Modeling based on a reparameterized Birnbaum-Saunders distribution for survival data

Jeremias da Silva Leão

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA UFSCar – USP

Jeremias da Silva Leão

Modeling based on a reparameterized Birnbaum-Saunders distribution for analysis of survival data

Doctoral dissertation submitted to the Departamento de Estatística – Des/UFSCar and to Instituto de Ciências Matemáticas e de Computação – ICMC-USP, in partial fulfillmente of the requirements for the degree of the Doctorate in Statistics – Programa Interinstitucional de Pós-Graduação em Estatística UFSCar – USP.

Advisor: Prof. Dra. Vera Lucia Damasceno Tomazella

Co-advisor: Prof. Dr. Victor Eliseo Leiva Sanchez

São Carlos
January 2017

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA UFSCar – USP

Jeremias da Silva Leão

Modelagem baseada na distribuição Birnbaum-Saunders reparametrizada para análise de dados de sobrevivência

Tese apresentada ao Departamento de Estatística – Des/UFSCar e ao Instituto de Ciências Matemáticas e de Computação – ICMC – USP, como parte dos requisitos para obtenção do título de Mestre ou Doutor em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística UFSCar – USP.

Orientador: Prof. Dra. Vera Lucia Damasceno Tomazella

Co-orientador: Prof. Dr. Victor Eliseo Leiva Sanchez

São Carlos
Janeiro de 2017

Ficha catalográfica elaborada pelo DePT da Biblioteca Comunitária UFSCar
Processamento Técnico
com os dados fornecidos pelo(a) autor(a)

Leão, Jeremias da Silva
L437m Modeling based on a reparameterized Birnbaum-Saunders distribution for analysis of survival data / Jeremias da Silva Leão. -- São Carlos : UFSCar, 2017.
117 p.

Tese (Doutorado) -- Universidade Federal de São Carlos, 2017.

1. Birnbaum-Saunders distribution. 2. Cure rate model. 3. Diagnostic analysis. 4. Frailty model. 5. Likelihood estimation. I. Título.

UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia

Programa Interinstitucional de Pós-Graduação em Estatística

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a defesa de tese de doutorado do candidato Jeremias da Silva Leão, realizada em 09/01/2017:

Profa. Dra. Vera Lucia Damasceno Tomazella
UFSCar

Prof. Dr. Filidor Edilfonso Vilca Labra
UNICAMP

p/

Prof. Dr. Helton Saulo Bezerra dos Santos
UFG

Prof. Dr. Manoel Ferreira dos Santos Neto
UFCG

Prof. Dr. Vicente Garibay Cancho
USP

Certifico que a sessão de defesa foi realizada com a participação à distância do membro Prof. Dr. Helton Saulo Bezerra dos Santos e, depois das arguições e deliberações realizadas, o participante à distância está de acordo com o conteúdo do parecer da comissão examinadora redigido no relatório de defesa do aluno Jeremias da Silva Leão.

Profa. Dra. Vera Lucia Damasceno Tomazella
Presidente da Comissão Examinadora
UFSCar

To my parents

ACKNOWLEDGEMENTS

This work could never have been completed without ample guidance, assistance and encouragement. In this respect, I would like to thank the following persons:

My supervisors Prof. Vera Tomazella and Prof. Victor Leiva for their patience, encouragement and guidance. I am indeed fortunate to learn from scholars of your calibre.

A special word of thanks to Helton Saulo for all the contribution during the development of the thesis.

The Departament of Statistics at UFSCar and USP.

The UFSCar/USP community, Isabel, Victor, Gláucia, Humberto and unknown servants, for support.

My UFSCar/USP friends, especially Edgar, José Clelto, Roberta, Daiane, Mauro, Pedro Ramos, Rafael Paixão, David and Andrey.

My professors Francisco Cysneiros, Leandro Rêgo, Maurício Motta, Luis Ernesto, Galvão, Vicente and Rafael Stern, for sharing some of your knowledge.

My UFPI friends, Máx, Valmária, Fernando, Cleide and Roney for the support when I left the institution.

My UFAM friends, Cardoso, James, Amazoneida, Joceli, José Raimundo and others for their support in my release to complete the doctorate degree.

My old friends, Hemílio, Josimar, Lutemberg, Marcelo, Manoel and William Marciano for unceasing friendship.

I would like to express my sincere thanks and gratitude to my parents, Antonia and Pedro, for all effort you put in my education. Thank you for your love and support. Also to my brother Paulo and my sister Miria as well as to my niece Ingrid and my sister-in-law Elizandra.

My wife, Themis and my daughters Lívia and Laís. Thank you for your love and companionship.

The committee members, my profound thanks.

Special thanks to all who contributed directly and indirectly to this achievement.

“ A vida é o dever que nós trouxemos para fazer em casa.
Quando se vê, já são seis horas!
Quando de vê, já é sexta-feira!
Quando se vê, já é natal...
Quando se vê, já terminou o ano...
Quando se vê perdemos o amor da nossa vida.
Quando se vê passaram 50 anos!
Agora é tarde demais para ser reprovado...
Se me fosse dado um dia, outra oportunidade, eu nem olhava o relógio.
Seguiria sempre em frente e iria jogando pelo caminho a casca dourada e inútil das horas...
Seguraria o amor que está a minha frente e diria que eu o amo...
E tem mais: não deixe de fazer algo de que gosta devido à falta de tempo.
Não deixe de ter pessoas ao seu lado por puro medo de ser feliz.
A única falta que terá será a desse tempo que, infelizmente, nunca mais voltará. ”
(Mário Quintana)

RESUMO

LEÃO, J. S.. **Modeling based on a reparameterized Birnbaum-Saunders distribution for analysis of survival data**. 2017. 117 f. Doctoral dissertation (Doctorate Candidate joint Graduate Program in Statistics DEs-UFSCar/ICMC-USP) – Instituto de Ciências Matemáticas e de Computação (ICMC/USP), São Carlos – SP.

Nesta tese propomos modelos baseados na distribuição Birnbaum-Saunders reparametrizada introduzida por Santos-Neto et al. (2012) e Santos-Neto et al. (2014), para análise dados de sobrevivência. Incialmente propomos o modelo de fragilidade Birnbaum-Saunders sem e com covariáveis observáveis. O modelo de fragilidade é caracterizado pela utilização de um efeito aleatório, ou seja, de uma variável aleatória não observável, que representa as informações que não podem ou não foram observadas tais como fatores ambientais ou genéticos, como também, informações que, por algum motivo, não foram consideradas no planejamento do estudo. O efeito aleatório (a "fragilidade") é introduzido na função de risco de base para controlar a heterogeneidade não observável. Usamos o método de máxima verossimilhança para estimar os parâmetros do modelo. Avaliamos o desempenho dos estimadores sob diferentes percentuais de censura via estudo de simulações de Monte Carlo. Considerando variáveis regressoras, derivamos medidas de diagnóstico de influência. Os métodos de diagnóstico têm sido ferramentas importantes na análise de regressão para detectar anomalias, tais como quebra das pressuposições nos erros, presença de outliers e observações influentes. Em seguida propomos o modelo de fração de cura com fragilidade Birnbaum-Saunders. Os modelos para dados de sobrevivência com proporção de curados (também conhecidos como modelos de taxa de cura ou modelos de sobrevivência com longa duração) têm sido amplamente estudados. Uma vantagem importante do modelo proposto é a possibilidade de considerar conjuntamente a heterogeneidade entre os pacientes por suas fragilidades e a presença de uma fração curada. As estimativas dos parâmetros do modelo foram obtidas via máxima verossimilhança, medidas de influência e diagnóstico foram desenvolvidas para o modelo proposto. Por fim, avaliamos a distribuição bivariada Birnbaum-Saunders baseada na média, como também introduzimos um modelo de regressão para o modelo proposto. Utilizamos os métodos de máxima verossimilhança e método dos momentos modificados, para estimar os parâmetros do modelo. Avaliamos o desempenho dos estimadores via estudo de simulações de Monte Carlo. Aplicações a conjuntos de dados reais ilustram as potencialidades dos modelos abordados.

Palavras-chave: Análise de diagnóstico, Distribuição Birnbaum-Saunders, Estimação de máxima verossimilhança, Modelos de fragilidade, Modelos de fração de cura.

ABSTRACT

LEÃO, J. S.. **Modeling based on a reparameterized Birnbaum-Saunders distribution for analysis of survival data**. 2017. 117 f. Doctoral dissertation (Doctorate Candidate joint Graduate Program in Statistics DEs-UFSCar/ICMC-USP) – Instituto de Ciências Matemáticas e de Computação (ICMC/USP), São Carlos – SP.

In this thesis we propose models based on a reparameterized Birnbaum-Saunder (BS) distribution introduced by Santos-Neto et al. (2012) and Santos-Neto et al. (2014), to analyze survival data. Initially we introduce the Birnbaum-Saunders frailty model where we analyze the cases (i) with (ii) without covariates. Survival models with frailty are used when further information is non-available to explain the occurrence time of a medical event. The random effect is the "frailty", which is introduced on the baseline hazard rate to control the unobservable heterogeneity of the patients. We use the maximum likelihood method to estimate the model parameters. We evaluate the performance of the estimators under different percentage of censured observations by a Monte Carlo study. Furthermore, we introduce a Birnbaum-Saunders regression frailty model where the maximum likelihood estimation of the model parameters with censored data as well as influence diagnostics for the new regression model are investigated. In the following we propose a cure rate Birnbaum-Saunders frailty model. An important advantage of this proposed model is the possibility to jointly consider the heterogeneity among patients by their frailties and the presence of a cured fraction of them. We consider likelihood-based methods to estimate the model parameters and to derive influence diagnostics for the model. In addition, we introduce a bivariate Birnbaum-Saunders distribution based on a parameterization of the Birnbaum-Saunders which has the mean as one of its parameters. We discuss the maximum likelihood estimation of the model parameters and show that these estimators can be obtained by solving non-linear equations. We then derive a regression model based on the proposed bivariate Birnbaum-Saunders distribution, which permits us to model data in their original scale. A simulation study is carried out to evaluate the performance of the maximum likelihood estimators. Finally, examples with real-data are performed to illustrate all the models proposed here.

Keywords: Birnbaum-Saunders distribution, Cure rate model, Diagnostic analysis, Frailty model, Likelihood estimation.

LIST OF FIGURES

LIST OF ALGORITHMS

LIST OF TABLES

CONTENTS

CHAPTER

1

INTRODUCTION

1.1 Introduction and bibliographical review

Analysis of lifetime data plays an important role in several fields of knowledge such as economics, biology, medicine, epidemiology, engineering, demography, among others. This area has been widely studied by many researchers, with works having been published in various areas of knowledge. In this sense we can make some questions like: What distinguishes survival analysis from other areas of statistics? Why do survival data need a special statistical theory? The reason is that we are observing something that develops dynamically over time. Thus we can highlight two points related to this development. The first, survival times are usually a mixture of discrete and continuous data that lend themselves to a different type of analysis than in the traditional discrete or continuous case. The mixture is the result of censoring and has an important effect on data analysis. To put it plainly, a censored observation contains only partial information about the random variable (RV) of interest. The Kaplan-Meier estimator, proposed by Kaplan and Meier (1958), of the survival function (SF) is a major step in the development of suitable models for such kind of data. The second is because most of the evaluations are made conditionally on what is known at the time of the analysis, and this changes over time. Frequently, as the population under study is changing, we only consider the individual risk to die for those who are still alive, but this means that many standard statistical approaches cannot be applied.

Cox (1972) proposed the proportional hazards model and since then this model has been widely used in survival analysis. One of the main reasons for this is because of the ease with which technical difficulties such as censoring and truncation are handled. This is due to the appealing interpretation of the hazard rate (HR) as a risk that changes over time. Naturally, the concept allows for the entering of covariates in order to describe their influence and to model different levels of risk for different subgroups which can be considered in modeling. However, in general, it is impossible to include all relevant risk factors, perhaps because we have no

information on individual values, which is often the case in demography. Furthermore, we may not know all relevant risk factors or it is impossible to measure them without great financial costs, something that is common in medical and biological studies. The neglect of covariates leads to (unobserved) heterogeneity. That is, the population consists of individuals with different risks. Another important regression model proposed in survival analysis was the accelerated failure time (AFT) model, see details in Lawless (2011). The Cox's model and its various generalizations are mainly used in medical and biostatistical fields, while the AFT model is primarily applied in reliability theory and industrial experiments.

In this thesis we focus on frailty and cure rate frailty models as well as on a bivariate model, where all these models are based on the reparameterized Birnbaum-Saunders distribution proposed by Santos-Neto et al. (2012). In the context of frailty modeling the frailty indicates that apparently similar patients can have different risks. Thus, different patients can possess distinct frailties and then frailer patients tend to experience the event of interest earlier than those who are less frail. Therefore, frailty models have been introduced into the statistical literature in an attempt to account for the existence of heterogeneity in a population under study. In essence this concept goes back to the work of Greenwood and Yule (1920) on "accident proneness". The RV "frailty" may be incorporated in the baseline HR additively or multiplicatively. The term frailty itself was introduced by Vaupel et al. (1979) in univariate frailty model. Several authors have studied these models, which represent a generalization of the Cox model; see Cox (1972) and Stare and O'Quigley (2004). The interested reader in frailty models is referred to Hougaard (2000), Duchateau and Janssen (2008) and Wienke (2011).

Some other studies about frailty models are the following. Aalen and Tretli (1999) studied incidence of testicular cancer with frailty models. Fan and Li (2002) considered variable selection in frailty models. Tomazella (2003) studied the problem as homogeneous and heterogeneous Poisson process with a frailty term. Androulakis et al. (2012) extended the gamma (GA) frailty model methodology proposed by Fan and Li (2002) to the uniform frailty model. Mallick and Ravishanker (2006) studied an additive stable frailty model for multivariate times to event data. Mallick et al. (2008) introduced bivariate positive stable considering dependent multivariate times-to-events with a Weibull baseline hazard. Barker and Henderson (2005) adapted the standard expectation maximization algorithm and analyzed its behavior in the univariate GA frailty model. As the frailty component of the model is random, a distribution can be assumed for it. Yu (2008) introduced a frailty into the mixture cure model to study recurrent event data with a cured fraction. Cai (2010) studied Bayesian semi-parametric frailty selection in multivariate event time data. Martinussen et al. (2011) describe an innovative approach to estimation in the Aalen additive GA frailty hazards model. Mazroui et al. (2013) proposed a multivariate frailty model that jointly analyzes two types of recurrent events with a dependent terminal event. Liu et al. (2014) proposed an accelerated intensity frailty model for recurrent events data and derived a test for the frailty variance. Enki et al. (2014) studied a new parametric time-varying shared frailty model to represent changes over time in population heterogeneity, for use with bivariate

current status data. Chan (2014) introduced a flexible individual frailty model for clustered right-censored data, in which covariate effects can be marginally interpreted as log failure odds ratios. Zhou et al. (2015) motivated by breast cancer data proposed a covariate-adjusted proportional hazards frailty model for the analysis of clustered right-censored data. Inspired by frailty-contagion approaches used in finance and insurance, Koch and Naveau (2015) proposed a multi-site precipitation simulator that, given appropriate regional atmospheric variables, can simultaneously handle dry events and heavy rainfall periods.

