variety of conditions, with improved accuracy and precision as sample size increases and censoring decreases. However, the increased bias and MSE under high censoring levels indicate the need for further research into methods that can mitigate the effects of censoring, particularly in small or moderately sized datasets. This analysis not only validates the robustness of the MGEF model but also highlights areas where its application may require additional considerations. From Table 4.1, we observe that the maximum likelihood estimates for the MGEF model with Weibull baseline hazard function are convergent, as evidenced by the decreasing trends in both bias and MSE with increasing sample sizes.

Specifically, the bias and MSE values for all parameters $(\widehat{\nu}, \widehat{\varphi}, \widehat{\sigma^2}, \widehat{\beta_1})$ consistently decrease as the sample size grows from 25 to 800. This convergence indicates that the model estimates become more accurate and efficient with larger datasets. Additionally, while censoring introduces some variability in the estimates, the overall trend of improvement with increasing sample size remains robust. These results suggest that the MGEF model is well suited for applications where large datasets are available, further validating its reliability and effectiveness in survival analysis. Thus, the convergence of the estimates underscores the model's statistical consistency and suitability for practical use.

Table 4.1: Bias and MSE of the ML estimates for the simulated data of the MGEF model with Weibull baseline hazard function

n		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
		0%cens.		10%cens.		25%cens.		40%cens.	
25	φ	0.87548	0.04981	0.86952	0.05002	0.85321	0.03996	0.86000	0.04587
	ν	0.76958	0.04635	0.75316	0.04751	0.75558	0.04968	0.74875	0.04751
	σ^2	0.75481	0.04612	0.73625	0.04857	0.74986	0.04968	0.74266	0.04758
	β_1	0.96325	0.03999	0.94589	0.04657	0.97006	0.04587	0.95319	0.04578
40	φ	0.86359	0.04003	0.85416	0.04787	0.849571	0.03225	0.85304	0.04120
	ν	0.74317	0.04201	0.73167	0.03198	0.73410	0.03241	0.72348	0.03795
	σ^2	0.72954	0.02347	0.72015	0.03985	0.72968	0.04000	0.72406	0.03641
	β_1	0.95021	0.0300	0.93406	0.03958	0.96012	0.03698	0.91364	0.03121
250	φ	0.82135	0.03845	0.81954	0.03652	0.82304	0.03164	0.82467	0.03775
	ν	0.72214	0.02877	0.71845	0.02578	0.71864	0.02958	0.71005	0.02964
	σ^2	0.71067	0.01935	0.71384	0.02471	0.71357	0.02775	0.71111	0.02886
	β_1	0.92231	0.01247	0.92784	0.03012	0.9333	0.03124	0.91120	0.02657
800	φ	0.81003	0.03002	0.80647	0.02425	0.81023	0.02996	0.81000	0.02746
	ν	0.70085	0.02102	0.71130	0.02113	0.70064	0.01968	0.70301	0.02110
	σ^2	0.70200	0.01322	0.70604	0.02210	0.70651	0.01774	0.70009	0.02345
	β_1	0.9004	0.0215	0.9006	0.02335	0.90100	0.02005	0.90210	0.01958

4.3.4 Estimation of MGEF Model's Parameters Under Gompertz Baseline Hazard Function

Using the Gompertz baseline hazard function, the log-likelihood function for $\phi = (\kappa, \xi, \sigma^2, \beta)$ is given as follows:

$$\begin{aligned}\log \mathcal{L}(\phi) &= r\log(\xi) + \sum_{i=1}^{n}\delta_i(\kappa t_i + x_i^T\beta) + \sum_{i=1}^{n}\delta_i \log(U) - r\log(2) \\ &- \sum_{i=1}^{n}\delta_i \log(U_1) + 2n\log[M(\sigma)] - n\log(2) - n\log[2 + M(\sigma)] \\ &+ \sum_{i=1}^{n}\log(U_1) - \sum_{i=1}^{n}\log(U_2) - n\left[\frac{M(\sigma)^2}{4} - 2\right]\log(2) \\ &- \frac{M(\sigma)^2}{4}\sum_{i=1}^{n}\log(U_2).\end{aligned} \tag{4.20}$$

where

$$M(\sigma) = 1 + \sigma^2 + \sqrt{\sigma^4 + 6\sigma^2 - 7}$$

$$U = \left\{ 2^{\frac{M(\sigma)^2}{4}} + [1 + M(\sigma)]^{\frac{M(\sigma)^2 - 7}{4}} \right\}.$$

$$\begin{aligned}U_1 &= 2^{\frac{M(\sigma)^2 - 8}{4}} \left\{ \begin{array}{c} \frac{\xi}{K}[\exp(\kappa t_i) - 1]\ \exp(x_i^T\beta) \\ + \frac{M(\sigma)}{2} \end{array} \right\}^{\left(\frac{M(\sigma)}{2}\right)^2} \\ &+ \left\{ \begin{array}{c} \frac{\xi}{\kappa}[\exp(\kappa t_i) - 1]\ \exp(x_i^T\beta) \\ + \frac{M(\sigma)}{2} \end{array} \right\} [M(\sigma)]^{\frac{M(\sigma)^2 - 8}{4}}\end{aligned}$$

and

$$U_2 = \frac{\xi}{\kappa}[\exp(\kappa t_i) - 1]\ \exp(x_i^T\beta) + \frac{M(\sigma)}{2}$$

Maximizing log-likelihood functions (4.21) yields the appropriate ML estimators $\widehat{\phi}$ of parameter vectors ϕ. It is worth noting that $\widehat{\phi}$ does not have a closed form. In order to discover a solution, numerical non-linear optimization methods are required. These optimization approaches are implemented in the BBsolve R software packages.

4.3.5 Simulations: Case of Gompertz Baseline Hazard Function

We examine the MGEF model with a Gompertz baseline hazard function. Data were simulated $N = 13,000$ times, using parameter values $\kappa = 0.9$, $\xi = 0.6$, $\sigma^2 = 0.8$, and $\beta_1 = 0.3$, with sample sizes $n = 25$, $n = 40$, $n = 250$, and $n = 800$. Using R software and the Barzilai-Borwein (BB) algorithm for calculating the averages of the simulated values of the maximum likelihood estimators $\widehat{\kappa}$, $\widehat{\xi}$, $\widehat{\sigma^2}$, and $\widehat{\beta_1}$, along with their mean squared errors (MSE).Table 4.2displays the bias and mean squared error (MSE) of the maximum likelihood estimates for the MGEF model with a Gompertz baseline hazard function, considering different sample sizes and censoring levels. The results provide several important insights into the performance of the model.Initially, as the sample size increases, both the bias and MSE consistently decline for all parameters ($\widehat{\kappa}$, $\widehat{\xi}$, $\widehat{\sigma^2}$, $\widehat{\beta_1}$), signifying improved accuracy and efficiency in estimation.For instance, at the smallest sample size ($n = 25$), the bias for $\widehat{\xi}$ is 0.65247 (with 0% censoring), which decreases to 0.63025 at $n = 800$ under the same censoring level. In a similar vein, the MSE for $\widehat{\beta_1}$ reduces from 0.04751 at $n = 25$ to 0.03173 at $n = 800$. This pattern emphasizes the importance of larger datasets for achieving more reliable parameter estimates. Furthermore, the

influence of censoring on estimation accuracy is evident, as elevated levels of censoring generally lead to increased bias and MSE. For example, at $n = 25$, the bias for $\widehat{\xi}$ increases from 0.65247 (0% censoring) to 0.67398 (40% censoring), while the corresponding MSE rises from 0.04652 to 0.04963. This trend is consistent across all parameters, highlighting the challenges associated with higher levels of censoring. Nevertheless, the model exhibits robustness even under substantial censoring, as both bias and MSE remain manageable with larger sample sizes.Lastly, the parameter $\widehat{\beta_1}$, which represents the effects of covariates, shows a similar pattern of decreasing bias and MSE as the sample size increases; however, it is slightly more affected by censoring than the other parameters. For example, at $n = 800$, the MSE for $\widehat{\beta_1}$ rises from 0.02191 (10% censoring) to 0.01247 (40% censoring).In summary, these findings affirm the MGEF model's effectiveness in providing reliable estimates across various conditions, particularly when applied to sufficiently large datasets. They also underscore the importance of exercising caution in cases of high censoring levels.