In general the methods in survival analysis implicitly assume that populations are homogeneous, meaning all individuals have the same risk of death, but as mentioned above, it is often important to consider the population as heterogeneous, i.e. a mixture of individuals with different hazards, for example. The frailty model is a random effects model for time-to-event data, where the frailty has a multiplicative effect on the baseline hazard function. It can be used for univariate (independent) lifetimes, i.e. to describe the influence of unobserved covariates in a proportional hazards model (heterogeneity). The variability of duration data is split into one part that depends on risk factors and is thus theoretically predictable, and one part that is initially unpredictable, even knowing all relevant information at that time. There are advantages in separating these sources of variability: heterogeneity can explain some unexpected results or give an alternative interpretation, for example crossing-over or levelling-off effects of HR. The introduction of a common random effect "the frailty" is a natural way of modeling the dependence of event times. The random effect explains the dependence in the sense that had we known the frailty, the events would have been independent. In other words, the lifetimes are conditionally independent given the frailty. This approach can be used for survival times of related individuals such as twins or family members, where independence cannot be assumed, or for recurrent events in the same individual or for times to several events for the same individual, such as onset of different diseases, relapse or death (competing risks).

Due to the randomness of the frailty term of a model, it is necessary to assume a distribution for it, called frailty distribution. Due to the way how the frailty term acts on the HR, natural candidates to the frailty distribution are the GA, inverse Gaussian (IG), lognormal (LN) and Weibull models; see Hougaard (1995). Particularly, since the seminal work presented by Vaupel et al. (1979), the GA frailty distribution has been used in most applications published up to date; see Balakrishnan and Peng (2006). However, compared with standard random effect models, frailty models pose additional difficulties in developing inferential methods, caused by incompleteness of data due to censoring and truncation and by the requirement for a specification of a baseline hazard (or a non-parametric baseline hazard). A good alternative to the GA distribution is the BS distribution. It has been widely considered in the literature due to its physical arguments, its attractive properties and its relationship with the normal distribution. The BS model was proposed by Birnbaum and Saunders (1969) and has been extensively applied for modeling failure times in engineering, although some novel applications have been considered in biological, environmental and financial sciences; see, for example, Desmond (1985), Kotz

et al. (2010a), Saulo et al. (2013) and Leiva et al. (2014b), Leiva et al. (2014c), Leiva et al. (2015), Leiva et al. (2015), Leiva et al. (2017). Some details about genesis and justification of the BS distribution for medical data are given in Chapter 2. Santos-Neto et al. (2012) introduced several parameterizations of the BS distribution. Specially, one of them is established in terms of the distribution mean, whereas its variance is a quadratic function of this mean. Thus, such a parameterization allows us to mimic a property of the GA frailty distribution early proposed by Vaupel et al. (1979), doing the BS frailty distribution to be a new alternative to frailty modeling.

Another important area of suvival analysis is related to cure rate models (or long-term survival models). The first work in this context was proposed by Boag (1949), Berkson and Gage (1952) and refers to the mixture cure model. In this case, the population is classified into the following two subpopulations: (a) individuals who are cured with certain probability and (b) individuals exposed to risk characterized by the complementary probability, which can be estimated by using a probability distribution, for instance, exponential, Gompertz or Weibull; see Kuk and Chen (1992), Koti (2003) and Shao and Zhou (2004). A new approach for this model was presented by Yakovlev and Tsodikov (1996) and Chen et al. (1999) and refers to the non-mixture cure model, which has its structure based on the assumption that the cumulative hazard function is bounded because of the existence of cured individuals.

Rodrigues et al. (2009) proposed an unified approach to long-term survival models. Several authors have been proposed cure models using this methodology. Some works are the following. Castro et al. (2009) studied a cure rate model where the number of competing causes of the event of interest follows the negative binomial distribution in a bayesian framework. Castro et al. (2010) applied the generalized additive models for location, scale, and shape (GAMLSS) framework to the fitting of long-term survival model. Cancho et al. (2012) proposed a cure rate survival model by assuming the number of competing causes to be Geometric distributed and the time to event of interest following a BS distribution. Cancho et al. (2011) presented a flexible cure rate model in a Bayesian approach. Eudes et al. (2013) studied the mixture model assuming the modified Weibull distribution. Louzada et al. (2014) introduce a Bayesian partition modeling for lifetime data in the presence of a cure fraction by considering a local structure generated by a tessellation. This modeling is based on a promotion time cure model structure but assuming that the number of competing causes follows a geometric distribution. Rodrigues et al. (2015) introduced a relaxed cure rate model as a natural and less restrictive extension of the popular Poisson cure rate model at the cost of an additional parameter. Cordeiro et al. (2016) proposed a cure rate survival model by assuming that the number of competing causes of the event of interest follows the negative binomial distribution and the time to the event of interest has the Birnbaum-Saunders distribution. Suzuki et al. (2016) studied a new survival model, called Poisson Inverse-Gaussian regression cure rate model, which enables different underlying activation mechanisms that lead to the event of interest. Yiqi et al. (2016) developed a Bayesian approach for the Weibull-Negative-Binomial regression model with cure rate under latent failure causes and presence of randomized activation mechanisms.

Cure rate models allow us to estimate separate covariate effects that may influence the cured fraction and the hazard of the at-risk population. If a cured fraction is not present, the analysis reduces to the standard considerations of survival data. Further details and examples of cure models and their utility are provided in Maller and Zhou (1996), Ibrahim et al. (2005) and Aalen et al. (2008). Cure models assume that the individuals experiencing the event of interest are homogeneous. However, this assumption may not be valid as unobserved heterogeneity among individuals may be present. In this sense, cure data can be analyzed utilizing statistical models that account for heterogeneity among individuals. A portion of the heterogeneity is explainable in terms of observed covariates. However, there remains a degree of heterogeneity induced by unobservable risk factors. Failing to account for this latter form of heterogeneity may lead to distorted results.

Several authors studied extensions of frailty models to account for a cured component. Longini and Halloran (1996) derived a statistical model for estimating vaccine efficacy that expresses the often unmeasured heterogeneous host response. Price and Manatunga (2001) studied survival data with cure frailty model, where the authors consider the GA frailty mixture and compound Poisson distribution. Yin (2005) proposed a Bayesian approach to model correlated or clustered failure time data incorporating a surviving fraction, considering two forms of cure rate frailty models. Wienke et al. (2006) analyzed three correlated frailty models for bivariate time-to-event data, where is assumed GA, lognormal and compound Poisson distributions. Yu (2008) studied the inclusion of frailty into the mixture cure model to model recurrent event data with a cure fraction. Peng and Zhang (2008b) proposed a mixture cure frailty model that generalizes the general mixture cure model by adding a frailty term in the latency distribution and investigated the identifiability of the mixture cure model and the frailty model. Rahimzadeh et al. (2011) develope a Bayesian approach for the estimation of two cure correlated frailty models that have been extended to the cure frailty models introduced by Yin (2005). Rondeau et al. (2013) studied a cure frailty models for survival data with recurrences for breast cancer and colorectal cancer. Gonzales et al. (2013) introduced the GA frailty mixture regression model. Calsavara et al. (2016) proposed a flexible cure rate model which is an extension of Cancho et al. (2011) model by incorporating a power variance function (PVF) frailty term in latent risk.

The use of bivariate distributions plays a fundamental role in many areas of knowledge, for example, in survival analysis and reliability. Here, we studied a bivariate Birnbaum-Saunders distribution parameterized by its means. The bivariate BS distribution was proposed by Kundu et al. (2010), where some other author have been extended this model as well as evaluate some properties of the model, for example Khosravi et al. (2015), Kundu et al. (2010), Kocherlakota (1986), Kundu et al. (2013), Díaz-Garcia and Leiva (2005), Vilca et al. (2014a), Vilca et al. (2014b),

1.2 Objectives of the thesis

The BS distribution has been receiving considerable attention due to its good properties. Santos-Neto et al. (2012) introduced several parameterizations for the BS distribution, where one of these reparameterizations indexes the BS distribution by its mean. Santos-Neto et al. (2014) present some mathematical properties and estimates by the maximum likelihood method, Moments and Modified moments method of the parameters from this new version of the BS distribution. According to this new BS model our gerenal objective is to study the BS frailty model. However, we can list some specific objetives

- to propose a new BS frailty model, which can be a good alternative to frailty modeling.
- to introduce a BS frailty regression model and its inference based on maximum likelihood (ML) method, and to derive influence diagnostics tools for this model. In addition, we want to apply the BS frailty regression model and its diagnostics to medical data to illustrate its potential applications and compare it with classical frailty models.
- to propose a cure rate Birnbaum-Saunders frailty model, where an important advantage of the proposed model is the possibility to jointly consider the heterogeneity among patients by their frailties and the presence of a cured fraction of them.
- to introduce a bivariate Birnbaum-Saunders distribution based on a parameterization given by Santos-Neto et al. (2012) and Santos-Neto et al. (2014) which has the mean as one of its parameters.

1.3 Organization of the chapters

The whole thesis is written with independent chapters and each chapter has new research contribution. The chapters are related in the sense that they talk about related areas involving frailty models and the Birnbaum-Saunders distribution. Chapter 2 presents briefly some concepts related to the genesis of the Birnbaum-Saunders distribution and the reparameterized version used in this work, as well as some features of the frailty and cure rate models. In Chapter 3, we

introduce the BS frailty model, which can be a good alternative to frailty modeling. We employ the Laplace transform to find the BS unconditional SF on the individual frailty. We use the ML method for estimating the model parameters. We investigate the asymptotic properties of the ML estimators and evaluate their performance by a Monte Carlo (MC) study. We illustrate the proposed model with uncensored and censored data. In Chapter 4, we present the BS frailty regression model and its inference based on ML methods, and to derive influence diagnostics tools for this model. In addition, we want to apply the BS frailty regression model and its diagnostics to medical data to illustrate its potential applications and compare it with classical frailty models. In Chapter 5, we propose a cure rate frailty model based on the Birnbaum-Saunders distribution as an alternative approach to modeling such data. An important advantage of the proposed cure rate frailty model is the possibility to jointly consider the heterogeneity among individuals and the presence of a cured component. We consider the ML method to estimate the model parameters and to derive influence tools. We assess local influence on the parameter estimates under different perturbation schemes. Numerical evaluation of the proposed model is considered by means of MC simulation studies and an application to a real medical data set from the medical area. In Chapter 6, we introduce a bivariate Birnbaum-Saunders distribution based on a parameterization of the Birnbaum-Saunders which has the mean as one of its parameters. We discuss the ML estimation of the model parameters and show that these estimators can be obtained by solving non-linear equations. We also discuss modified moment (MM) estimation for the unknown parameters which are easy to compute and can therefore be used as initial values to calculate the ML estimates. We derive the asymptotic distributions of these estimators and carry out a simulation study to evaluate the performance of all these estimators. The probability coverages of confidence intervals are also discussed. We then derive a regression model based on the proposed bivariate Birnbaum-Saunders distribution, which permits us to model data in their original scale. In addition, two examples are performed to illustrate the proposed methods here. Finally, we present a discussion, conclusions and future research in Chapter 7.

1.4 Products of the thesis

This thesis allowed the following products to be obtained:

- Leao, J., Leiva, V., Tomazella, V., Saulo H. (2017) Birnbaum-Saunders frailty regression models: Diagnostics and application to medical data. *Biometrical Journal* (in press);
- Leao, J., Leiva, V., Tomazella, V., Saulo H. (2016) A Birnbaum-Saunders frailty model for survival data (under review for *Brazilian Journal of Probability and Statistics*).
- Leao, J., Leiva, V., Tomazella, V., Saulo H. (2016) A cure rate frailty model based on the reparameterized Birnbaum-Saunders distribution (submitted).

- Saulo, H., Leao, J., Leiva, V., Tomazella, V. (2016) On a bivariate Birnbaum-Saunders distribution parameterized by its means (submitted).

- Leao, J., Leiva, V., Tomazella, V., Saulo H. (2015) Birnbaum-Saunders frailty regression models: Diagnostics and application to medical data. Awarded as outstanding presentation for postgraduate students, Univesidad Adolfo Ibáñez - Second International Workshop on "Statistical Models for Business, Engineering and Sciences".

CHAPTER

2

BACKGROUND

2.1 Introduction

In this chapter, we present some features of the BS distribution beginning with its genesis. Then, we justify the use of the BS model in medical data. We also describe briefly the parameterization used in the course of this thesis. We describe breifly the bivariate BS distribution in its original form. The interested reader in BS distribution is referred to Leiva (2016) and references therein. We present the frailty model, discuss how to obtain the unconditional HR and SF. Moreover, we discuss unified cure rate model and present some of its features; see Rodrigues et al. (2009).

2.2 Birnbaum-Saunders distribution

The Birnbaum-Saunders distribution is right-skewed (asymmetrical), continuous and unimodal. It is also known as the fatigue life distribution and has received considerable attention due to its theoretical arguments, its attractive properties and its relation with the normal distribution; see the seminal paper by Birnbaum and Saunders (1969). As justification the authors used a physical argument originated from renewal theory, via idealization of the number of cycles necessary to force a fatigue crack to grow past a critical value; see, for example, Mann et al. (1974).

As a fatigue life distribution, the BS model considers a material specimen that is exposed to a sequence of m cyclic loads, $\{l_i, i = 1,2,\ldots,m, m \in \mathbb{N}\}$; for more details about this type of load; see Saunders (2007). The loading scheme can be depicted as follows:

$$\underbrace{l_1,\ldots,l_m}_{\text{Cycle 1}}\ \underbrace{l_{m+1},\ldots,l_{2m}}_{\text{Cycle 2}}\ \cdots\ \underbrace{l_{jm+1},\ldots,l_{jm+m}}_{\text{Cycle (j+1)}},$$

where $l_{jm+i} = l_{km+i}$, for $j \neq k$. Birnbaum and Saunders (1969) considered that the loading is continuous (see more details in Section 2.1), which implies that the load function, say $l_i(\cdot)$, evaluated at the unit interval gives the amount of stress imposed on the specimen, that is,

$$l_{i-1}(0) = l_i(1) = l_{i+1}(0),\ i = 1,\ldots,m,\ m \in \mathbb{N}.$$

Thus, at the imposition of each load, l_i, the crack is extended by a random amount. Crack extensions by cycle cannot be observed in practice and we only know the instant when the failure occurs. Having explained the physical framework of the genesis of the Birnbaum-Saunders distributions it is now necessary to make the statistical assumptions. Birnbaum and Saunders (1969) used the knowledge of certain type of materials failure due to fatigue to develop their model. The fatigue process that they used was based on the following:

(D1) A material specimen is subjected to cyclic loads or repetitive shocks, which produce a crack or wear-out in this specimen;

(D2) The failure occurs when the size of the crack in the material specimen exceeds certain level of resistance (threshold), denoted by ω,

(D3) The sequence of loads imposed in the material specimen is the same from a cycle to another one;

(D4) The incremental crack extension due to a load l_i, say X_i, during the jth cycle is a RV whose distribution is governed by all the loads l_j, for $j < i$, and by the actual crack extensions that have preceded it in cycle alone;

(D5) The total size of the crack due to the jth cycle, say Y_i, is a RV that follows a statistical distribution of mean μ_0 and variance σ_0^2, and

(D6) The sizes of cracks in different cycles are mutually independent. Note that the total crack extension due to the $(j+1)$th cycle of load is

$$Y_{j+1} = X_{jm+1} + X_{jm+2} + \cdots + X_{jm+m};\ j,m = 0,1,2,\ldots.$$

As mentioned by Mann et al. (1974), assumption (D4) is rather restrictive and may not be valid for certain applications. This assumption ensures that, regardless of the dependence among the successive random extensions due to the loads in a particular cycle, the total random crack extensions are independent from cycle to cycle. The plausibility of this assumption in aeronautical fatigue studies is briefly stated by Birnbaum and Saunders (1969).

The BS model looks for the distribution of the smallest n, say n^*, such that the sum

$$S_n = \sum_{j=1}^{n} Y_j, \tag{2.1}$$

of n positive RV exceeds the given threshold ω, that is,

$$n^* = \inf\{n \in \mathbb{N}; S_n = \sum_{j=1}^{n} Y_j > \omega\}.$$

In the simplest case, the BS distribution is derived by supposing that the Y_j are independent and identically distributed RV, then applying the central limit theorem and then by regarding n^* as a continuous RV T.

Specifically, based on the central limit theorem, Equation (2.1), and the assumptions (D5) and (D6) made by Birnbaum and Saunders (1969), as $n \to \infty$, it is possible to establish that

$$S_n \dot{\sim} N(n\mu_0, n\sigma_0^2). \tag{2.2}$$

Let N be the number of required cycles until the failure. Given that $Y_j > 0$ for all $j \geq 1$, the damage is irreversible and so by complementarity, we have $\{N > n\} \equiv \{S_n \leq \omega\}$ and we have $\{N \leq n\} \equiv \{S_n > \omega\}$. Then, from assumption (D5) and and (2.1), we have that $\mathrm{E}(S_n) = n\mu_0$ and $\mathrm{Var}(S_n) = n\sigma_0^2$. Therefore, by standardizing (2.2), we get

$$\begin{aligned} P(N \leq n) &\approx P\left(\frac{S_n - n\mu_0}{\sigma_0\sqrt{n}} > \frac{\omega - n\mu_0}{\sigma_0\sqrt{n}}\right) \\ &= P\left(\frac{S_n - n\mu_0}{\sigma_0\sqrt{n}} \leq \frac{n\mu_0 - \omega}{\sigma_0\sqrt{n}}\right) \\ &= \Phi\left(\frac{n\mu_0 - \omega}{\sigma_0\sqrt{n}}\right) \\ &= \Phi\left(\frac{\sqrt{\omega\mu_0}}{\sigma_0}\left[\frac{n}{\omega/\mu_0} - \frac{\omega/\mu_0}{n}\right]\right). \end{aligned} \tag{2.3}$$

Birnbaum and Saunders (1969) used Equation (2.3) to define a continuous life distribution, idealizing the discrete variate N through a continuous variate T and the discrete argument n by means of the continuous t, that is, the number of cycles until the failure, N , is replaced by the total time until that the failure occurs, T, and nth cycle by the time t. Thus, taking

$$\alpha = \frac{\sigma_0}{\sqrt{\omega\mu_0}} \text{ and } \beta = \frac{\omega}{\mu_0},$$

and

$$a_t(\alpha, \beta) = a_t = \frac{1}{\alpha}\left[\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right], \tag{2.4}$$

we obtain that

$$F_T(t) = \Phi\left(a_t(\alpha, \beta)\right),\ t > 0, \alpha > 0, \beta > 0,$$

which is the cumulative distribution function (CDF) of the BS distribution with shape and scale parameters, α and β, respectively. This means that we are admitting as definition that a random variable T follows the BS distribution with shape and scale parameters $\alpha > 0$ and $\beta > 0$, respectively, if it can be written as

$$T = \beta \left[\frac{\alpha}{2} Z + \sqrt{\left(\frac{\alpha}{2} Z\right)^2 + 1} \right], \tag{2.5}$$

where Z is a random variable following the standard normal distribution, such that

$$Z = \frac{1}{\alpha} \left[\sqrt{\frac{T}{\beta} - \sqrt{\frac{\beta}{T}}} \right] \sim \mathrm{N}(0,1) \tag{2.6}$$

The derivation of Birnbaum and Saunders (1969) supposes some aspects about the growth of a crack that are questionable. Desmond (1985) gave a more general derivation for this distribution as well as derived the BS distribution using a biological model discussed by Cramér (1947). In the next subsection we present some of these arguments presented in Desmond (1985) that justify the use of the BS model in medical data.

2.3 Justifying the BS distribution for medical data

Assuming a BS distribution to model medical data based on an empirical fitting can be a reasonable argument. However, the argument may be strengthened if we justify why the BS distribution might be suitable for such a modeling. Cramér's biological model, linked to the proportionate-effect model, allows us to justify the BS distribution within a medical setting.

Consider a RV related to the size of a human organ. The size may be considered as the joint effect of a large number of independent causes. The causes act sequentially through the time of organ growth. If the effects of causes are summed and assumed as RVs, then the sum follows an asymptotic normal distribution due to the central limit theorem. However, the causes do not seem to jointly operate by simple addition. It seems more natural to assume that each cause provides an impulse. Thus, the effect depends on both the impulse strength and the organ size attained at the instant when the impulse is working. Specifically, let $Y_1, \ldots, Y_n$ be independent RVs corresponding to the magnitude of n impulses, which act sequentially according to their sub-indices. In addition, let X_j be the size of an organ, which is produced by the impulses. Assume that X_{j+1} increases proportionally to the $(j+1)$th impulse Y_{j+1} and to some function $g(X_j)$ of the organ size as

$$X_{j+1} = X_j + Y_{j+1}\, g(X_j), \quad j = 0, 1, \ldots. \tag{2.7}$$

Therefore, X_{j+1} is the accumulated size of the organ after application of the impulse Y_{j+1}. From (2.7), the LN distribution can be obtained when $g(X) = X$, whereas the BS distribution

can be obtained when $g(X) = 1$. Relationship defined in (2.7) has been proposed in biological and fatigue contexts; see Desmond (1985). The BS distribution was built to model fatigue life of material specimens subject to cyclic stress. It provokes a damage that is accumulated over time by a sum of numerous small damages. When the damage exceeds a rupture threshold of the specimen, it fails; see Birnbaum and Saunders (1969). Table 1 provides a conceptual analogy between fatigue and medical settings.

Table 1 – Conceptual analogy between material fatigue and organ growth.

Concept / Process	Specimen	Cause	Threshold	Effect	RV
Fatigue	Material	Damage	Rupture	Failure	Fatigue life
Growth	Organ	Impulse	To die	Death	Time of death

We put Frost & Dugdale's model, often used in engineering; see Frost and Dugdale (1958), in a medical setting by

$$\frac{\mathrm{d}a}{\mathrm{d}n} = \frac{S^3 a}{c} = c_1 f(\Delta K), \tag{2.8}$$

where a is the organ size, n the number of impulses, S the strength applied in each impulse, c a constant, ΔK the range of a strength intensity factor, $f(\cdot)$ an empirically determined function and c_1 an experimental constant. One can relate ΔK in (2.8) to the organ size a by means of

$$\Delta K = b\Delta S a^{1/2}, \tag{2.9}$$

where b is a geometrically related parameter and ΔS the strength amplitude applied in each impulse. Based on (2.8) and (2.9), and approximating $f(\cdot)$ by $f(b\Delta S a^{1/2}) \approx c_0 + c_2 g(a)$, with $g(\cdot)$ being a function of the organ size, we have

$$\frac{\mathrm{d}a}{\mathrm{d}n} \approx c_0 + c_3\, g(a), \tag{2.10}$$

where c_3 contains constants c_1, c_2, and c_0 is the initial organ size, with c_0, a and c_3 being considered as RVs to make it closer to reality.

Note the similarity between differential-equation model (2.10) and proportionate-effect model (2.7). Retake (2.7), apply the central limit theorem and consider the increment $\Delta X_j = X_{j+1} - X_j$ in the $(j+1)$th impulse gives a small contribution to the organ growth. Then, summation can be changed by integration. Thus, we get

$$\sum_{j=1}^{n} Y_j = \sum_{j=1}^{n} \frac{\Delta X_j}{g(X_j)} \approx \int_{X_0}^{X_n} \frac{\mathrm{d}x}{g(x)} = \log(g(X_n)) - \log(g(X_0)),$$

follows approximately a normal distribution, where X_0 is the initial organ size and X_n its final size.

To obtain the BS distribution in this setting, suppose that the mean of X_j is η and its variance ρ^2. This generalizes Assumption 2 of Birnbaum and Saunders (1969) and conducts to

$$I(X(t)) = \int_{X_0}^{X(t)} \frac{1}{g(x)}\,\mathrm{d}x \sim \mathrm{N}(t\eta, t\rho^2), \tag{2.11}$$

where $X(t)$ is the organ size at time t. Assume now that $X_c > X_0$ is a critical organ size at which death occurs. Then, $T = \inf\{t\colon X(t) > X_c\}$ is the time of death. Therefore, from (2.11) and using the equivalent events $\{T \le t\}$ and $\{X(t) > X_c\}$, it follows that the CDF of T is

$$F_T(t) = \Phi((t\eta - I(X_c))/\sqrt{t}\rho), \tag{2.12}$$

where Φ is the CDF of the standard normal distribution. From (2.12), note that choice of the function $g(X)$ in the model given in (2.7) determines the dependence of the organ size on the previous size. A power function for $g(X)$ could be a reasonable choice depending on the type of organ. Suppose that $g(X) = X^{\delta}$, where δ could be a parameter related to the type of organ. In this case, from (2.11), note that $X(t)^{1-\delta} \sim \mathrm{N}(X_0^{1-\delta} + [1-\delta]t\eta, [1-\delta]^2 t\rho^2)$ and then the CDF of T is

$$F_T(t) = \begin{cases} \Phi\left(\frac{X_c^{1-\delta} - X_0^{1-\delta} + [\delta-1]t\eta}{[\delta-1]\sqrt{t}\rho}\right), & \text{if } \delta > 1; \\ \Phi\left(\frac{X_0^{1-\delta} - X_c^{1-\delta} + [1-\delta]t\eta}{[1-\delta]\sqrt{t}\rho}\right), & \text{if } \delta < 1. \end{cases} \tag{2.13}$$

Therefore, the LN distribution is obtained from (2.13) by letting $\delta \to 1$, whereas the BS distribution results for $\delta = 0$. However, although the case $\delta = 1$ corresponds to the proportionate-effect model, the life distribution itself is of the BS type and not of LN type, see Desmond (1985).

2.4 Parametrizations of the BS distribution

Santos-Neto et al. (2012) proposed several parameterizations of the BS distribution, which allow diverse features of data modeling to be considered. One of such parameterizations is indexed by the parameters $\mu = \beta(1+\alpha^2/2)$ and $\delta = 2/\alpha^2$, where $\alpha > 0$ and $\beta > 0$ are the original BS parameters; see Birnbaum and Saunders (1969), $\mu > 0$ is a scale parameter and the mean of the distribution, whereas $\delta > 0$ is a shape and precision parameter. The notation $U \sim \mathrm{BS}(\mu, \delta)$ is used when the RV U follows such a distribution. This parameterization of the BS distribution permits us to mimic a property of the GA distribution, which was the first distribution used in a frailty model Vaupel et al. (1979) as follows. The mean and variance of $U \sim \mathrm{BS}(\mu, \delta)$ are $\mathrm{E}[U] = \mu$ and $\mathrm{Var}[U] = \mu^2/\phi$, respectively, where $\phi = (\delta+1)^2/(2\delta+5)$. Then, as mentioned, δ can be interpreted as a precision parameter, that is, for fixed values of μ, when $\delta \to \infty$, the variance of T tends to zero. Also, for fixed μ, if $\delta \to 0$, then $\mathrm{Var}[U] \to 5\mu^2$. Note that $\mathrm{Var}[U] = \mu^2/\phi$ is similar to the variance function of the GA distribution, which has a quadratic relation with its mean. Therefore, a frailty model based on the BS distribution, in its reparameterized form, can be a good alternative to the GA frailty model.

If $U \sim \text{BS}(\mu, \delta)$, then its probability density function (PDF) is

$$f_U(u;\mu,\delta) = \frac{\exp(\delta/2)\sqrt{\delta+1}}{4u^{\frac{3}{2}}\sqrt{\pi\mu}}\left(u+\frac{\delta\mu}{\delta+1}\right)\exp\left(-\frac{\delta}{4}\left(\frac{u(\delta+1)}{\delta\mu}+\frac{\delta\mu}{u(\delta+1)}\right)\right), \quad u>0. \tag{2.14}$$

It is possible to show that $kU \sim \text{BS}(k\mu, \delta)$, with $k > 0$, and $1/U \sim \text{BS}(\mu^*, \delta)$, where $\mu^* = (\delta+1)/(\delta\mu)$, that is, the BS distribution, in its original and reparameterized forms, is closed under scaling and reciprocation. From (2.14), the SF and HR of U are, respectively,

$$\begin{aligned} S_U(u;\mu,\delta) &= \frac{1}{2}\Phi(u+\delta(u-\mu)/(2\sqrt{u(1+\delta)\mu})), \quad u>0, \\ h_U(u;\mu,\delta) &= \frac{\exp(-(-\delta\mu+\delta u+u)^2/(4(\delta+1)\mu u))(\delta\mu+\delta u+u)}{(\pi\mu(\delta+1))^{\frac{1}{2}}2\mu^{\frac{1}{2}}u^{\frac{3}{2}}\Phi((u+\delta(u-\mu))/(2\sqrt{u(1+\delta)\mu}))}, \quad u>0, \end{aligned}$$

where Φ is the CDF of the N(0, 1) distribution. Figure 1 displays some shapes for the PDF, SF and HR of $U \sim \text{BS}(\mu = 1, \delta)$. Note that a unimodal behavior is detected for the PDF, as well as different degrees of asymmetry and kurtosis, whereas the HR has increasing and decreasing shapes, such as the GA distribution, but also an inverse bathtub shape.