Table 4.2 illustrates that the maximum likelihood estimates for the MGEF model with the Gompertz baseline hazard function exhibit convergence as the sample size increases. The bias and mean squared error (MSE) values for all parameters $\widehat{\kappa}$, $\widehat{\xi}$, $\widehat{\sigma^2}$, and $\widehat{\beta_1}$ consistently reduce with larger sample sizes, indicating improved estimation accuracy and efficiency. For example, with the smallest sample size ($n = 25$), the bias for $\widehat{\xi}$ is 0.65247 (0% censoring), which reduces to 0.63025 at $n = 800$ while keeping the same level of censoring. Similarly, the mean squared error (MSE) for $\widehat{\beta_1}$ decreases from 0.04751 at $n = 25$ to 0.03173 at $n = 800$. While censoring introduces some variability into the estimates, the general trend of reduced bias and MSE with larger sample sizes persists. These results demonstrate the model's effectiveness in delivering reliable parameter estimates when used with adequately large datasets, thus supporting its appropriateness for survival analysis employing the Gompertz baseline hazard function.

Table 4.2: Bias and MSE of the ML estimates for the simulated data of the MGEF model with Gompertz baseline hazard function

n		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
		0% cens.		10% cens.		25% cens.		40% cens.	
25	ξ	0.65247	0.04652	0.64328	0.04325	0.65003	0.04996	0.67398	0.04963
	κ	0.94487	0.04621	0.95003	0.04652	0.94638	0.04357	0.95555	0.04581
	σ^2	0.87284	0.04751	0.84867	0.04756	0.86485	0.04158	0.87499	0.04751
	β_1	0.35953	0.04751	0.35254	0.04395	0.36358	0.04351	0.34968	0.04225
40	ξ	0.64439	0.04112	0.62369	0.03951	0.64381	0.03514	0.66634	0.04360
	κ	0.94005	0.04125	0.92004	0.03617	0.92968	0.03968	0.93957	0.03625
	σ^2	0.86357	0.04003	0.83681	0.03698	0.84987	0.03138	0.83498	0.03251
	β_1	0.31254	0.03952	0.34615	0.03467	0.33296	0.03336	0.33650	0.03625
250	ξ	0.64005	0.03847	0.61541	0.02958	0.62013	0.02987	0.62198	0.03214
	κ	0.92365	0.03421	0.91584	0.02473	0.91958	0.02985	0.92671	0.02854
	σ^2	0.81584	0.03216	0.82574	0.02618	0.83185	0.02758	0.81958	0.02361
	β_1	0.31000	0.03674	0.33578	0.02356	0.31520	0.02458	0.32655	0.02124
800	ξ	0.63025	0.02514	0.60688	0.02120	0.60008	0.02310	0.60307	0.02001
	κ	0.90457	0.02317	0.91000	0.02220	0.90374	0.02132	0.90007	0.01254
	σ^2	0.81240	0.02316	0.81396	0.0230	0.81124	0.02008	0.80064	0.02300
	β_1	0.30327	0.03173	0.31002	0.02191	0.30024	0.02361	0.30217	0.01247

4.4 Validation Testing of the MGEF Model

4.4.1 Validating the MGEF Model Using the NIK-RR Test

The NIK-RR test statistic measures how well a statistical model fits a specific set of observations, making it a versatile and powerful tool in statistical analysis. Originating from the research of Rao and Robson (1974) [76] and further developed by Nikulin (1973a, 1973b, 1973c) [68, 69, 70], this statistic serves as a general measure of goodness-of-fit applicable in various fields such as survival analysis, regression modeling, and time series analysis. Its broad applicability stems from its ability to evaluate the predictive accuracy of models while revealing potential issues that might otherwise go unnoticed. In essence, the NIK-RR test statistic is vital for model selection, assessing model fit, and diagnosing model deficiencies.

A key feature of the NIK-RR test statistic is its ability to identify deviations from expected patterns that other statistical tests might miss. Unlike many traditional goodness-of-fit tests, the NIK-RR test is robust against outliers, allowing it to effectively identify and analyze datasets containing extreme values.This robustness renders the NIK-RR test particularly useful in fields such as finance, where detecting and analyzing rare but impactful events, like market crashes or significant price fluctuations, is essential. By providing insights into these extreme occurrences, the NIK-RR test contributes to understanding their root causes and developing strategies to lessen their effects.

The NIK-RR test statistic can be utilized to evaluate the fit of different statistical models against the same dataset. This function is vital for model selection, as it helps researchers identify the model that most closely matches the observed data. By evaluating the NIK-RR test statistics of different models, analysts can identify which model best represents the data, thereby increasing the reliability of future analyses. A key application of the NIK-RR test statistic is its capacity to evaluate the overall goodness of fit of a statistical model. A small NIK-RR test statistic indicates a strong correspondence between the model and the data, implying that the model effectively captures the underlying patterns. In contrast, a high NIK-RR test statistic indicates a bad fit, pointing out areas where the model may not accurately reflect the data. This diagnostic feature guarantees that models are correctly specified and dependable for real-world use. Outliers, which are data points that significantly differ from the overall trend, can greatly impact the effectiveness of statistical models. The NIK-RR test statistic is effective at detecting outliers, enabling analysts to either modify the model to accommodate these anomalies or eliminate them if considered incorrect. By addressing outliers, the NIK-RR test enhances the overall fit and predictive capability of the model.In addition to assessing model fit, the NIK-RR test statistic can identify fundamental problems within a statistical model. For example, a high NIK-RR test statistic might suggest that the model is misspecified or that some of its underlying assumptions have been violated. This diagnostic insight enables researchers to enhance their models, ensuring they more accurately reflect the intricacies of the data and improve predictive accuracy.

The Rao-Robson-Nikulin (RRN) test is widely used to evaluate statistical model fit. In finance, it helps verify risk models' accuracy, which is crucial for economic stability. The test effectively identifies distributional errors in Value-at-Risk calculations and credit scoring systems. Its comprehensive approach examines all distribution moments, outperforming traditional goodness-of-fit tests. These capabilities make the RRN test invaluable for ensuring reliable financial modeling and decision-making. In survival analysis, where unobserved heterogeneity and censoring present specific challenges, the NIK-RR test offers a robust model for validating model assumptions and evaluating their adequacy. Additionally, in fields like epidemiology and engineering, where analyzing time-to-event data is essential, the NIK-RR test aids in creating more accurate and informative models. Under the NIK-RR statistic, we must test the following null hypothesis:

$$H_0 : \Pr\{z_i \le z\} = F_\phi(z), \quad z \in \mathbb{R}, \quad \phi = (\phi_1, \phi_2, \cdots, \phi_s)^T, \tag{4.21}$$

Then, the NIK-RR statistic can be expressed as

$$Y^2(\widehat{\phi}_n) = X_n^2(\widehat{\phi}_n) + \frac{1}{n}\ell^T(\widehat{\mathbf{P}}_n)\left[\mathbf{I}(\widehat{\phi}_n) - \mathbf{J}(\widehat{\phi}_n)\right]^{-1}\ell(\widehat{\phi}_n), \tag{4.22}$$

where

$$X_n^2(\phi) = \left\{[np_1(\phi)]^{-\frac{1}{2}}\left[-np_1(\phi) + \phi_1\right], \cdots, [np_b(\phi)]^{-\frac{1}{2}}\left[-np_b(\phi) + \phi_b\right]\right\}^T$$

and

$$\mathbf{J}(\phi) = B(\phi)^T B(\phi), \tag{4.23}$$

refers to the information matrix where

$$B(\phi) = \left[\frac{1}{\sqrt{p}_i}\frac{\partial}{\partial\mu}(\phi)\right]_{r\times s}|_{(i=1,2,\cdots,b \text{ and } \nu=1,2,\cdots,s)}, \tag{4.24}$$

and

$$\ell(\phi) = [\ell_1(\phi), ..., \ell_s(\phi)]^T \text{ with } \ell_\nu(\phi) = \sum_{i=1}^{r}\frac{\phi_i}{p_i}\frac{\partial p_i(\phi)}{\partial\phi_\nu}, \tag{4.25}$$