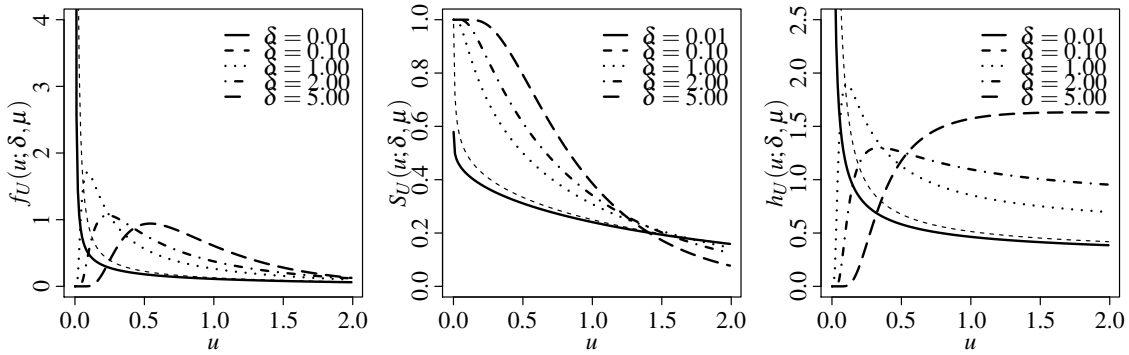

Figure 1 – Plots of PDF, SF and HR of the BS distribution for $\mu = 1$ and different values of δ.

The use of the BS distribution has the following appealing advantages:

- Based on its genesis, it is possible to make an analogy in the modeling of medical data; see, for example, Desmond (1985).
- Its parameterization based on the mean (μ), such as in (2.14), allows us to analyze data in their original scale, avoiding, for instance, problems of interpretation in models which employ a logarithmic transformation of the data; see Leiva et al. (2014a) and Santos-Neto et al. (2014).
- In the context of frailty models, it can be very competitive in terms of fitting.
- It belongs to the class of log-symmetric distributions, such as the case of the generalized BS, LN, log-logistic, log-Laplace, log-Student-t, log-power-exponential, log-slash and F distributions; see Vanegas and Paula (2016a) and Vanegas and Paula (2016b). The log-symmetric class of distributions arises when an RV has the same distribution as its

reciprocal or as ordinary symmetry of the distribution of the logged RV; see Jones (2008). One can obtain the BS, LN and log-logistic frailty models as particular cases of log-symmetric frailty models. However, besides the BS frailty model which is proposed in this thesis, the only other popular log-symmetric frailty model that belongs to this class is the LN one. But unlike the BS frailty model, the LN model does not have an explicit Laplace transform; see Wienke (2011). This explicit form is useful to obtain the PDF and the unconditional SF and HR.

- It is flexible in terms of bimodality when the logarithm of a BS RV is taken into account. Note that:

  (i) If $U \sim \text{BS}(1,\delta)$, $Y = \log(U) \sim \text{log-BS}(\sqrt{2/\delta}, \log(\delta/(\delta+1)))$; see Rieck and Nedelman (1991);

  (ii) If $U \sim \text{LN}(1,\sigma^2)$, $Y = \log(U) \sim \text{N}(1,\sigma^2)$; see Crow and Shimizu (1988);

  (iii) If $U \sim \text{IG}(1,\sigma^2)$, $Y = \log(U) \sim \text{log-IG}(\sigma^2, 0)$; see Kotz et al. (2010b);

  (iv) If $U \sim \text{GA}(1/\zeta, 1/\zeta)$, $Y = \log(U) \sim \text{log-GA}(1/\zeta, 1/\zeta)$; see Johnson et al. (1995).

  Some properties of the log-BS distribution are as follows. If $Y \sim \text{log-BS}(\sqrt{2/\delta}, \log(\delta\mu/(\delta+1)))$, then: (a) $U = \exp(Y) \sim \text{BS}(\mu,\delta)$; (b) $\text{E}(Y) = \log(\delta\mu/(\delta+1))$; (c) there is no closed form for the variance of Y, but based upon an asymptotic approximation for the log-BS moment generating function, it follows that, as $\delta \to \infty$, $\text{Var}(Y) = 2/\delta - 1/\delta^2$, whereas that, in contrast, as $\delta \to 0$, $\text{Var}(Y) = 4(\log^2(2/\sqrt{\delta}) + 2 - 2\log(2/\sqrt{\delta}))$; and (d) the distribution of Y is symmetric around μ, unimodal for $\delta \geq 0.5$ and bimodal for $\delta < 0.5$; see Rieck and Nedelman (1991) and Leiva (2016). Figure 2 shows some shapes for the PDF of $Y = \log(U)$ in each aforementioned distribution. Note that the bimodality property is only found in the log-BS case.

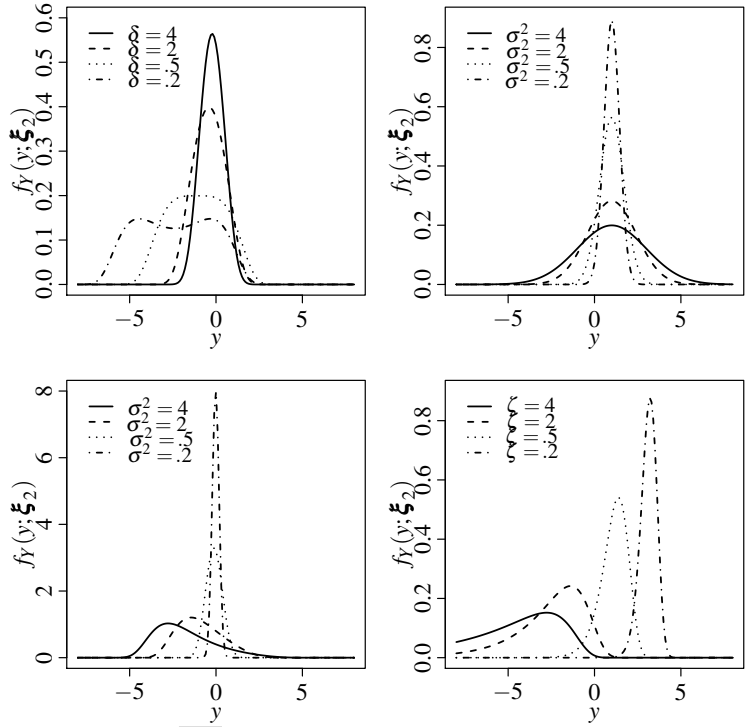

Figure 2 – PDF plots of the log-BS($\sqrt{2/\delta}, \log(\delta/(\delta+1))$), N$(1,\sigma^2)$, log-IG$(\sigma^2,0)$ and log-GA$(1/\zeta,1/\zeta)$ distributions.

2.5 Bivariate Birnbaum-Saunders distribution

Kundu et al. (2010) introduced the bivariate Birnbaum-Saunders (BBS) distribution, where the authors discussed maximum likelihood estimation and modified moment estimation of the model parameters. Recently, Khosravi et al. (2015) observed that the bivariate BS proposed in Kundu et al. (2010) can be written as the weighted mixture of bivariate inverse Gaussian distribution and its reciprocals; see Kocherlakota (1986). They also introduced a mixture of two bivariate BS distributions and discussed its various properties. Kundu et al. (2013) extended to the multivariate case, the generalized BS distribution introduced by Díaz-Garcia and Leiva (2005). Other bivariate and multivariate distributions related to the BS model can be found in Vilca et al. (2014a), Vilca et al. (2014b), Kundu et al. (2015), Kundu (2015) and Jamalizadeh and Kundu (2015).

If the random vector $\boldsymbol{T} = (T_1, T_2)^\top$ is BBS distributed with parameter vectors $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^\top$ and $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$, and correlation coefficient ρ, denoted by $\boldsymbol{T} \sim \text{BBS}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \rho)$, then its joint CDF and PDF are given by

$$F_{\mathrm{BBS}}(\boldsymbol{t};\boldsymbol{\alpha},\boldsymbol{\beta},\rho)=\Phi_2\left(\frac{1}{\alpha_1}\left(\sqrt{\frac{t_1}{\beta_1}}-\sqrt{\frac{\beta_1}{t_1}}\right),\frac{1}{\alpha_2}\left(\sqrt{\frac{t_2}{\beta_2}}-\sqrt{\frac{\beta_2}{t_2}}\right);\rho,g_c\right),\quad \boldsymbol{t}>\mathbf{0},\tag{2.15}$$

$$\begin{aligned} f_{\mathrm{BBS}}(\boldsymbol{t};\boldsymbol{\alpha},\boldsymbol{\beta},\rho) &= \phi_2\left(\frac{1}{\alpha_1}\left(\sqrt{\frac{t_1}{\beta_1}}-\sqrt{\frac{\beta_1}{t_1}}\right),\frac{1}{\alpha_2}\left(\sqrt{\frac{t_2}{\beta_2}}-\sqrt{\frac{\beta_2}{t_2}}\right);\rho\right) \\ &\quad\times\frac{1}{2\alpha_1}\left\{\left(\frac{\beta_1}{t_1}\right)^{\frac{1}{2}}+\left(\frac{\beta_1}{t_1}\right)^{\frac{3}{2}}\right\}\frac{1}{2\alpha_2}\left\{\left(\frac{\beta_2}{t_2}\right)^{\frac{1}{2}}+\left(\frac{\beta_2}{t_2}\right)^{\frac{3}{2}}\right\},\quad \boldsymbol{t}>\mathbf{0}, \end{aligned}\tag{2.16}$$

where $\alpha_k>0$ and $\beta_k>0$ for $k=1,2$, $-1<\rho<1$, and $\phi_2(\cdot,\cdot;\rho)$ is a normal joint PDF given by

$$\phi_2(u,v;\rho)=\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left\{\frac{1}{(1-\rho^2)}(u^2+v^2-2\rho uv)\right\}.$$

Kundu et al. (2010) present some properties of the BBS distribution, for example

Proposition 1 If $\boldsymbol{T}=(T_1,T_2)^\top\sim\mathrm{BBS}(\boldsymbol{\alpha},\boldsymbol{\beta},\rho)$, then

a) $\boldsymbol{T}^{-1}=(T_1^{-1},T_2^{-1})^\top\sim\mathrm{BBS}(\alpha_1,1/\beta_1,\alpha_2,1/\beta_2,\rho)$;
b) $\boldsymbol{T}_1^{-1}=(T_1^{-1},T_2)^\top\sim\mathrm{BBS}(\alpha_1,1/\beta_1,\alpha_2,\beta_2,-\rho)$;
c) $\boldsymbol{T}_2^{-1}=(T_1,T_2^{-1})^\top\sim\mathrm{BBS}(\alpha_1,\beta_1,\alpha_2,1/\beta_2,-\rho)$;

Proposition 2 If $\boldsymbol{T}=(T_1,T_2)^\top\sim\mathrm{BBS}(\boldsymbol{\alpha},\boldsymbol{\beta},\rho)$, then

a) The conditional PDF of T_1, given $T_2=t_2$ is given by

$$\begin{aligned} f_{T_1|T_2=t_2}(t_1) &= \frac{1}{\alpha_1\beta_1\sqrt{2\pi(1-\rho^2)}}\left[\left(\frac{\beta_1}{t_1}\right)^{1/2}+\left(\frac{\beta_1}{t_1}\right)^{3/2}\right] \\ &\quad\times\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{1}{\alpha_1}\left(\sqrt{\frac{t_1}{\beta_1}}-\sqrt{\frac{\beta_1}{t_1}}\right)-\frac{\rho}{\alpha_2}\left(\sqrt{\frac{t_2}{\beta_2}}-\sqrt{\frac{\beta_2}{t_2}}\right)\right]^2\right\} \end{aligned}$$

b) The conditional CDF of T_1, given $T_2=t_2$ is given by

$$P[T_1\le t_1|T_2=t_2] = \Phi\left\{\frac{\frac{1}{\alpha_1}\left(\sqrt{\frac{t_1}{\beta_1}}-\sqrt{\frac{\beta_1}{t_1}}\right)-\frac{\rho}{\alpha_2}\left(\sqrt{\frac{t_2}{\beta_2}}-\sqrt{\frac{\beta_2}{t_2}}\right)}{\sqrt{1-\rho^2}}\right\}$$

2.6 Frailty models

Consider an unobserved source of heterogeneity that is not readily captured by a covariate in a univariate frailty model. This extends the Cox model, such that the HR of a patient depends

on an unobservable RV U, which acts multiplicatively on the baseline HR. Therefore, the conditional HR of the lifetime T, given $U = u_i$ for the patient i at time t, is

$$h_{T|U=u_i}(t;\boldsymbol{\xi}_1,\boldsymbol{\xi}_2) = u_i h_0(t;\boldsymbol{\xi}_1), \quad i = 1,\ldots,n, \quad t > 0, \tag{2.17}$$

where u_i is the frailty of the patient i and h_0 is a baseline HR, that is, we consider the case with a proportional HR. In (2.17), note that $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ are vectors of the model parameters related to the lifetime and frailty distributions, respectively, of the patient i. In addition, observe that (2.17) is known as the Clayton model; see Clayton (1991). From (2.17), a patient i is called "standard" if his/her frailty is $u_i = 1$; "twice as likely to die" if his/her frailty is $u_i = 2$, at any particular time and in relation to the standard patient; and "one-half as likely to die" if his/her frailty is $u_i = 1/2$; see Vaupel et al. (1979). The corresponding conditional SF of T is $S_{T|U=u_i}(t;\boldsymbol{\xi}_1,\boldsymbol{\xi}_2) = (S_0(t;\boldsymbol{\xi}_1))^{u_i}$, for $i = 1,\ldots,n$ and $t > 0$, which represents the probability of the patient i to be alive at time t given the random effect $U_i = u_i$.

If values of covariates in the model given in (2.17) are introduced similarly to the Cox model, we have

$$h_{T|U=u_i}(t;\boldsymbol{x},\boldsymbol{\xi}) = u_i h_0(t;\boldsymbol{\xi}_1)\exp(\boldsymbol{x}_i^\top \boldsymbol{\varphi}), \quad i = 1,\ldots,n, \quad t > 0, \tag{2.18}$$

where $\boldsymbol{x}_i^\top = (1, x_{1i},\ldots,x_{pi})$ is a vector containing the values of p covariates for the patient i, $\boldsymbol{\varphi} = (\varphi_0,\varphi_1,\ldots,\varphi_p)^\top$ is the vector of regression coefficients to be estimated, and $\boldsymbol{\xi} = (\boldsymbol{\xi}_1^\top,\boldsymbol{\xi}_2^\top,\boldsymbol{\varphi}^\top)^\top$. Therefore, the frailty model given in (2.18) is a generalization of the proportional hazard model, which is obtained when the frailty distribution degenerates at $U = 1$ for all patients. The corresponding conditional SF can be obtained from (2.18) as

$$S_{T|U=u_i}(t;\boldsymbol{x},\boldsymbol{\xi}) = \exp(-u_i H_0(t;\boldsymbol{\xi}_1)\exp(\boldsymbol{x}^\top\boldsymbol{\varphi})), \quad i = 1,\ldots,n, \quad t > 0, \tag{2.19}$$

where $H_0(t;\boldsymbol{\xi}_1) = \int_0^t h_0(s;\boldsymbol{\xi}_1)\,\mathrm{d}s$ is the baseline cumulative hazard rate (CHR).