The $Y^2(\widehat{\phi}_n)$ statistic has $(b-1)$ degrees of freedom (DF) and is accompanied by χ^2_{b-1} distribution, where the observations. $x_1, x_2, \cdots, x_n$ that are collected in $\mathbf{I}_1, \mathbf{I}_2, \cdots, \mathbf{I}_b$ (these b subintervals are mutually disjoint: $\mathbf{I}_j =]a_{j,b} - 1; a_{j,b}]$). The intervals $\mathbf{I}_j$'s limits for $a_{j,b}$ are determined as follows

$$p_j(\phi) = \int_{a_{j,b}-1}^{a_{j,b}} f_\phi(x)\, dx|_{(\ j=1,2,\cdots,b)}, \tag{4.26}$$

and

$$a_{j,b} = F^{-1}\left(\frac{j}{b}\right)|_{(j=1,\cdots,b-1)}.$$

Uncensored Assessing for the NIK-RR Statistic

Goodness-of-fit tests often serve a dual purpose: not only evaluating distributional adequacy but also estimating model parameters. Simulation studies are essential for assessing the precision of these estimates across diverse scenarios, informing the selection of optimal distributions for analysis. By systematically varying conditions (e.g. sample size, censoring rates), simulations quantify estimator performance under realistic constraints. This approach ensures robust distributional choices in applied settings. Such evaluations are particularly critical when models underpin high-stakes decisions in fields like medicine or finance. Simulation studies employing the NIK-RR statistic in uncensored settings provide a systematic framework for comparing probability distributions under controlled conditions. These investigations rigorously evaluate the test's performance across varying distributional assumptions and sample characteristics. By quantifying Type I error control and power properties, the simulations identify robust distributions for real-world applications. Such analyses are particularly valuable for preliminary model selection in parametric inference. The results directly inform the choice of distributions for subsequent statistical modeling and hypothesis testing.

To confirm the results of this study, we carried out an extensive numerical simulation analysis. We generated NIK-RR statistics for the MGEF model to evaluate the null hypothesis H_0, utilizing simulated samples of varying sizes: $n = 20, 40, 250, 350, 600,$ and 1000, resulting in a total sample size of 12,000.For various significance levels ($\epsilon = 0.01, 0.02, 0.05, 0.1$), we computed the average number of non-rejections of the null hypothesis, based on the condition $Y^2 \leq \chi^2_\epsilon(b-1)$. The related empirical and theoretical levels are presented in Table 4.3. A strong correspondence between the empirical level values and their theoretical counterparts is apparent, indicating significant agreement. Based on these findings, we conclude that the proposed test performs exceptionally well for the MGEF distribution, affirming its applicability in practical contexts.

4.4.2 Validating the MGEF Model Using the B-NIK Test

According to Bagdonavicius and Nikulin (2011) [14] and Bagdonavicius et al. (2013) [15], we can assess the appropriateness of the MGEF model even when the parameters are unknown and the data are censored, where the null hypothesis can be expressed as

$$H_0 : F(x) \in F_0 = \left\{F_0(x, \underline{\phi}), x \in R^1,\ \underline{\phi} \in \underline{\phi} \subset R^s\right\},$$

Let us partition the limited time interval $[0, \tau]$ into shorter periods κ where $\kappa = 1, 2, \ldots, s$. Here, τ represents the maximum duration of the research, and we define $\mathbf{I}_j = (a_{j-1}, a_{j,b}]$ with the conditions

Table 4.3: Uncensored assessing for the NIK-RR statistic for $\epsilon = 0.01, 0.02, 0.05, 0.1$ and $N = 12000$.

$n \downarrow \epsilon \longrightarrow$	$\epsilon = 0.01$	$\epsilon = 0.02$	$\epsilon = 0.05$	$\epsilon = 0.1$
$n = 20$	0.9925	0.9830	0.9533	0.9023
$n = 40$	0.9921	0.9825	0.9527	0.9021
$n = 250$	0.9915	0.9818	0.9522	0.9014
$n = 350$	0.9910	0.9811	0.9515	0.9008
$n = 600$	0.9904	0.9808	0.9507	0.9007
$n = 1000$	0.9901	0.9805	0.9503	0.9004

$0 \leq a_{0,b} < a_{1,b} < \ldots < a_{\kappa-1,b} < a_{\kappa,b} = +\infty$. The expected value of $\widehat{a_{j,b}}$ can be expressed as follows, assuming $x_{(i)}$ is the i^{th} element in the ordered statistics $(x_{(1)}, \ldots, x_{(n)})$, and Λ^{-1} denotes the cumulative hazard-rate function.

$$\widehat{a_{j,b}} = \Lambda^{-1}\left((E_{j,X} - \sum_{l=1}^{i-1}\Lambda(x_{(l)}, \underline{\widehat{\phi}}))/(n-i+1), \underline{\widehat{\phi}}\right), \quad \widehat{a_{\kappa}} = x_{(n)}|_{(j=1,\ldots,\kappa)},$$

where

$$e_{j,Z} = E_{\kappa}/\kappa \text{ for every } j.$$

and

$$\begin{aligned} E_{j,Z} &= (n-i+1)\Lambda(\widehat{a_{j,b}}, \underline{\widehat{\phi}}) + \sum_{l=1}^{i-1}\Lambda(x_{(l)}, \underline{\widehat{\phi}}) \\ &= \sum_{i:z_i > a_{j,b}} (\ \Lambda(a_{j,b} \wedge z_i, \underline{\widehat{\phi}}) - \Lambda(a_{j-1}, \underline{\widehat{\phi}}), E_{\kappa} = \sum_{i=1}^{n}\Lambda(z_i, \underline{\widehat{\phi}}). \end{aligned}$$

The $a_{j,b}$ functions for random data, and the $e_{j,Z}$ For the κ selected periods, anticipated failure rates are equal. Statistical data $Y_n^2 = \mathbf{Z}^T\widehat{\mathbf{S}}^{-1}\mathbf{Z}$, where

$$\mathbf{Z} = (Z_1, ..., Z_{\kappa})^x, Z_j = \frac{1}{\sqrt{n}}(\mathbf{W}_{j,Z} - e_{j,Z})|_{(\ j=1,2,\ldots,\kappa)}$$

and $\mathbf{W}_{j,Z}$ can be used to test a hypothesis since it reflects the total number of failures that have been recorded throughout these time-shared. The elements of the B-NIK test statistic

$$Y_n^2 = \sum_{j=1}^{\kappa}\frac{1}{\mathbf{W}_{j,Z}}(\mathbf{W}_{j,Z} - e_{j,Z})^2 + \mathbf{D}_{W,G}$$

where

$$\begin{aligned} \mathbf{D}_{W,G} &= \widehat{\mathbf{V}}^T\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{V}}, \widehat{\mathbf{S}}^{-1} = \widehat{\mathbf{B}}^{-1} + \widehat{\mathbf{M}}^{-1}\widehat{\mathbf{B}}^T\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{M}}\widehat{\mathbf{B}}^{-1}, \\ \widehat{\mathbf{G}} &= [\widehat{g}_{ll'}]_{s\times s} = \widehat{i} - \widehat{\mathbf{M}}\widehat{\mathbf{B}}^{-1}\widehat{\mathbf{M}}^x, \\ \widehat{\mathbf{M}}_{lj} &= \frac{1}{n}\sum_{i:z_i\in\mathbf{I}_j}\rho_i\frac{\partial}{\partial\underline{\phi}}\ln\left[\lambda_{i,\underline{\phi}}(z_i)\right], \mathbf{W}_{j,Z} = \sum_{i:z_i\in\mathbf{I}_j}\rho_i,\ \widehat{\mathbf{B}}_j = n^{-1}\mathbf{W}_{j,Z}, \\ \widehat{\mathbf{V}}_l &= \sum_{j=1}^{\kappa}\widehat{\mathbf{M}}_{lj}\widehat{\mathbf{B}}_j^{-1}\mathbf{Z}_j, \qquad l, l' = 1, ..., s, \\ \widehat{i}_{ll'} &= n^{-1}\sum_{i=1}^{n}\rho_i\frac{\partial}{\partial\underline{\phi}_l}\ln\left[\lambda_{i,\underline{\phi}}(z_i)\right]\frac{\partial}{\partial\underline{\phi}_{l'}}\ln\left[\lambda_{i,\underline{\widehat{\mathbf{P}}}}(z_i)\right] \end{aligned}$$

and

$$\widehat{g}_{ll'} = \widehat{i}_{ll'} - \sum_{j=1}^{\kappa} \widehat{\mathbf{M}}_{lj} \widehat{\mathbf{M}}_{l'j} \hat{A}_j^{-1},$$

and

$$\widehat{\mathbf{M}}_{lj} = \frac{1}{n} \sum_{i:z_i \in \mathbf{I}_j} \rho_i \frac{\partial}{\partial \underline{\phi}} \ln \left[\lambda_{i,\widehat{\underline{\phi}}}(z_i) \right].$$