Suppose that the lifetime is not completely observed and may be subject to right censoring. Let v_i denote the censoring time, y_i the time to event of interest and u_i the frailty for the patient i, respectively. We observe $t_i = \min\{y_i, v_i\}$, that is, if the censoring indicator $\varsigma_i = 1$, $t_i = y_i$ is the lifetime of the patient i; otherwise, if $\varsigma_i = 0$, $t_i = v_i$ is the right censoring time of the patient i; for $i = 1,\ldots,n$. Then, from (2.18) and (2.19), the corresponding likelihood function is

$$L(\boldsymbol{\xi};\boldsymbol{t},\boldsymbol{\varsigma},\boldsymbol{x},\boldsymbol{u}) = L(\boldsymbol{\xi}) = \prod_{i=1}^{n}\Big(u_i h_0(t;\boldsymbol{\xi}_1)\exp(\boldsymbol{x}^\top\boldsymbol{\varphi})\Big)^{\varsigma_i}\exp\Big(-u_i H_0(t;\boldsymbol{\xi}_1)\exp(\boldsymbol{x}^\top\boldsymbol{\varphi})\Big), \tag{2.20}$$

where $\boldsymbol{\xi}$ is defined in (2.18), $\boldsymbol{t} = (t_1,\ldots,t_n)^\top$ are the lifetimes of the patients, $\boldsymbol{\varsigma} = (\varsigma_1,\ldots,\varsigma_n)^\top$ is the vector of their censoring indicators, and $\boldsymbol{u} = (u_1,\ldots,u_n)^\top$ is the vector of their frailties. Now, conditional on the unobserved frailties $\boldsymbol{u}$, the likelihood function given in (2.20) forms the basis for the parameter estimation. The frailties $\boldsymbol{u}$ must be integrated out (in closed form or by numerical or stochastic integration, depending on the frailty distribution) to get a likelihood

function (not depending on unobserved quantities) of the type

$$L(\boldsymbol{\xi};\boldsymbol{t},\boldsymbol{\varsigma},\boldsymbol{x}) = L(\boldsymbol{\xi}) = \prod_{i=1}^{n} (h_T(t_i;\boldsymbol{x},\boldsymbol{\xi}))^{\varsigma_i} S_T(t_i;\boldsymbol{x},\boldsymbol{\xi}), \tag{2.21}$$

where h_T and S_T are the unconditional HR and SF, respectively, defined next.

2.6.1 Unconditional hazard and survival functions

The unconditional (population) SF of T can be obtained by integrating $S_{T|U=u_i}(t;\boldsymbol{x})$ given in (2.19) on the frailty U. It may be viewed as the (unconditional) SF of patients randomly drawn from the population under study; see Klein and Moeschberger (2003), Aalen et al. (2008) and Wienke (2011). Unconditional HF and SF can be obtained with the Laplace transform; see Hougaard (1984). Then, when seeking distributions for the frailty RV U, it is natural to use frailty distributions with an explicit Laplace transform, because it facilitates the use of standard ML methods for parameter estimation. To get the unconditional SF, we need to integrate out the frailty component as

$$S_T(t;\boldsymbol{x},\boldsymbol{\xi}) = \int_0^\infty S_{T|U=u}(t;\boldsymbol{x},\boldsymbol{\xi}) f_U(u;\boldsymbol{\xi}_2)\,\mathrm{d}u, \tag{2.22}$$

where $\boldsymbol{\xi}$ is defined in (2.18), $S_{T|U=u}(t;\boldsymbol{x},\boldsymbol{\xi}) = \exp(-uH_0(t;\boldsymbol{\xi}_1)\exp(\boldsymbol{x}^\top\boldsymbol{\varphi}))$ is the conditional SF as given in (2.19) and f_U is the corresponding frailty PDF. The Laplace transform of real argument s of a function f is

$$Q(s) = \int_0^\infty \exp(-sx) f(x)\,\mathrm{d}x. \tag{2.23}$$

Let $f = f_U$ be the frailty PDF and $s = H_0(t;\boldsymbol{\xi}_1)\exp(\boldsymbol{x}^\top\boldsymbol{\varphi})$. Then, according to (2.23), we obtain the Laplace transform of the unconditional SF of T as

$$S_T(t;\boldsymbol{x},\boldsymbol{\xi}) = \int_0^\infty \exp(-uH_0(t;\boldsymbol{\xi}_1)\exp(\boldsymbol{x}^\top\boldsymbol{\varphi})) f_U(u;\boldsymbol{\xi}_2)\,\mathrm{d}u = Q(H_0(t;\boldsymbol{\xi}_1)\exp(\boldsymbol{x}^\top\boldsymbol{\varphi})). \tag{2.24}$$

Note that (2.24) conducts to the same form as the unconditional SF given in (2.22); see Vaupel et al. (1979) and Wienke (2011). The frailty RVs U_i are usually assumed to be independent with identical frailty distribution. As mentioned, the frailty distribution can be GA, IG, LN or Weibull. We consider a reparameterized version of the BS distribution introduced by Santos-Neto et al. (2012), Santos-Neto et al. (2014), because it allows us to mimic a property of the GA distribution, traditionally used in frailty models.

2.7 Cure rate models

The unified long-term survival model has its formulation based on a biological context as in Yakovlev and Tsodikov (1996) and Chen et al. (1999); further details can be seen in Rodrigues et al. (2009). For a subject in the population, we represent by N the number of competing causes related to the occurrence of an event of interest. Given $N = n$ the promotion time for the jth competing cause is denoted by Z_j, $j = 1,\ldots,n$. We assume that, conditional on N, the Z_j's are independent and identically distributed (IID). We suppose also that N is independent of $(Z_1,\ldots,Z_n)$. The observable time to event is defined as $T = \min\{Z_1, Z_2, \ldots, Z_N\}$ for $N \geq 1$, and $T = \infty$ if $N = 0$, which leads to a cure fraction p_0. According to Rodrigues et al. (2009), the long-term survival function of the RV T is given by

$$\begin{aligned} S_p(t) =& P(T \geq t) = P(N = 0) + \sum_{n=1}^{\infty} P(Z_1 > t, \ldots, Z_N > t | N = n) P(N = n) \\ =& \sum_{n=0}^{\infty} P(N = n)[S_T^*(t)]^n = \mathbb{A}_N[S_T^*(t)], \end{aligned} \tag{2.25}$$

where $S_T^*(\cdot)$ denotes the common survival function of the unobserved lifetimes and $\mathbb{A}_N[\cdot]$ is the probability generating function of the RV N, which converges when $u = S_T^*(t) \in [0,1]$. Thus, various results can be obtained for each choice of the generating function of the distribution of N and $S_T^*(t)$. More details about this model can be found in Rodrigues et al. (2009).

In this work, we assume that the unobserved latent variable N has a negative binomial distribution Piegorsch (1990), Saha and Paul (2005) with probability mass function given by

$$P(N = n) = \frac{\Gamma(n + \phi^{-1})}{n!\Gamma(\phi^{-1})} \left(\frac{\phi\theta}{1 + \phi\theta}\right)^n (1 + \phi\theta)^{-1/\phi}, \tag{2.26}$$

with $n = 0, 1, \ldots$, $\theta > 0$, $\phi \geq -1$ and $1 + \phi\theta > 0$, so that $\mathbb{E}(N) = \theta$ and $\mathrm{Var}(N) = \theta + \phi\theta^2$.

As discussed by Tournoud and Ecochard (2007), the parameters of the negative binomial distribution have biological interpretations. The mean of the number of competing causes is represented by θ, whereas ϕ is the dispersion parameter. The variance of the number of initiated cells is flexible: there is under-dispersion in the Poisson model when $-1/\theta < \phi < 0$, whereas for $\phi > 0$ over-dispersion is present. The negative binomial model comprises some well-known models when the parameter ϕ is fixed. For instance, if $\phi \to 0$ the probability function of the Poisson distribution is obtained, when $\phi = -1$ the Bernoulli distribution and if $\phi = 1$ the geometric distribution. Therefore, considering the number of competing causes to be negative binomial distributed and $S_T^*(\cdot)$ a proper SF, we have that the long-term SF of the RV T is given by

$$S_p(t) = \{1 + \phi\theta(1 - S_T^*(t))\}^{-1/\phi}, \tag{2.27}$$

where $p_0 = \lim_{t\to\infty} S_p(t) = (1 + \phi\theta)^{-1/\phi} > 0$, p_0 represents the proportion of cured or immune individuals in the population. From (2.25) we can obtain the PDF associated with the long-term

SF which is given by

$$f_p(t) = f_T^*(t)\left(\frac{\mathrm{d}\mathbb{A}(u)}{\mathrm{d}u}|_{u=S_T^*(t)}\right), \tag{2.28}$$

and the long-term hazard function is defined as

$$h_p(t) = \frac{f_p(t)}{S_p(t)} = f_T^*(t)\frac{\frac{\mathrm{d}\mathbb{A}(u)}{\mathrm{d}u}|_{u=S_T^*(t)}}{S_p(t)}. \tag{2.29}$$

From (2.28) and (2.29) the corresponding PDF and HR become

$$f_p(t) = \theta f_T^*(t)\{1+\phi\theta[1-S_T^*(t)]\}^{-1/\phi-1}, \tag{2.30}$$

and

$$h_p(t) = \theta f_T^*(t)\{1+\phi\theta[1-S_T^*(t)]\}^{-1}. \tag{2.31}$$

Notice that $f_p(t)$ and $h_p(t)$ are improper functions, since $f_p(t)$ is not a proper SF.

CHAPTER

3

A BIRNBAUM-SAUNDERS FRAILTY MODEL FOR SURVIVAL DATA

3.1 Introduction

In this chapter, we introduce the BS frailty model, discuss aspects of model identifiability, estimate its parameters, introduce two residuals, conduct a simulation study to evaluate the behavior of the parameter estimators and illustrate the potentiality of the proposed model with two real-world data sets. Here, we consider BS frailty model without observed covariates, but they will be explicitly considered in the next chapter.

3.2 Birnbaum-Saunders frailty model

In this section, we discuss some model identifiability issues and how to estimate the model parameters via the ML method and to infer about these parameters.

3.2.1 Model identifiability and features

In univariate frailty models, an important aspect is its identifiability. In the context of proportional hazard models, when working with frailty, it is necessary that the random effect distribution has finite mean for the model to be identifiable; see Elbers and Ridder (1982). Thus, in order to keep the identifiability of the model, it is convenient to take the distribution with mean equal to one. We assume that the frailty U has a BS distribution with parameters $\mu = 1$ and δ, where $\mathrm{E}[U] = 1$ and $\mathrm{Var}[U] = (2\delta+5)/(\delta+1)^2$. The variance quantifies the amount of heterogeneity among patients.

From (2.22), the Laplace transform for the BS distribution with parameters $\mu = 1$ and δ is given by

$$Q(s) = \frac{\exp\left(\frac{\delta}{2}\left(1-\sqrt{\delta+4s+1}/\sqrt{\delta+1}\right)\right)\left(\sqrt{\delta+4s+1}+\sqrt{\delta+1}\right)}{2\sqrt{\delta+4s+1}}. \tag{3.1}$$

From (2.22) and evaluating (3.1) at $s = H_0(t)$, we obtain the unconditional SF under the BS frailty as

$$S_T(t) = \frac{\exp(\frac{\delta}{2}(1-\sqrt{\delta+4H_0(t)+1}/\sqrt{\delta+1}))(\sqrt{\delta+4H_0(t)+1}+\sqrt{\delta+1})}{2\sqrt{\delta+4H_0(t)+1}}. \tag{3.2}$$

Then, from (3.1), the corresponding unconditional HR is given by

$$h_T(t) = h_0(t)\left(\frac{\delta(\delta+\sqrt{\delta+1}\sqrt{\delta+4H_0(t)+1}+4H_0(t)+3)+2}{(\delta+4H_0(t)+1)(\delta+\sqrt{\delta+1}\sqrt{\delta+4H_0(t)+1}+1)}\right). \tag{3.3}$$

We assume that the baseline HR $h_0(t)$ is specified up to a few unknown parameters, which are related to a distribution assumed for the baseline hazard. For example, we can suppose an exponential, LN or Weibull distribution. However, assuming a parametric distribution is not always desirable, because such a assumption is often difficult to verify. Note that the exponential distribution has been extensively used to model the baseline HR due to its simplicity or when the HR must be constant for each patient; see Lawless (2011). Therefore, we use the exponential distribution as baseline hazard, which has $h_0(t) = \gamma$ and $H_0(t) = \gamma t$, for $t > 0$. Thus, from (3.3), the unconditional HR under BS frailty reduces to

$$h_T(t) = \frac{\gamma(\delta(\delta+\sqrt{\delta+1}\sqrt{\delta+4\gamma t+1}+4\gamma t+3)+2)}{(\delta+4\gamma t+1)(\delta+\sqrt{\delta+1}\sqrt{\delta+4\gamma t+1}+1)}, \tag{3.4}$$

where γ is the average HR of each patient. From (3.2), the unconditional SF under BS frailty is

$$S_T(t) = \frac{\exp\left(\frac{1}{2}\delta\left(1-\sqrt{\delta+4\gamma t+1}/\sqrt{\delta+1}\right)\right)\left(\sqrt{\delta+1}+\sqrt{\delta+4\gamma t+1}\right)}{2\sqrt{\delta+4\gamma t+1}}. \tag{3.5}$$

Note that (3.2) and (3.3) can be easily applied to different baselines other than the exponential one. In fact, in Section 3.4 we also consider a Weibull baseline.