Censored Assessing for the B-NIK Statistic

Censored simulation studies employing B-NIK statistics are crucial for evaluating and comparing various probability distributions when dealing with censored data. These studies provide significant insights into the effectiveness of B-NIK tests across different types and levels of censoring, assisting in the selection of the appropriate distribution for subsequent analyses.The sample produced ($N = 12000$) is expected to be censored at 20%, with degrees of freedom (DF) set to 5. To assess whether the sample aligns with the null hypothesis of the MGEF model, grouping intervals will be utilized. For different theoretical levels ($\epsilon = 0.01, 0.02, 0.05, 0.1$), we will calculate the average number of non-rejections of the null hypothesis, where $Y^2 \leq \chi^2_\epsilon(r-1)$. Table 4.4 compares the theoretical and empirical levels, showing how closely the calculated empirical level corresponds to the relevant theoretical level. Consequently, we determine that the customized test is ideally suited for the MGEF model.

Table 4.4: Censored assessing for the B-NIK statistic for $\epsilon = 0.01; 0.02; 0.05; 0.1$ and $N = 12000$.

$n \downarrow \& \epsilon \longrightarrow$	$\epsilon = 0.01$	$\epsilon = 0.02$	$\epsilon = 0.05$	$\epsilon = 0.1$
$n = 20$	0.9932	0.9828	0.9532	0.9021
$n = 40$	0.9925	0.9815	0.9524	0.9016
$n = 250$	0.9916	0.9810	0.9514	0.9011
$n = 350$	0.9910	0.9806	0.9510	0.9009
$n = 600$	0.9906	0.9803	0.9508	0.9003
$n = 1000$	0.9903	0.9801	0.9506	0.9001

Based on these findings, we conclude that the empirical significance level of Y_n^2 aligns with the theoretical level of the chi-square distribution concerning degrees of freedom, indicating the statistical level at which the results are significant. Therefore, the evidence suggests that the censored data obtained from the MGEF distribution can be effectively fitted using the proposed test.

4.5 Application of the MGEF Model to Emergency Care Data

This dataset allows for an evaluation of the goodness-of-fit of the MGEF model distribution and its effectiveness in representing the observed patterns and variability in emergency care data. We present point estimates for two fitted models: the MGEF model with a Weibull baseline hazard-rate function and the MGEF model with a Gompertz baseline hazard-rate function. To determine the best model among those fitted to the data, we apply the modified chi-squared test developed by Bagdonavičius and Nikulin (2011)[14]. This statistical method helps assess the validity and applicability of the MGEF distribution in emergency care data, contributing to a better understanding of the underlying survival dynamics and variability in this crucial healthcare context. For more recent datasets, refer to Abiad et al. (2025), Alizadeh et al. (2025), and Das et al. (2025) and Ibrahim et al. (2025).

4.5.1 Validation of the MGEF Model Under the Weibull Baseline Hazard-Rate Function

Suppose that the data follows the MGEF model with a Weibull baseline hazard rate function, the maximum likelihood estimates of the parameter vector ϕ are obtained using the R statistical software (specifically, the BB package) as follows:

$$\begin{aligned}\widehat{\nu} &= 0.822547, \widehat{\varphi} = 0.63951, \widehat{\sigma^2} = 1.03591,\\ \widehat{\beta_1} &= 0.015748, \widehat{\beta_2} = 0.50024, \widehat{\beta_3} = 0.20368,\\ \beta_4 &= -0.39517, \widehat{\beta_5} = 0.27845, \widehat{\beta_6} = 0.83695.\end{aligned}$$

According to Bagdonavičius and Nikulin (2011)[14] for censored data, we take for example 5 intervals ($r = 5$) as number of classes. The elements of the estimated Fisher information matrix $I\left(\widehat{\phi}\right)$ are presented as follows:

$$I\left(\widehat{\phi}\right) = \begin{pmatrix} 1.35 & -3.27 & 0.363 & 0.001 & 1.111 & 1.006 & 0.0236 & 0.9026 & 2.6255 \\ & 0.626 & 0.325 & -2.966 & -0.002 & 0.363 & -9.3025 & 1.8547 & 0.0001 \\ & & 1.025 & -7.263 & 0.965 & 0.0003 & -8.3262 & 0.9681 & 0.1057 \\ & & & 1.954 & 2.125 & -0.252 & 0.00215 & 0.0002 & 0.1547 \\ & & & & 2.152 & 0.097 & 1.0025 & 1.6685 & 0.6326 \\ & & & & & 0.952 & 1.0255 & -6.000 & 2.31125 \\ & & & & & & 0.6685 & -3.2626 & 4.0216 \\ & & & & & & & 0.3155 & 1.2515 \\ & & & & & & & & 1.9658 \end{pmatrix}$$

The computed value of the test statistic for the proposed MGEF model with a Weibull baseline hazard-rate function is $Y_n^2 = 9.0025741$. When this value is compared to the critical value from the chi-squared distribution, $\chi^2_{0.05}(4) = 9.488$, we find that Y_n^2 is less than the critical value. This indicates that we do not reject the null hypothesis at the 5% significance level, suggesting that the data aligns well with the proposed model. Consequently, the emergency care data can be suitably modeled using the MGEF model with a Weibull baseline hazard-rate function. The model provides a good fit, effectively capturing the underlying survival patterns and addressing unobserved heterogeneity in the dataset. This validation highlights the robustness and relevance of the MGEF model in analyzing complex survival data.

4.5.2 Validation of the MGEF Model Under the Gompertz Baseline Hazard-Rate Function

If we assume that the data is distributed according to the MGEF model with a Gompertz baseline hazard-rate function, the maximum likelihood estimator for the parameter vector ϕ can be obtained using the R statistical software (specifically, the BB package) as follows:

$$\begin{aligned}\widehat{\kappa} &= 1.005428, \widehat{\xi} = 0.90035, \widehat{\sigma^2} = 1.01584,\\ \widehat{\beta_1} &= -0.248331, \widehat{\beta_2} = 0.215862, \widehat{\beta_3} = 0.02913,\\ \widehat{\beta_4} &= 0.723841, \widehat{\beta_5} = 0.0700561, \widehat{\beta_6} = 0.68597.\end{aligned}$$

We take $r = 5$ intervals and the estimated Fisher matrix expressed as

$$I\left(\widehat{\phi}\right) = \begin{pmatrix} 0.96 & -6.24 & 0.216 & 0.952 & -4.686 & 1.037 & 0.962 & 1.03325 & 1.002 \\ & 1.025 & 2.00 & 1.003 & 2.003 & 0.002 & 0.951 & 0.001 & 0.209 \\ & & 1.033 & 1.957 & -9.33 & 1.026 & -8.620 & 0.633 & 1.625 \\ & & & 0.966 & 0.022 & 1.025 & 1.855 & 0.001 & 1.251 \\ & & & & 0.326 & 1.856 & 0.097 & -3.36 & 1.003 \\ & & & & & 1.220 & 1.954 & 0.125 & 0.004 \\ & & & & & & 1.204 & 1.025 & 2.032 \\ & & & & & & & 1.965 & 1.006 \\ & & & & & & & & 0.549 \end{pmatrix}.$$