3.3 Estimation of parameters

Assuming that the time to the event of interest is not completely observed and it may be subject to right censoring. Let $\boldsymbol{\xi} = (\delta, \gamma)^\top$ denote the parameter vector of the BS frailty model of the time to event in (3.5). From n pairs of times and censoring indicators $(t_1, \varsigma_1), \ldots, (t_n, \varsigma_n)$,

the corresponding likelihood function under uninformative censoring can be expressed as

$$L(\boldsymbol{\xi}) = \prod_{i=1}^{n}\left(\frac{\gamma(\delta\left(\delta+\sqrt{\delta+1}\sqrt{\delta+4\gamma t_i+1}+4\gamma t_i+3\right)+2)}{(\delta+4\gamma t_i+1)\left(\delta+\sqrt{\delta+1}\sqrt{\delta+4\gamma t_i+1}+1\right)}\right)^{\varsigma_i} \times\left(\frac{\exp(\frac{1}{2}\delta(1-\frac{\sqrt{\delta+4\gamma t_i+1}}{\sqrt{\delta+1}})(\sqrt{\delta+1}+\sqrt{\delta+4\gamma t_i+1})}{2\sqrt{\delta+4\gamma t_i+1}}\right). \tag{3.6}$$

Therefore, the log-likelihood function for the BS frailty model obtained from (3.6) is given by

$$\begin{aligned}\ell(\boldsymbol{\xi}) = {} & \frac{n\delta}{2}-\frac{\delta}{2\sqrt{\delta+1}}\sum_{i=1}^{n}\sqrt{\delta+4\gamma t_i+1}-\sum_{i=1}^{n}\varsigma_i\log(\delta+4\gamma t_i+1)\\ & -\sum_{i=1}^{n}\log(2\sqrt{\delta+4\gamma t_i+1})+\sum_{i=1}^{n}\log(\sqrt{\delta+1}+\sqrt{\delta+4\gamma t_i+1})\\ & +\sum_{i=1}^{n}\varsigma_i\log(\gamma(\delta(\delta+\sqrt{(\delta+1)(\delta+4\gamma t_i+1)}+4\gamma t_i+3)+2))\\ & -\sum_{i=1}^{n}\varsigma_i\log(\delta+\sqrt{(\delta+1)(\delta+4\gamma t_i+1)}+1).\end{aligned} \tag{3.7}$$

Then, the first derivatives of the log-likelihood function with respect to the two parameters can be obtained; see Appendix A. The ML equations for δ and γ must be solved with an iterative method for non-linear optimization problems. Specifically, the ML estimates of the BS frailty model parameters can be obtained by using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton non-linear optimization algorithm with numeric derivatives; see Nocedal and Wright (2006) and Lange (2010). The BFGS method is implemented in the `R` software by the functions `optim` and `optimx`; see <www.R-project.org> and R Core Team (2016).

Standard regularity conditions; see for example Cox and Hinkley (1979), Serfling (1980), Lehmann and Casella (2006), Migon et al. (2014), are fulfilled for the proposed model, whenever the parameters are within the parameter space. It is well known the ML estimators are asymptotically normally distributed. Thus, for the BS frailty model, we have

$$\widehat{\boldsymbol{\xi}} \xrightarrow{D} \mathrm{N}_2(\boldsymbol{\xi},\boldsymbol{\Sigma}_\xi),$$

where $\boldsymbol{\Sigma}_\xi$ is the asymptotic variance-covariance matrix of $\widehat{\boldsymbol{\xi}}$ and $\xrightarrow{D}$ denotes convergence in distribution. Therefore, an approximate $100\times(1-\varpi)\%$ confidence interval (CI) for $\boldsymbol{\xi}$ is

$$\mathscr{R} = \{\boldsymbol{\xi}\in\mathbb{R}^2 : |\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi}|^{\top}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\xi}}^{-1}|\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi}|\leq\chi^2_{2;1-\varpi}\},\ 0<\varpi<1, \tag{3.8}$$

where $\chi^2_{2;1-\varpi}$ denotes the $100\times(1-\varpi)$th quantile of the chi-squared distribution with two degrees of freedom and $\widehat{\boldsymbol{\Sigma}}_\xi$ is an estimate of $\boldsymbol{\Sigma}_\xi$. Confidence bands for the BS frailty model

parameters can be obtained by means of the region provided in (3.8). The observed information matrix for the BS frailty model is given in Appendix A, which is useful for computing $\widehat{\boldsymbol{\Sigma}}_{\xi}$.

The goodness-of-fit of the BS frailty model is assessed by two residuals. The first is a generalized Cox-Snell (GCS) residual given by

$$r_i^{\mathrm{GCS}} = -\log(\widehat{S}_T(t_i)), \quad i = 1,\ldots,n, \tag{3.9}$$

where $\widehat{S}_T(t_i)$ is the fitted survival function of the ith lifetime. The GCS has a unit exponential (EXP(1)) distribution when the frailty model is correctly specified, regardless of the frailty model specification. The second type is the randomized quantile (RQ), which is usually used in GAMLSS models; see Dunn and Smyth (1996). The RQ is given by

$$r_i^{\mathrm{RQ}} = \Phi^{-1}(\widehat{S}_T(t_i)), \quad i = 1,\ldots,n, \tag{3.10}$$

where $\Phi^{-1}(\cdot)$ is the standard normal CDF and $\widehat{S}_T(t_i)$ is as in (3.9). The RQ residual defined in (3.10) has a standard normal distribution if the frailty model is correctly specified for any frailty model specification.

3.3.1 A simulation study

We here carry out a simulation study to evaluate the performance of the ML estimators of the BS frailty model parameters with exponential baseline HR. The simulation scenario considered the following: sample sizes $n \in \{30, 150, 400, 600\}$, values of the true parameter $\delta \in \{0.25, 0.50, 1.50, 2.50\}$, 5000 MC replications, and without loss of generality, we assume $\gamma = 1.0$ in all cases. The percentage of censored observations were $\{0, 10, 25, 40\}$.

Note that, based on the probability integral transform, the BS frailty CDF follows a U(0,1) distribution. Then, the BS frailty SF is U(0,1) distributed as well. Simulation studies to evaluate the performance of the ML estimators can be carried out by generating random numbers from the BS frailty model from Algorithm 1.

Algorithm 1 – Generator of random numbers from the BS frailty model.

1: Obtain a random number m_i from $M \sim \mathrm{U}(0,1)$.
2: Fix values of δ and γ.
3: Equate m_i to the SF and obtain the lifetime y_i by solving numerically the equation

$$\frac{\exp(\frac{1}{2}\delta(1-\sqrt{\delta+4\gamma y_i+1}/\sqrt{\delta+1}))(\sqrt{\delta+1}+\sqrt{\delta+4\gamma y_i+1})}{2\sqrt{\delta+4\gamma y_i+1}} = m_i.$$

4: Obtain the censored time c_i from $C \sim \mathrm{U}(a,b)$, where $a, b > 0$.
5: Obtain $t_i = \min\{y_i, c_i\}$.
6: If $y_i < c_i$ then $\varsigma_i = 1$, otherwise $\varsigma_i = 0$.
7: Repeat Steps 1 to 6 until the required number of data has been generated.

In Step #3, we use the function `uniroot` of the `R` software to obtain the root of the equation; see Brent (2013). For each value of the parameter, sample size and percentage of censored observations, we report the empirical values for the bias (B) and mean squared error (MSE) of the ML estimators in Table 2. From this table, note that, as the sample size increases, the ML estimators become more efficient, as expected. We can also note that, as the percentage of censored observations increases, the performance of the estimator of δ, the shape parameter, deteriorates, especially for $n = 30$. From Figures 3-4 we notice that the estimators of parameter δ tend to be unbiased and consistent. Generally, all of these results show the good performance of the proposed model.

Table 2 – empirical bias (with MSEs in parentheses) of the ML estimators of δ and γ from the BS frailty model under different censoring proportions.

		BS frailty model			
n		0%		10 %	
	δ	$B(\widehat{\delta})$	$B(\widehat{\gamma})$	$B(\widehat{\delta})$	$B(\widehat{\gamma})$
30	0.25	0.0180 (0.0474)	0.1243 (0.2448)	−0.0201 (0.0501)	−0.1300 (0.2488)
	0.50	0.0287 (0.0887)	0.1129 (0.2317)	−0.0431 (0.1004)	−0.1108 (0.2217)
	1.50	0.2515 (0.3429)	0.0538 (0.1522)	−0.2842 (0.3788)	−0.0530 (0.1651)
	2.50	0.5139 (0.6520)	0.0652 (0.1484)	−0.5949 (0.7340)	−0.0667 (0.1540)
150	0.25	0.0076 (0.0281)	0.0198 (0.1422)	−0.0054 (0.0300)	−0.0295 (0.1418)
	0.50	0.0173 (0.0583)	0.0186 (0.1326)	−0.0081 (0.0632)	−0.0171 (0.1299)
	1.50	0.1292 (0.2013)	0.0528 (0.1122)	−0.1501 (0.2262)	−0.0514 (0.1070)
	2.50	0.3081 (0.4130)	0.0603 (0.1096)	−0.3475 (0.4552)	−0.0600 (0.1071)
400	0.25	0.0056 (0.0241)	0.0088 (0.1109)	−0.0035 (0.0240)	−0.0016 (0.1080)
	0.50	0.0142 (0.0485)	0.0008 (0.1043)	−0.0075 (0.0507)	−0.0064 (0.1067)
	1.50	0.0793 (0.1428)	0.0361 (0.0958)	−0.0998 (0.1652)	−0.0407 (0.0974)
	2.50	0.2050 (0.2933)	0.0548 (0.0968)	−0.2460 (0.3399)	−0.0491 (0.0993)
600	0.25	0.0048 (0.0221)	0.0039 (0.1050)	−0.0048 (0.0222)	−0.0001 (0.1039)
	0.50	0.0120 (0.0463)	0.0120 (0.0988)	−0.0065 (0.0470)	−0.0104 (0.0954)
	1.50	0.0586 (0.1225)	0.0259 (0.0906)	−0.0745 (0.1390)	−0.0302 (0.0879)
	2.50	0.1601 (0.2406)	0.0166 (0.0888)	−0.2071 (0.2941)	−0.0259 (0.0906)
		BS frailty model			
n		25%		40 %	
	δ	$B(\widehat{\delta})$	$B(\widehat{\gamma})$	$B(\widehat{\delta})$	$B(\widehat{\gamma})$
30	0.25	−0.0322 (0.0625)	−0.1496 (0.2632)	−0.0524 (0.0853)	−0.1550 (0.2654)
	0.50	−0.0618 (0.1211)	−0.1121 (0.2298)	−0.1007 (0.1635)	−0.1300 (0.2465)
	1.50	−0.3342 (0.4377)	−0.0549 (0.1776)	−0.4296 (0.5395)	−0.0683 (0.1888)
	2.50	−0.7106 (0.8647)	−0.0651 (0.1605)	−0.8870 (1.0495)	−0.0606 (0.1744)
150	0.25	−0.0081 (0.0361)	−0.0385 (0.1483)	−0.0147 (0.0432)	−0.0405 (0.1501)
	0.50	−0.0163 (0.0718)	−0.0139 (0.1322)	−0.0316 (0.0866)	−0.0293 (0.1332)
	1.50	−0.2016 (0.2779)	−0.0554 (0.1152)	−0.2488 (0.3380)	−0.0498 (0.1118)
	2.50	−0.4300 (0.5518)	−0.0612 (0.1121)	−0.5518 (0.6830)	−0.0598 (0.1118)
400	0.25	−0.0017 (0.0273)	−0.0042 (0.1139)	−0.0028 (0.0315)	−0.0039 (0.1131)
	0.50	−0.0032 (0.0563)	−0.0068 (0.1068)	−0.0075 (0.0661)	−0.0191 (0.1068)
	1.50	−0.1292 (0.1974)	−0.0507 (0.0986)	−0.1692 (0.2431)	−0.0562 (0.0994)
	2.50	−0.3097 (0.4059)	−0.0497 (0.0984)	−0.3984 (0.5129)	−0.0486 (0.0998)
600	0.25	−0.0011 (0.0248)	−0.0087 (0.1011)	−0.0027 (0.0285)	−0.0157 (0.1054)
	0.50	−0.0019 (0.0539)	−0.0060 (0.0969)	−0.0017 (0.0591)	−0.0149 (0.0946)
	1.50	−0.1041 (0.1700)	−0.0372 (0.0921)	−0.1561 (0.2256)	−0.0267 (0.0936)
	2.50	−0.2681 (0.3623)	−0.0265 (0.0931)	−0.3406 (0.4450)	−0.0426 (0.0946)

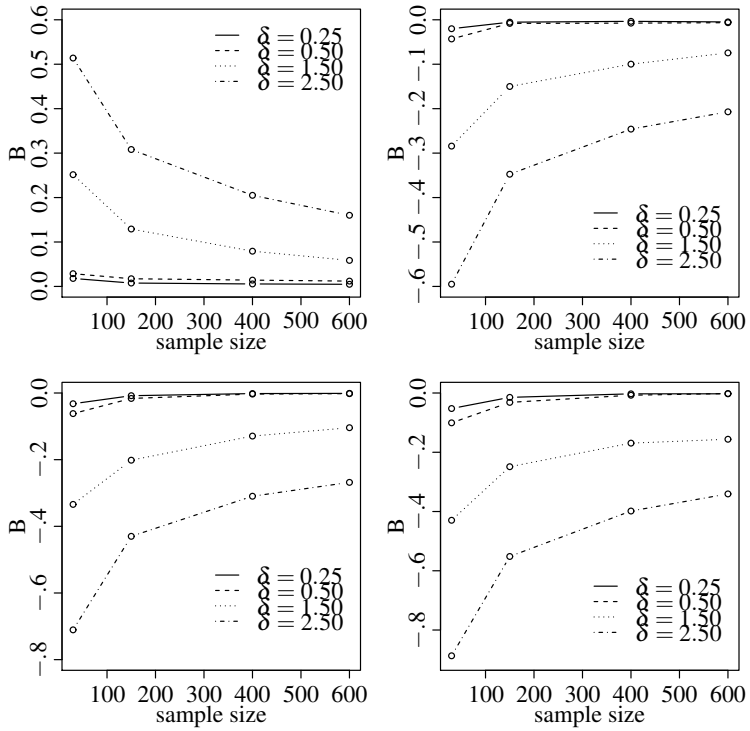

Figure 3 – Bias from different values of δ and sample sizes.

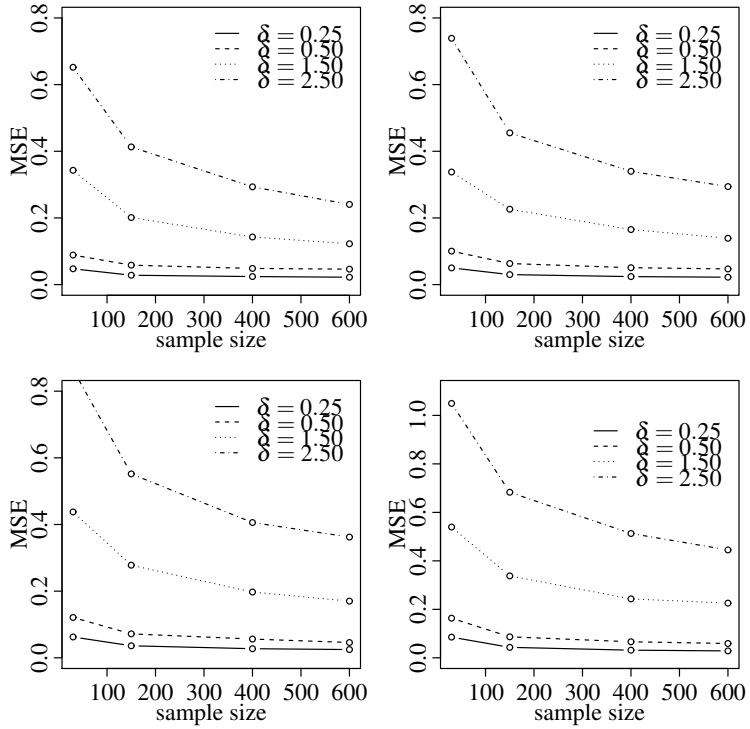

Figure 4 – MSE from different values of δ and sample sizes.