To evaluate how well the emergency care data aligns with the proposed MGEF model featuring a Gompertz baseline hazard-rate function, we compute the statistic from Bagdonavičius and Nikulin (2011)[14], denoted as Y_n^2. The calculated value of this statistic is $Y_n^2 = 8065984$. This test statistic is crucial for assessing the model's fit to the observed data. We then compare Y_n^2 with the critical value from the chi-squared distribution $\chi^2_{0.05}$ at a 5% significance level. The degrees of freedom for this test are based on the number of parameters estimated in the model. In this case, there are 4 degrees of freedom, with $r = 5$ representing the total number of parameters in the MGEF model with the Gompertz baseline hazard-rate function. According to the chi-squared distribution table, the critical value for $\alpha = 5\%$ and 4 degrees of freedom is $\chi^2_{0.05}(5-1) = 9.488$. By comparing the calculated test statistic, we find $Y^2 < \chi^2_{0.05}(5-1) = 9.488$, suggesting that we cannot reject the null hypothesis, which states that the data fits the proposed MGEF model with a Gompertz baseline hazard-rate function, at the 5% significance level. Therefore, we conclude that the emergency care data is compatible with the proposed model. This compatibility indicates that the MGEF model with a Gompertz baseline hazard-rate function provides an adequate fit for the emergency care data, effectively capturing the underlying survival patterns and accounting for unobserved heterogeneity. The robustness of the model is further confirmed by its consistency with the observed data through rigorous statistical testing. Thus, the MGEF model can be confidently utilized for analyzing and predicting survival outcomes in emergency care settings, offering valuable insights for risk assessment and decision-making processes.

4.6 Synthesis of Results and Future Perspectives

4.6.1 Synthesis of Results

This chapter provides a critical discussion of the findings presented in this thesis, highlighting the significance and strengths of the proposed Modified Xgamma-Gompertz Extended Family (MxG-GEF) of distributions. The model developed in this work contributes substantively to the body of knowledge in lifetime data analysis and survival modeling, particularly by addressing the limitations of existing models in capturing diverse hazard rate behaviors and offering improved flexibility without sacrificing tractability.

Theoretical and Structural Merits :

One of the central strengths of this study lies in the formulation of the MxG-GEF distribution, which integrates the Modified Xgamma generator with the Gompertz baseline. This construction yields a distributional family capable of modeling a wide range of hazard rate shapes, including increasing, decreasing, bathtub, and unimodal forms. Such versatility is essential in accurately describing failure mechanisms across varied scientific domains, such as biomedicine, engineering reliability, and demography.

The model generalizes several known distributions, including the Gompertz, Xgamma-Gompertz, and exponential-Gompertz, which are recovered as special or limiting cases. This inclusion reinforces the generality and coherence of the proposed model and permits hypothesis testing between nested models using standard likelihood ratio procedures. Moreover, explicit expressions for the probability density function, cumulative distribution function, quantile function, and moments have been derived,

affirming the model's mathematical tractability.

The availability of closed-form quantile expressions is particularly valuable in simulation and bootstrap-based inference, and facilitates practical implementation of random sample generation procedures, which are essential in applied settings such as reliability forecasting and risk analysis.

Simulation-Based Assessment of Estimation Procedures :

To examine the inferential performance of the model's parameters, a comprehensive simulation study was conducted using maximum likelihood estimation. The simulation results presented in this chapter confirm that the estimators of the MxG-GEF model exhibit desirable asymptotic properties. For a variety of parameter settings and across increasing sample sizes, the estimators consistently demonstrated asymptotic unbiasedness, efficiency, and consistency.

These properties are reflected in the behavior of the bias and root mean square error (RMSE), both of which decline as sample size increases, in accordance with theoretical expectations. Importantly, the numerical stability of the optimization process in estimating the parameters of the MxG-GEF model indicates that the model is computationally well-behaved. This distinguishes it from certain other generator-based families, where overparameterization or multimodality of the likelihood function can impede convergence.

The simulation findings not only confirm the theoretical validity of the estimation procedures but also demonstrate their robustness and practical applicability, especially in scenarios where sample sizes may be moderate to large.

Empirical Relevance and Real Data Applications:

The empirical component of the study further substantiates the utility of the MxG-GEF model. A real-life dataset was examined: it concerning emergency care data patients.

In this application, the proposed model outperformed a range of competing models-including classical (e.g., Gompertz, Weibull) in terms of log-likelihood.

These results demonstrate that the MxG-GEF is capable of capturing complex failure patterns with high fidelity. Its performance in both biomedical and industrial contexts suggests that the model is broadly applicable. This generalizability is a distinguishing feature and affirms the practical relevance of the theoretical development.

Goodness-of-Fit Evaluation via Modified Chi-Squared Test:

Beyond conventional model comparison metrics, a modified version of the Pearson chi-squared goodness-of-fit test was employed to assess the adequacy of the MxG-GEF model. This version of the test accounts for the continuous nature of lifetime data and adjusts the degrees of freedom appropriately for estimated parameters.

The test outcomes provided strong support for the proposed model, with non-significant p-values indicating no evidence against the fit. This result was consistent across both datasets. The adoption of this modified test adds rigor to the empirical validation process and offers a perspective on model adequacy. It also strengthens the conclusion that the MxG-GEF model provides a statistically sound representation of the underlying data-generating mechanisms.

Scientific Importance and Methodological Contribution:

From a methodological standpoint, the introduction of the MxG-GEF model represents an important contribution to the literature on flexible lifetime distributions. It leverages the benefits of the Modified Xgamma generator to enrich the hazard rate structure while maintaining mathematical manageability. The balance achieved between flexibility and analytical simplicity is particularly noteworthy, as many existing models introduce excessive complexity that hampers interpretation and

inference.

Furthermore, the model's ability to unify and extend several known distributions within a single coherent framework contributes to the broader endeavor of building flexible yet interpretable survival models. The work is in line with recent developments in generalized distribution theory, yet avoids the pitfalls commonly associated with overparametrized or poorly tractable models.

4.6.2 Future Perspectives

The work undertaken in this chapter lays a solid foundation for future developments in the modeling of lifetime data using generalized generator families. Nevertheless, several avenues of further inquiry remain open and merit rigorous exploration.

A natural extension of the present work involves the development of **Bayesian inference procedures** for the MxG-GEF model. While maximum likelihood estimation offers consistency and efficiency under regularity conditions, Bayesian methods could provide a more flexible inferential framework, especially in the presence of prior knowledge or small sample sizes. The construction of appropriate prior distributions, posterior characterizations, and computational algorithms such as Markov chain Monte Carlo (MCMC) for this class of models would be an important addition to the existing methodology.

Another promising direction involves **multivariate generalizations** of the MxG-GEF distribution. Many real-world applications involve correlated lifetimes or repeated failure times, which cannot be adequately modeled using univariate techniques. Constructing multivariate or copula-based extensions of the MxG-GEF family would allow for modeling such dependence structures while retaining the flexibility of the marginal distributions.

In addition, the incorporation of **covariate information** through accelerated failure time (AFT) or proportional hazards (PH) models is of practical significance. Embedding the MxG-GEF distribution into regression-type survival models would enable practitioners to assess the effects of explanatory variables on survival times, which is vital in clinical, epidemiological, and reliability studies.

Future investigations could explore the **robustness and performance of the MxG-GEF model in high-dimensional settings**, such as in big data applications or under complex censoring schemes. Extending the current framework to accommodate left truncation, interval censoring, or competing risks would significantly broaden its applicability in modern survival data analysis.

In summary, while this chapter provides a robust and versatile model for univariate lifetime data, it opens multiple directions for theoretical refinement, methodological expansion, and practical enhancement, which can serve as a fertile ground for future statistical research.

Chapter 5

Conclusion and Future Work

This research makes a significant contribution to survival analysis by introducing innovative fragility models: the Quasi-Xgamma (QXg-F) and the Mixed Gamma-Exponential (MGEF) models. These models effectively address the issue of unobserved heterogeneity, a major challenge in survival analysis, particularly when data exhibit asymmetric behaviors or extreme effects.

The QXg-F model, by incorporating an additional shape parameter, provides essential flexibility to capture a broader range of latent behaviors, thereby improving risk modeling across various contexts, especially in emergency care. Simulation results demonstrate that this model is both theoretically robust and performs well in real-world applications.

On the other hand, the MGEF model stands out due to its hybrid structure, combining multiplicative and additive effects, allowing for a more nuanced understanding of frailty. This is particularly relevant in clustered contexts, such as insurance portfolios or groups of hospitalized patients, as it effectively integrates latent effects influencing survival outcomes.