3.4 Applications to real data

In this section, we illustrate the proposed methodology for modeling unobserved frailty by applying it to two real-world medical data sets. The first (uncesored) of them from a leukemia cancer study introduced by Feigl and Zelen (1965), whereas the second (censored) from a lung cancer trial studied in Kalbfleisch and Prentice (2011). Given the frailty component, the times of death are independent and follow a proportional hazard model. We compare the proposed BS frailty model, with the Weibull model without frailty, GA frailty model and IG model with all of them having a Weibull baseline. We assess the impact of the frailty model on the frailty variance. Then, we find the model that provides an adequate fit to the data. To make sure that the GA and IG models are identifiable, we consider $U \sim \text{GA}(1/\zeta, 1/\zeta)$ and $U \sim \text{IG}(1, \sigma^2)$; see Wienke (2011).

3.4.1 First case study: leukemia cancer data

The data set corresponds to the survival times of 33 patients who died from acute myelogenous leukemia. Also, their measures of white blood cell count at the time of diagnosis were recorded. The patients were divided into 2 groups, according to the presence or absence of a morphological characteristic of white blood cells. Patients termed as AG positive were identified by the presence of significant granulation of the leukemic cells in the bone marrow at the time of diagnosis. Besides the covariates white blood cell count and groups positive/negative AG, other factors related to leukemia lifetime correspond to certain chemical agents and genetic factors, which were not measured. It motivates the use of the frailty models to capture the influence of such factors.

Exploratory data analysis

Table 3 provides a descriptive summary of the observed lifetime (in weeks) that includes median (MD), mean ($\bar{t}$), standard deviation (SD), coefficient of variation (CV), coefficient of skewness (CS), coefficient of kurtosis (CK), and minimum ($t_{(1)}$) and maximum ($t_{(n)}$) values. From this table, we observe the positively skewed nature and moderate kurtosis level of the data distribution. The skewed nature is confirmed by the histogram of Figure 5(left).

The shape of a HR is an important point to decide whether a particular distribution is suitable or not for a data set. A manner to characterize the shape of a HR is by means of the scaled total time on test (TTT) function. We can detect the type of HR that the data have and then choose a suitable distribution. Let $h(t) = f(t)/[1-F(t)]$ be the HR of a RV T, where $f(\cdot)$ and $F(\cdot)$ are the PDF and CDF of T. Then, the TTT function is given by $W(u) = H^{-1}(u)/H^{-1}(1)$, for $0 \leq u \leq 1$, where $H^{-1}(u) = \int_0^{F^{-1}(u)} [1-F(z)]\text{d}z$, with $F^{-1}(\cdot)$ denoting the inverse CDF of T. A plot of the points $[k/n, W_n(k/n)]$ can approximate $W(\cdot)$, with $W_n(k/n) = [\sum_{i=1}^{k} t_{(i)} + \{n-k\}t_k]/\sum_{i=1}^{n} t_{(i)}$, for $k = 1, \ldots, n$, and $t_{(i)}$ denoting the ith observed order statistic; see, for example, Figure 1 in Azevedo et al. (2012) for different theoretical shapes for the scaled TTT curves. Figure 5(center)

suggests a decreasing HR for the observed lifetime. Therefore, the Weibull distribution is a good choice as a baseline function, since it allows us to model constant, increasing and decreasing HR. Moreover, this distribution is one of the most used models in survival and reliability analysis due to its good properties and flexibility in data modeling.

Figure 5(right) presents the usual and adjusted boxplots. The latter is important in cases where the data follow a skewed distribution, since a significant number of observations can be classified as atypical when they are not; see Hubert and Vandervieren (2008). From Figure 5(right), we note that potential outliers considered by the usual boxplot are not outliers when we consider the adjusted boxplot.

Table 3 – Descriptive statistics for the observed lifetime.

$t_{(1)}$	MD	$\bar{t}$	SD	CV	CS	CK	$t_{(n)}$	n
1.00	22.00	40.88	46.70	1.14	1.16	3.12	156	33

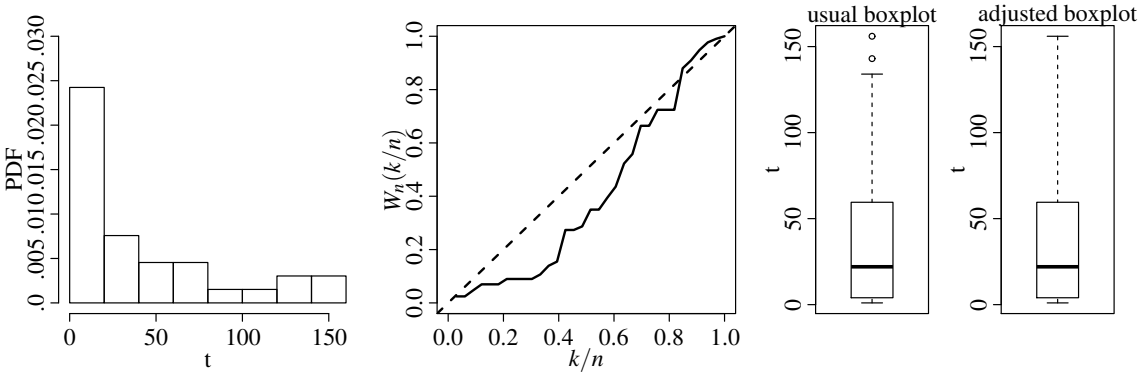

Figure 5 – Histogram, TTT plot and boxplots for the leukemia cancer data

Estimation and model selection

Table 4 reports the ML estimates of the BS frailty model from leukemia cancer data. The results of the Weibull model without frailty, the GA and IG frailty model are given as well. In this table, values for Akaike (AIC) and Bayesian (BIC) information criteria, are provided. Table 4 indicates that the BS frailty model with Weibull baseline provides the best overall fit in terms of AIC and BIC. The estimated variance of the BS, GA and IG frailty models are, respectively, given by

$$\widehat{\mathrm{Var}}(U) = \frac{2\widehat{\delta}+5}{(\widehat{\delta}+1)^2}, \quad \widehat{\mathrm{Var}}(U) = \widehat{\zeta} \quad \text{and} \quad \widehat{\mathrm{Var}}(U) = \widehat{\sigma}^2. \tag{3.11}$$

Table 4 – ML estimates (with –SEs– in parentheses) and model selection measures for the fit to leukemia data.

Parameter	Weibull	BS frailty	GA frailty	IG frailty
γ	0.2804 (0.2370)	0.2289 (0.0547)	0.0285 (0.0542)	0.1215 (0.0429)
κ	1.2878 (0.1380)	2.8283 (0.5312)	0.7888 (0.6958)	1.3909 (0.1804)
δ	-	0.0184 (0.0175)	-	-
ζ	-	-	0.0176 (2.1091)	-
σ^2	-	-	-	12.1083 (5.1661)
log-likelihood	−153.6000	−149.2648	−153.6221	−152.8119
AIC	311.1737	304.5297	313.2455	311.6238
BIC	318.6628	309.0192	317.7350	316.1133

Based on Table 4 and Equation (3.11), the estimated frailty variances from the BS, GA and IG frailty models are 4.8564, 0.0176 and 12.1083, respectively. Notice that overall models capture the unobserved heterogeneity and from Table 4 the BS model provides a better fit compared to the GA and IG models.

In Figure 6, we present the fitted SF by Kaplan-Meier (KM). It allows us to compare the empirical SF of the data for the BS, GA and IG frailty models. It is important to highlight that the overall results suggest that the BS frailty model is better than the GA and IG frailty models, indicating the potential of the new model in describing frailty data.

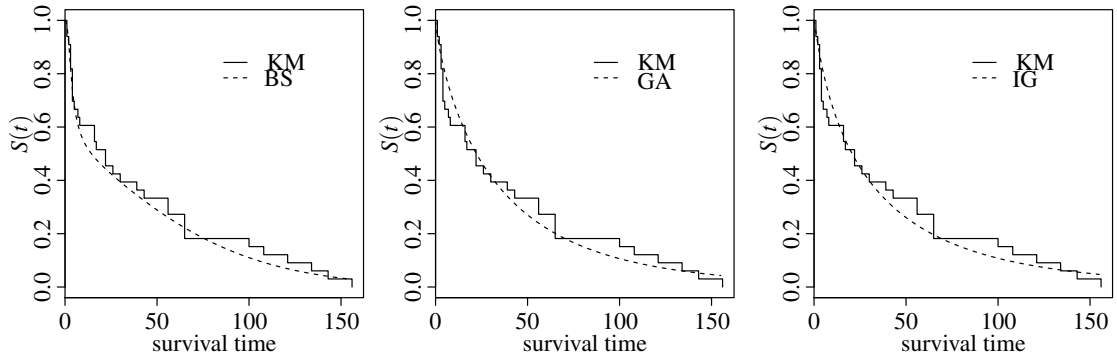

Figure 6 – Fitted SFs by KM method and BS, GA and IG frailty models for the leukemia cancer data.

Figure 7 shows the QQ plots with simulated envelope for the GCS and QS residuals. These plots allow us to check graphically whether the GCS and QS residuals follow the EXP(1) and standard normal distributions or not, respectively. From Figure 7, we note that these residuals present a good agreement with their corresponding target distributions, respectively.

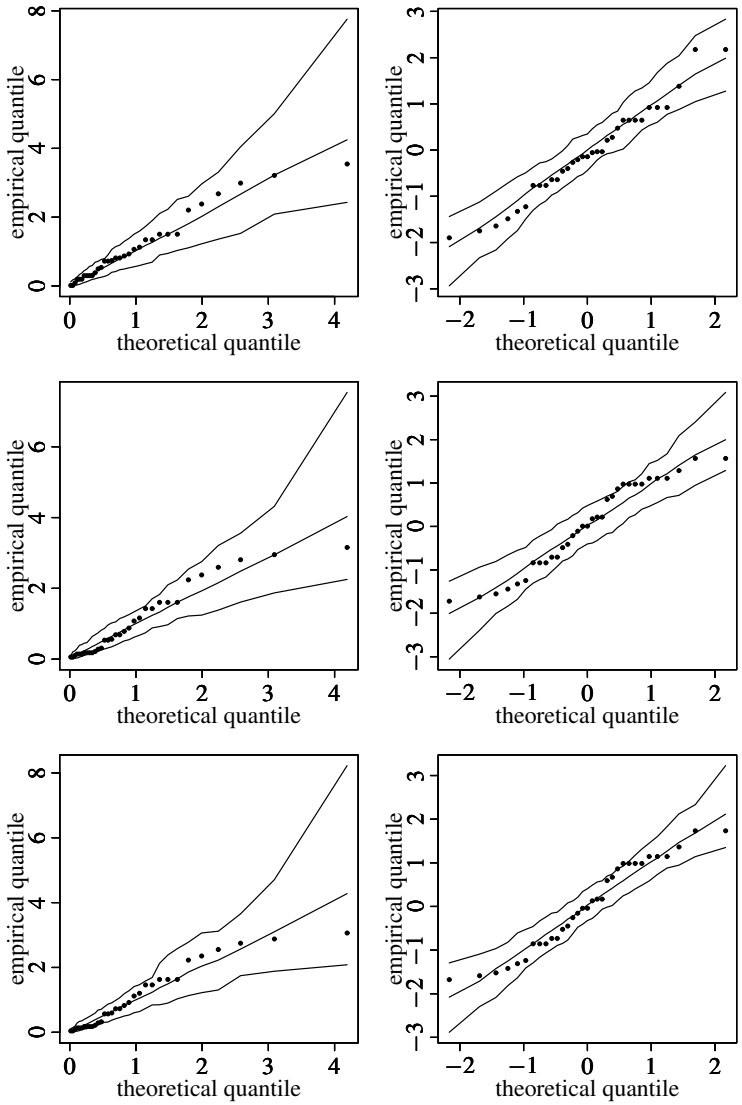

Figure 7 – QQ plot with envelope for GCS and RQ residuals for the BS, GA and IG frailty models, respectively, for leukemia cancer data

3.4.2 Second case study: lung cancer data

The data set corresponds to the survival times on 137 advanced lung cancer patients from the Veterans' Administration lung cancer trial; see Kalbfleisch and Prentice (2011). The explanatory variables recorded when the patient is taken on study include: performance status; a measure of general medical status; age (years); time in months from diagnosis to starting on study; and previous therapy. The percentage of censored observations was 6.57%.

Exploratory data analysis

Table 5 reports the descriptive statistics of the observed lifetime (in days) from the lung cancer trial. The CK and CS indicate the positively skewed nature and high kurtosis level of the data distribution.

Table 5 – Descriptive statistics for the observed lifetime.

MD	$\bar{t}$	SD	CV	CS	CK	$t_{(1)}$	$t_{(n)}$	n
80.00	121.60	157.82	1.40	3.13	15.55	1.00	999.00	137

Figure 8 shows the TTT plot, histogram and boxplots for the lung cancer data. Note that the skewed nature is confirmed by the histogram of Figure 8(left). The TTT plot displayed in Figure 8(center) suggests a decreasing HR for the observed lifetime, which justifies the use of the Weibull distribution as a baseline function. From Figure 8(center), we note that potential outliers considered by the usual boxplot are not outliers when we consider the adjusted boxplot.

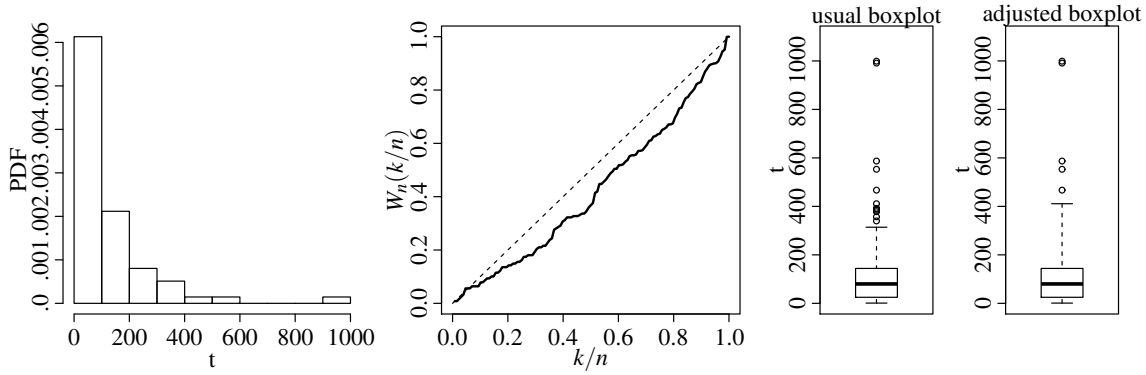

Figure 8 – Histogram, TTT plot and boxplots for the lung cancer data

Estimation and model selection

The estimates of the model parameters (with SEs in parentheses), AICs and BICs are summarized in Table 6. Comparing the information criteria, we notice that the BS frailty model has the smallest AIC and BIC values, suggesting that it is a good fit for this data.