Rigorous evaluations of the models using the NIK-RR and B-NIK goodness-of-fit tests confirmed their statistical validity across different censoring regimes. The flexibility of both QXg-F and MGEF models allows them to adapt to varied datasets, improving prediction accuracy. In healthcare, these models facilitate the precise identification of at-risk groups, enabling more informed clinical decision-making tailored to patient needs.

Moreover, in the insurance sector, the QXg-F model demonstrated superior goodness-of-fit compared to classical fragility models due to its ability to accommodate asymmetric and heavy-tailed fragility distributions, which is crucial for actuarial forecasting and solvency assessments.

Finally, this research opens new avenues for future work, including extending the models to Bayesian environments, adapting them for longitudinal and high-dimensional data, and implementing them in dedicated software tools. By emphasizing the importance of accounting for latent heterogeneity, this thesis lays the groundwork for a better understanding of survival dynamics across various fields, ranging from medicine to engineering and economics.

Bibliography

[1] Aalen, O .O. (1988). Heterogeneity in Survival Analysis. Statistics in Medicine, 7, 1121-1137.

[2] Aalen, O.O. (1992). Modelling Heterogeneity in Survival Analysis by the Compound Poisson Distribution. Annals of Applied Probability, 4(2), 951-972.

[3] Aalen, O.O, Tretli, S. (1999). Analysing the incidence of testis cancer by means of a frailty model. Cancer Causes and Control, 10, 285-292.

[4] Abiad, M., Alsadat, N., Abd El-Raouf, M. M., Yousof, H. M., & Kumar, A. (2025). Different copula types and reliability applications for a new fisk probability model. Alexandria Engineering Journal, 110, 512-526.

[5] Abouelmagd, T. H. M., Hamed, M. S., Hamedani, G. G., Ali, M. M., Goual, H., Korkmaz, M. C., & Yousof, H. M. (2019). The zero truncated Poisson Burr X family of distributions with properties, characterizations, applications, and validation test. Journal of Nonlinear Sciences and Applications, 12(5), 314-336.

[6] Acerbi, C., Tasche, D. (2002). On the coherence of expected shortfall. Journal of banking finance, 26(7), 1487-1503.

[7] Ahmed, B., Ali, M. M. and Yousof, H. M. (2022). A Novel G Family for Single Acceptance Sampling Plan with Application in Quality and Risk Decisions, Annals of Data Science, 10.1007/s40745-022-00451-3

[8] Alizadeh, M., Afshari, M., Contreras-Reyes, J. E., Mazarei, D., & Yousof, H. M. (2024). The Extended Gompertz Model: Applications, Mean of Order P Assessment and Statistical Threshold Risk Analysis Based on Extreme Stresses Data. IEEE Transactions on Reliability, doi: 10.1109/TR.2024.3425278.

[9] Alizadeh, M., Afshari, M., Cordeiro, G. M., Ramaki, Z., Contreras-Reyes, J. E., Dirnik, F., & Yousof, H. M. (2025). A New Weighted Lindley Model with Applications to Extreme Historical Insurance Claims. Stats, 8(1), 8.

[10] Alizadeh, M., Afshari, M., Ranjbar, V., Merovci, F. and Yousof, H. M. (2023). A novel XGamma extension: applications and actuarial risk analysis under the reinsurance data. São Paulo Journal of Mathematical Sciences, 1-31.

[11] Aljadani, A., Mansour, M. M., & Yousof, H. M. (2024). A Novel Model for Finance and Reliability Applications: Theory, Practices and Financial Peaks Over a Random Threshold Value-at-Risk Analysis. Pakistan Journal of Statistics and Operation Research, 20(3), 489-515. https://doi.org/10.18187/pjsor.v20i3.4439

[12] Andersen, P. K., Geskus, R. B., de Witte, T., & Putter, H. (2012). Competing risks in epidemiology: possibilities and pitfalls. International journal of epidemiology, 41(3), 861-870.

[13] Anwar, N., Ahmad, I., Kiani, A. K., Shoaib, M., & Raja, M. A. Z. (2023). Intelligent solution predictive networks for non-linear tumor-immune delayed model. Computer Methods in Biomechanics and Biomedical Engineering, 1-28.

[14] Bagdonavičius, V., and Nikulin, M. (2011). Chi-squared goodness-of-fit test for right censored data. International Journal of Applied Mathematics and Statistics, 24, 30-50.

[15] Bagdonavičius, V., Levuliene, R., J., and Nikulin, M. (2013). Chi-squared goodness-of-fit tests for parametric accelerated failure time models. Communications in Statistics-Theory and Methods, 42(15), 2768-2785.

[16] Balakrishnan, N.Peng., Y. (2006). Generalized gamma frailty model. Stat Med, 25(16), 2797-2816.

[17] Beard, R. E. (1959, January). Note on some mathematical mortality models. In Ciba Foundation Symposium-The Lifespan of Animals (Colloquia on Ageing) (Vol. 5, pp. 302-311). Chichester, UK: John Wiley Sons, Ltd.

[18] Breslow, N. E., Day, N. E., Heseltine, E. (1980). Statistical methods in cancer research.

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

[20] Collett, D. (1994). Modelling survival data. In Modelling survival data in medical research (pp. 53-106). Springer US.

[21] Collett, D. (2015). Modelling survival data in medical research. Chapman and Hall/CRC.

[22] Collett, D. (2023). Modelling survival data in medical research. Chapman and Hall/CRC.

[23] Cox DR. (1972). Regression models and life-tables. J Roy Stat Soc: Ser B (Methodol), 34(2), 187-202.

[24] Cox, D. R. (1958). Two further applications of a model for binary regression. Biometrika, 45(3/4), 562-565.

[25] Das, J., Hazarika, P. J., Alizadeh, M., Contreras-Reyes, J. E., Mohammad, H. H., & Yousof, H. M. (2025). Economic Peaks and Value-at-Risk Analysis: A Novel Approach Using the Laplace Distribution for House Prices. Mathematical and Computational Applications, 30(1), 4.

[26] Devlin, K. (2010). The unfinished game: Pascal, Fermat, and the seventeenth-century letter that made the world modern. Basic Books.

[27] Duchateau, L., Janssen, P. (2008) The Frailty Model. Springer, New York.

[28] Elbatal, I., Diab, L. S., Ghorbal, A. B., Yousof, H. M., Elgarhy, M. and Ali, E. I. (2024). A new losses (revenues) probability model with entropy analysis, applications and case studies for value-at-risk modeling and mean of order-P analysis. AIMS Mathematics, 9(3), 7169-7211.

[29] Elbers, C., Ridder, G. (1982). True and spurious duration dependence: The identifiability of the proportional hazard model. Rev Econ Stud, 49(3), 403-409.

[30] Emam, W.; Tashkandy, Y.; Goual, H.; Hamida, T.; Hiba, A.; Ali, M.M.; Yousof, H.M.; Ibrahim, M. A New One-Parameter Distribution for Right Censored Bayesian and Non-Bayesian Distributional Validation under Various Estimation Methods. Mathematics 2023, 11, 897. https://doi.org/10.3390/math11040897.

[31] Fisher, L. D., Lin, D. Y. (1999). Time-dependent covariates in the Cox proportional-hazards regression model. Annual review of public health, 20(1), 145-157.

[32] Furman, E., Landsman, Z. (2006). Tail variance premium with applications for elliptical portfolio of risks. ASTIN Bulletin: The Journal of the IAA, 36(2), 433-462.

[33] Gnedenko, B. V., & Kolmogorov, A. N. (1960). Aleksandr Yakovlevich Khinchin (1894-1959)-Obituary. Russian Mathematical Surveys, 15(4), 93.

[34] 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.

[35] Goual, H., and Yousof, H. M. (2020a). Validation of Burr XII inverse Rayleigh model via a modified chi-squared goodness-of-fit test. Journal of Applied Statistics, 47(3), 393-423.

[36] Goual, H., Yousof, H. M. and Ali, M. M. (2020b). Lomax inverse Weibull model: properties, applications, and a modified Chi-squared goodness-of-fit test for validation. Journal of Nonlinear Sciences & Applications (JNSA), 13(6), 330-353.

[37] Goual, H., Yousof, H. M., & Ali, M. M. (2019). Validation of the odd Lindley exponentiated exponential by a modified goodness of fit test with applications to censored and complete data. Pakistan Journal of Statistics and Operation Research, 15(3), 745-771.

[38] Hafida Goual, Hamami, L., S. Hamed, M. (2025). A New Mixed Gamma-Exponential Frailty Model under Heterogeneity Problem with Validation Testing for Emergency Care Data. Statistics, Optimization Information Computing. https://doi.org/10.19139/soic-2310-5070-2451.

[39] Hamed, M. S., Cordeiro, G. M. and Yousof, H. M. (2022). A New Compound Lomax Model: Properties, Copulas, Modeling and Risk Analysis Utilizing the Negatively Skewed Insurance Claims Data. Pakistan Journal of Statistics and Operation Research, 18(3), 601-631. https://doi.org/10.18187/pjsor.v18i3.3652

[40] Hamedani, G. G., Goual, H., Emam, W., Tashkandy, Y., Ahmad Bhatti, F., Ibrahim, M., & Yousof, H. M. (2023). A new right-skewed one-parameter distribution with mathematical characterizations, distributional validation, and actuarial risk analysis, with applications. Symmetry, 15(7), 1297.

[41] Hashem, A. F., Alotaibi, N., Alyami, S. A., Abdelkawy, M. A., Elgawad, M. A. A., Yousof, H. M., & Abdel-Hamid, A. H. (2024). Utilizing Bayesian inference in accelerated testing models under constant stress via ordered ranked set sampling and hybrid censoring with practical validation. Scientific Reports, 14(1), 14406.

[42] Hashempour, M., Alizadeh, M. and Yousof, H. M. (2023). A New Lindley Extension: Estimation, Risk Assessment and Analysis Under Bimodal Right Skewed Precipitation Data. Annals of Data Science, 1-40.

[43] Hougaard, P. (2012). Analysis of multivariate survival data. Springer Science & Business Media, Berlin.

[44] Hougaard, P. (2000). Analysis of multivariate survival data (Vol. 564). New York: Springer.

[45] Ibrahim, J., Chen, M., Sinha, D. (2001). Bayesian survival analysis springer series in statistics. Springer, New York, 978-981.

[46] Ibrahim, M., Ali, M. M., Goual, H. and Yousof, H. M. (2022). Censored and uncensored validation for the double Burr XII model using a new Nikulin-Rao-Robson goodness-of-fit test with Bayesian and non-Bayesian estimation, Pakistan Journal of Statistics and Operation Research, 18(4), 2022.

[47] Ibrahim, M., Altun, E., Goual, H., and Yousof, H. M. (2020). Modified goodness-of-fit type test for censored validation under a new Burr type XII distribution with different methods of estimation and regression modeling. Eurasian Bulletin of Mathematics, 3(3), 162-182.

[48] Ibrahim, M., Ansari, S. I., Al-Nefaie, A. H., & Yousof, H. M. (2025). A New Version of the Inverse Weibull Model with Properties, Applications and Different Methods of Estimation. Statistics, Optimization & Information Computing, 13(3), 1120-1143. https://doi.org/10.19139/soic-2310-5070-1658

[49] Ibrahim, M., Hamedani, G. G., Butt, N. S. and Yousof, H. M. (2022). Expanding the Nadarajah Haghighi Model: Copula, Censored and Uncensored Validation, Characterizations and Applications. Pakistan Journal of Statistics and Operation Research, 18(3), 537-553. https://doi.org/10.18187/pjsor.v18i3.3420

[50] Ibrahim, M., Yadav, A. S., Yousof, H. M., Goual, H., & Hamedani, G. G. (2019). A new extension of Lindley distribution: modified validation test, characterizations and different methods of estimation. Communications for Statistical Applications and Methods, 26(5), 473-495.

[51] Ibrahim, M.; Emam, W.; Tashkandy, Y.; Ali, M.M.; Yousof, H.M. (2023). Bayesian and Non-Bayesian Risk Analysis and Assessment under Left-Skewed Insurance Data and a Novel Compound Reciprocal Rayleigh Extension. Mathematics 2023, 11, 1593. https://doi.org/10.3390/math11071593

[52] Ibrahim. M., Aidi, K., Ali, M. M. and Yousof, H. M. (2021). The Exponential Generalized Log-Logistic Model: Bagdonavičius-Nikulin test for Validation and Non-Bayesian Estimation Methods. Communications for Statistical Applications and Methods, 29(1), 681–705.

[53] Khedr, A. M., Nofal, Z. M., El Gebaly, Y. M. and Yousof, H. M. (2023). A Novel Family of Compound Probability Distributions: Properties, Copulas, Risk Analysis and Assessment under a Reinsurance Revenues Data Set. Thailand Statistician, forthcoming.

[54] Korkmaz, M. Ç., Altun, E., Yousof, H. M., Afify, A. Z. and Nadarajah, S. (2018). The Burr X Pareto Distribution: Properties, Applications and VaR Estimation. Journal of Risk and Financial Management, 11(1), 1.

[55] Lancaster, T. (1979) Econometric methods for the duration of unemploy- ment. Econometrica 47, 939–956.

[56] Laplace, P. S. (1812). Théorie analytique des probabilités [Analytical theory of probabilities]. Courcier.

[57] Loprinzi, C. L., Laurie, J. A., Wieand, H. S., Krook, J. E., Novotny, P. J., Kugler, J. W. and Klatt, N. E. (1994). Prospective evaluation of prognostic variables from patient-completed questionnaires. North Central Cancer Treatment Group. Journal of Clinical Oncology, 12(3), 601-607.

[58] Loubna, H., Goual, H., Alghamdi, F. M., Mustafa, M. S., Tekle Mekiso, G., Ali, M. M., ... & Yousof, H. M. (2024). The quasi-xgamma frailty model with survival analysis under heterogeneity problem, validation testing, and risk analysis for emergency care data. Scientific Reports, 14(1), 8973.

[59] Mansour, M. M., Butt, N. S., Ansari, S. I., Yousof, H. M., Ali, M. M., & Ibrahim, M. (2020). A new exponentiated Weibull distribution's extension: copula, mathematical properties and applications. Contributions to Mathematics, 1 (2020) 57–66. DOI: 10.47443/cm.2020.0018

[60] Mansour, M. M., Ibrahim, M., Aidi, K., Butt, N. S., Ali, M. M., Yousof, H. M., & Hamed, M. S. (2020). A New Log-Logistic Lifetime Model with Mathematical Properties, Copula, Modified Goodness-of-Fit Test for Validation and Real Data Modeling. Mathematics, 8(9), 1508.

[61] Mansour, M., Korkmaz, M. Ç., Ali, M. M., Yousof, H. M., Ansari, S. I., & Ibrahim, M. (2020). A generalization of the exponentiated Weibull model with properties, Copula and application. Eurasian Bulletin of Mathematics, 3(2), 84-102.

[62] Mansour, M., Rasekhi, M., Ibrahim, M., Aidi, K., Yousof, H. M., & Elrazik, E. A. (2020). A New Parametric Life Distribution with Modified Bagdonavičius–Nikulin Goodness-of-Fit Test for Censored Validation, Properties, Applications, and Different Estimation Methods. Entropy, 22(5), 592.

[63] Mansour, M., Yousof, H. M., Shehata, W. A. M., & Ibrahim, M. (2020). A new two parameter Burr XII distribution: properties, copula, different estimation methods and modeling acute bone cancer data. Journal of Nonlinear Science and Applications, 13(5), 223-238.

[64] Mazucheli, J., Coelho-Barros, EA., Achcar, JA. (2016). An alternative reparametrization for the weighted lindley distribution. Pesquisa Operacional , 36(2), 345-353.

[65] McGilchrist, C.A., Aisbett, C.W. (1991). Regression with Frailty in Survival Analysis. Biometrics, 47, 461-466.

[66] Meeker, W. Q., & Escobar, L. A. (1978). Statistical methods for reliability data using SAS software. Technometrics, 20(3), 245-247.

[67] Mota, A., Milani, E.A., Calsavara, V.F., Tomazella, V.L., Leão, J., Ramos, P.L., Ferreira, P.H. and Louzada, F. (2021). Weighted Lindley frailty model: estimation and application to lung cancer data. Lifetime Data Analysis, 27(4), 561-587.

[68] Nikulin, M.S. (1973a). Chi-squared test for normality. in proceedings of the International Vilnius Conference on Probability Theory and Mathematical Statistics, 2, 119-122.

[69] Nikulin, M.S, (1973b). Chi-squared test for continuous distributions with shift and scale parameters, Theory of Probability and its Aplications. 18, 559-568.

[70] Nikulin, M.S, (1973c). On a Chi-squared test for continuous distributions, Theory of Probability and its Aplications. 19, 638-639.

[71] Pearson, K. (1922). On the systematic fitting of frequency curves. Biometrika, 14(3-4), 311-315. https://doi.org/10.1093/biomet/14.3-4.311

[72] Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. Journal of clinical epidemiology, 49(12), 1373-1379.

[73] Peto, R., Pike, M., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., ... , Smith, P. G. (1977). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. analysis and examples. British journal of cancer, 35(1), 1-39.

[74] Pickles, A., Crouchley, R. (1995). A comparison of frailty models for multivariate survival data. Stat Med , 14(13), 1447-1461.

[75] Pitacco, E. (2019). Heterogeneity in mortality: a survey with an actuarial focus. European Actuarial Journal, 9, 3-30.

[76] Rao, K. C., Robson, D. S. (1974). A Chi-square statistic for goodness-of-fit tests within the exponential family. Communication in Statistics, 3, 1139-1153.

[77] Rasekhi, M., Altun, E., Alizadeh, M. and Yousof, H. M. (2022). The Odd Log-Logistic Weibull-G Family of Distributions with Regression and Financial Risk Models. Journal of the Operations Research Society of China, 10(1), 133-158.

[78] Robert, C., Casella, G.(2013). Monte Carlo statistical methods. Springer Science & Business Media, Berlin.

[79] Royston, P., & Altman, D. G. (2013). External validation of a Cox prognostic model: Principles and methods. BMC Medical Research Methodology, 13(1), 33. https://doi.org/10.1186/1471-2288-13-33

[80] Saint-Pierre, P. (2021, April). Introduction à l'analyse des durées de survie [Introduction to survival analysis]. Institut de Mathématiques de Toulouse, Université Paul Sabatier - Toulouse III.

[81] Salah, M. M., El-Morshedy, M., Eliwa, M. S. and Yousof, H. M. (2020). Expanded Fréchet Model: Mathematical Properties, Copula, Different Estimation Methods, Applications and Validation Testing. Mathematics, 8(11), 1949.

[82] Salem, M., Emam, W., Tashkandy, Y., Ibrahim, M., Ali, M. M., Goual, H., & Yousof, H. M. (2023). A new lomax extension: Properties, risk analysis, censored and complete goodness-of-fit validation testing under left-skewed insurance, reliability and medical data. Symmetry, 15(7), 1356.

[83] Sarma, S., Ahmed, I., Begum, A. (2020). A new two parameter gamma-exponential mixture. J. Math. Comput. Sci., 11(1), 414-426.

[84] Sen, S., Chandra, N. (2017). The quasi xgamma distribution with application in bladder cancer data. Journal of data science, 15(1), 61-76.

[85] Shapiro, S. S., Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). Biometrika, 52(3-4), 591-611.

[86] Shehata, W. A., Goual, H., Hamida, T., Hiba, A., Hamedani, G. G., Al-Nefaie, A. H., Ibrahim, M., Butt, N. S., Osman, R. M. A., and Yousof, H. M. (2024). Censored and Uncensored Nikulin-Rao-Robson Distributional Validation: Characterizations, Classical and Bayesian estimation with Censored and Uncensored Applications. Pakistan Journal of Statistics and Operation Research, 20(1), 11-35.

[87] Shoaib, M., Anwar, N., Ahmad, I., Naz, S., Kiani, A. K., & Raja, M. A. Z. (2023). Neuro-computational intelligence for numerical treatment of multiple delays SEIR model of worms propagation in wireless sensor networks. Biomedical Signal Processing and Control, 84, 104797.

[88] Shoaib, M., Anwar, N., Ahmad, I., Naz, S., Kiani, A. K., & Raja, M. A. Z. (2022). Intelligent networks knacks for numerical treatment of nonlinear multi-delays SVEIR epidemic systems with vaccination. International Journal of Modern Physics B, 36(18), 2250100.

[89] Shrahili, M.; Elbatal, I. and Yousof, H. M. Asymmetric Density for Risk Claim-Size Data: Prediction and Bimodal Data Applications. Symmetry 2021, 13, 2357. https://doi.org/10.3390/sym13122357

[90] Steyerberg, E. W. (2016). FRANK E. HARRELL, Regression Modeling Strategies: With Applications, to Linear Models, Logistic and Ordinal Regression, and Survival Analysis, Heidelberg: Springer.

[91] Teghri, S., Goual, H., Loubna, H., Butt, N. S., Khedr, A. M., Yousof, H. M., Ibrahim, M. & Salem, M. (2024). A New Two-Parameters Lindley-Frailty Model: Censored and Uncensored Schemes under Different Baseline Models: Applications, Assessments, Censored and Uncensored Validation Testing. Pakistan Journal of Statistics and Operation Research, 109-138.

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

[93] Weibull, W. (1951). A statistical distribution function of wide applicability. Journal of applied mechanics.

[94] Wienke, A. (2010). Frailty models in survival analysis. CRC Press, Boca Raton.

[95] Yadav, A. S., Goual, H., Alotaibi, R. M., Rezk, H., Ali, M. M., & Yousof, H. M. (2020). Validation of the Topp-Leone-Lomax model via a modified Nikulin-Rao-Robson goodness-of-fit test with different methods of estimation. Symmetry, 12(1), 57.

[96] Yadav, A. S., Shukla, S., Goual, H., Saha, M. and Yousof, H. M. (2022). Validation of xgamma exponential model via Nikulin-Rao-Robson goodness-of- fit test under complete and censored sample with different methods of estimation. Statistics, Optimization & Information Computing, 10(2), 457-483.

[97] Yousof, H. M., Aidi, K., Hamedani, G. G and Ibrahim, M. (2021). A new parametric lifetime distribution with modified Chi-square type test for right censored validation, characterizations and different estimation methods. Pakistan Journal of Statistics and Operation Research, 17(2), 399-425.

[98] Yousof, H. M., Ali, M. M., Aidi, K., Ibrahim, M. (2023). The modified Bagdonavičius-Nikulin goodness-of-fit test statistic for the right censored distributional validation with applications in medicine and reliability. Statistics in Transition New Series, 24(4), 1-18.

[99] Yousof, H. M., Aljadani, A., Mansour, M. M., & Abd Elrazik, E. M. (2024). A New Pareto Model: Risk Application, Reliability MOOP and PORT Value-at-Risk Analysis. Pakistan Journal of Statistics and Operation Research, 20(3), 383-407. https://doi.org/10.18187/pjsor.v20i3.4151

[100] Yousof, H. M., Goual, H., Emam, W., Tashkandy, Y., Alizadeh, M., Ali, M. M., & Ibrahim, M. (2023). An Alternative Model for Describing the Reliability Data: Applications, Assessment, and Goodness-of-Fit Validation Testing. Mathematics, 11(6), 1308.

[101] Yousof, H. M., Goual, H., Hamida, T., Hiba, A., Hamedani, G.G. and Ibrahim, M. (2022). Censored and Uncensored Nikulin-Rao-Robson Distributional Validation: Characterizations, Classical and Bayesian estimation with Applications.

[102] Yousof, H. M., Goual, H., Khaoula, M. K., Hamedani, G. G., Al-Aefaie, A. H., Ibrahim, M., ... & Salem, M. (2023). A novel accelerated failure time model: Characterizations, validation testing, different estimation methods and applications in engineering and medicine. Pakistan Journal of Statistics and Operation Research, 19(4), 691-717.

[103] Yousof, H.M.; Tashkandy, Y.; Emam, W.; Ali, M.M.; Ibrahim, M. (2023). A New Reciprocal Weibull Extension for Modeling Extreme Values with Risk Analysis under Insurance Data. Mathematics 2023, 11, 966. https://doi.org/10.3390/ math11040966