Table 6 – ML estimates (with estimated asymptotic standard errors –SEs– in parentheses) and model selection measures for the fit to lung cancer data.

Parameter	Weibull	BS frailty	GA frailty	IG frailty
γ	0.2087 (0.1078)	0.0136 (0.0059)	0.0100 (0.0018)	0.0232 (0.0071)
κ	1.1735 (0.0669)	1.1468 (0.2782)	0.9749 (0.1193)	1.4177 (0.1524)
δ	-	2.0174 (2.7096)	-	-
ζ	-	-	0.2405 (0.2139)	-
σ^2	-	-	-	4.9577 (3.7818)
log-likelihood	−748.1000	−746.8067	−747.1860	−790.0269
AIC	1500.2000	1499.6130	1500.3720	1586.0540
BIC	1506.0400	1508.3730	1509.1320	1594.8140

The QQ plots with simulated envelope for the GCS and QS residuals are displayed in Figure 9. These graphical plots show the notorious superiority, in terms of fitting to the data, of the BS frailty model with Weibull baseline over all other models.

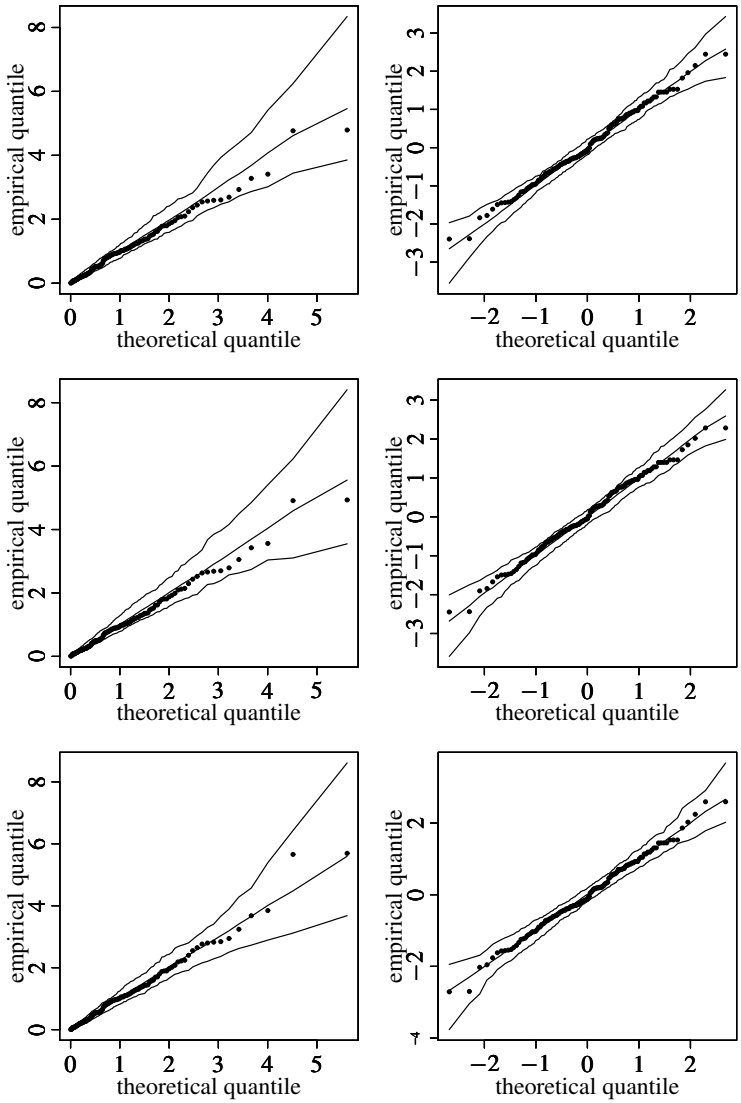

Figure 9 – QQ plot with envelope for GCS and RQ residuals for the BS, GA and IG frailty models, respectively, with lung cancer data.

The estimated frailty variances from BS, GA and IG frailty models are 0.9970, 0.2405 and 4.9577, respectively. The IG frailty model gives larger variance compared to the BS and GA frailty model, but overall the estimated frailty variances are different from zero indicating the models capture unobserved heterogeneity. However from Table 6 the BS model provides a better fit compared with GA and IG models.

Figure 10 shows the fitted SF by KM. The fitted SFs presented in Figures 10 these graphical plots show the notorious agreement, in terms of fitting to the data, of the BS frailty model.

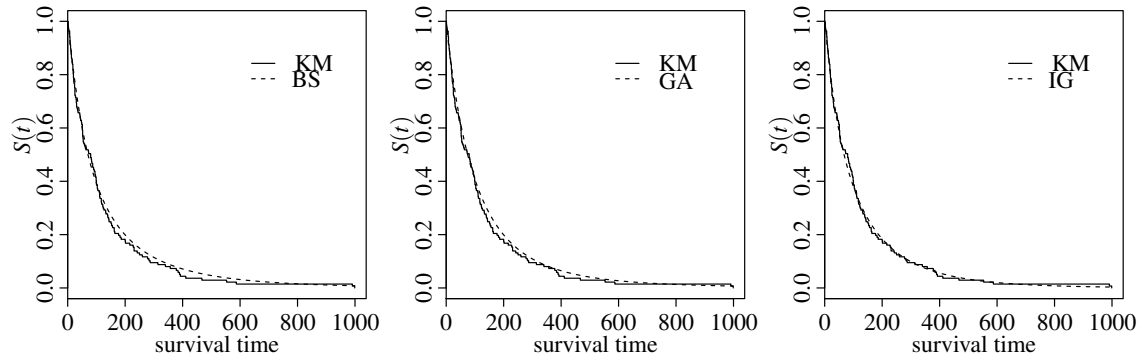

Figure 10 – Fitted SFs by KM method and BS, GA and IG frailty models for the lung cancer data.

3.5 Concluding remarks

In this chapter we have introduced a new frailty model based on reparameterized Birnbaum-Saunders distribution. The Birnbaum-Saunders distribution is employed to describe the unobserved frailty. Due to its genesis, justification, properties and features, the new model can be a good alternative in frailty modeling. A Monte Carlo simulation study has shown that the maximum likelihood estimators tend to their true values of the model parameters and their distributions to normality, when the sample size increased, as expected. We have proposed two types of residuals to assess the goodness-of-fit. Also, we have derived analytically the observed information matrix, which facilitates the direct computation of the corresponding estimated asymptotic standard errors. We have applied the proposed model to two real-world data sets, corresponding to the survival times of patients who died from acute myelogenous leukemia, and the survival times on advanced lung cancer patients, respecitvely. The applications have shown the potentiality of the new model. The precision parameter of the Birnbaum-Saunders distribution, which measures the effect of the time in the Birnbaum-Saunders frailty model, has shown to be significant and positive. Then, we have the effect of increasing the individual hazard rate, anticipating the occurrence of death, which does not happen in the model without frailty.

CHAPTER

4

BIRNBAUM-SAUNDERS FRAILTY REGRESSION MODELS: DIAGNOSTICS AND APPLICATION TO MEDICAL DATA

4.1 Introduction

The objectives of this chapter are: (i) to introduce a BS frailty regression model and its inference based on ML methods; and (ii) to derive influence diagnostics tools for this model. Diagnostic analysis is an efficient way to detect influential cases and evaluate their effect on the model inference. To the best of our knowledge, influence diagnostic tools have not been considered in frailty models. The natural method to assess the effect of an observation on the estimation is the case deletion, which is considered as a global influence method. However, it excludes all information from a case and can be hard to known whether that case has some influence on a specific aspect of the model. To overcome this problem, one can use the local influence method; see Cook (1987). This method allows us to detect locally influential cases and provides a sensitivity measure under perturbations on the data or the model. The local influence method has been extended to various regression models; see Osorio et al. (2007), Espinheira et al. (2008), Paula et al. (2009) and Leiva et al. (2014).

4.2 Birnbaum-Saunders frailty regression model

By taking (2.24) and the Laplace transform of the BS distribution, given in (3.1), with parameters $(1,\delta)$ and evaluated at $s = H_0(t)\exp(\eta)$, where $\eta = \boldsymbol{x}^\top\boldsymbol{\varphi}$ is the linear predictor, $\boldsymbol{\varphi} = (\varphi_1,\ldots,\varphi_p)^\top$ is a vector of regression parameters and $H_0(t)$ indicates the baseline CHR,

we obtain the unconditional SF under the BS frailty with values $\boldsymbol{x}$ for its covariates as

$$S_T(t;\boldsymbol{x},\boldsymbol{\xi}) = \frac{\exp\left(\frac{\delta}{2}\left(1-\frac{\sqrt{\delta+4H_0(t;\boldsymbol{\xi}_1)\exp(\eta)+1}}{\sqrt{\delta+1}}\right)\right)\left(\sqrt{\delta+4H_0(t;\boldsymbol{\xi}_1)\exp(\eta)+1}+\sqrt{\delta+1}\right)}{2\sqrt{\delta+4H_0(t;\boldsymbol{\xi}_1)\exp(\eta)+1}}. \tag{4.1}$$

Then, from (3.1) and (4.1), the corresponding unconditional HR of T is

$$h_T(t;\boldsymbol{x},\boldsymbol{\xi}) = h_0(t;\boldsymbol{\xi}_1)\exp(\eta)\times \frac{\delta(\delta+\sqrt{\delta+1}\sqrt{\delta+4H_0(t;\boldsymbol{\xi}_1)\exp(\eta)+1}+4H_0(t;\boldsymbol{\xi}_1)\exp(\eta)+3)+2}{(\delta+4H_0(t;\boldsymbol{\xi}_1)\exp(\eta)+1)(\delta+\sqrt{\delta+1}\sqrt{\delta+4H_0(t;\boldsymbol{\xi}_1)\exp(\eta)+1}+1)}. \tag{4.2}$$

We assume that h_0 is specified up to a few unknown parameters, which are related to a distribution assumed for the baseline HR. For example, we can suppose an exponential, LN or Weibull distribution, among others. However, assuming a parametric distribution is not always desirable, because such an assumption is often difficult to verify. Note that the Weibull distribution has been extensively used to model the baseline HR due to its flexibility or when the HR must be constant, increasing or decreasing for each patient; see Lawless (2011). Therefore, we use the Weibull distribution as baseline hazard, which has $h_0(t;\gamma,\kappa) = \gamma\kappa t^{\kappa-1}$ and $H_0(t;\gamma,\kappa) = \gamma t^{\kappa}$, for $t>0$, where $\kappa>0$ and $\gamma>0$ are shape and scale parameters, respectively. Note that the baseline HR h_0: (i) increases if $\kappa>1$; (ii) is constant (exponential model) if $\kappa=1$; and (iii) decreases if $\kappa<1$. This parameterization is commonly used in statistical models for medicine; see Collett (2015). Thus, from (4.1)-(4.2) and for $\boldsymbol{\xi}=(\gamma,\kappa,\delta,\eta)^\top$, the unconditional HR and SF of T under BS frailty reduce to

$$\begin{aligned} h_T(t;\boldsymbol{x},\boldsymbol{\xi}) &= \frac{\gamma\kappa t^{\kappa-1}\exp(\eta)\,(\delta(\delta+\Delta(t;\boldsymbol{\xi})+4\gamma t^{\kappa}\exp(\eta)+3)+2)}{\Delta^2(t;\boldsymbol{\xi})(\delta+\Delta(t;\boldsymbol{\xi})+1)}, \\ S_T(t;\boldsymbol{x},\boldsymbol{\xi}) &= \frac{\exp((\delta/2)(1-\Delta^*(t;\boldsymbol{\xi})/\sqrt{\delta+1}))(\Delta^*(t;\boldsymbol{\xi})+\sqrt{\delta+1})}{2\Delta^*(t;\boldsymbol{\xi})}, \end{aligned} \tag{4.3}$$

where $\Delta(t;\boldsymbol{\xi}) = \sqrt{(\delta+1)(\delta+4\gamma t^{\kappa}\exp(\eta)+1)}$ and $\Delta^*(t;\boldsymbol{\xi}) = \sqrt{\delta+4\gamma t^{\kappa}\exp(\eta)+1}$.

4.2.1 Estimation of parameters

Suppose that the lifetime is not completely observed and may be subject to right censoring. Then, from n patients with pairs of lifetimes and censoring indicators $(t_1,\varsigma_1),\ldots,(t_n,\varsigma_n)$, the corresponding log-likelihood function under uninformative censoring, taken from (2.21), can be expressed as

$$\ell(\boldsymbol{\xi}) = \log(L(\boldsymbol{\xi})) = \sum_{i=1}^{n}\ell_i(\boldsymbol{\xi}), \tag{4.4}$$

where

$$\begin{aligned}\ell_i(\boldsymbol{\xi}) &= \varsigma_i(\log(\kappa)+\log(\gamma)+(\kappa-1)\log(t_i)+\eta)+(\delta/2)(1-(\Delta^*(t_i;\boldsymbol{\xi}))/\sqrt{\delta+1}) \\ &\quad +\log(\Delta^*(t_i;\boldsymbol{\xi})+\sqrt{\delta+1})-\log(2\Delta^*(t_i;\boldsymbol{\xi})) \qquad (4.5)\\ &\quad +\varsigma_i\log(\delta(\delta+\Delta(t_i;\boldsymbol{\xi})+4\gamma t_i^{\kappa}\exp(\eta)+3)+2)-2\varsigma_i\log(\Delta^*(t_i;\boldsymbol{\xi})(\delta+\Delta(t_i;\boldsymbol{\xi})+(4.6)\end{aligned}$$

with ς_i and t_i being the elements of the vectors $\boldsymbol{\varsigma}$ and $\boldsymbol{t}$, respectively, with the latter defined in (2.20), $\boldsymbol{\xi}=(\gamma,\kappa,\delta,\boldsymbol{\varphi}^\top)^\top$ denoting here the parameter vector of the BS frailty regression model given in (4.3), and $\eta=\boldsymbol{x}^\top\boldsymbol{\varphi}$ defined in (2.20). The parameter vector $\boldsymbol{\xi}$ may be estimated by numerical maximization of the log-likelihood function ℓ given in (4.4) using the `R` software by its functions `optim` and `optimx`; see <www.R-project.org> and R Core Team (2016).

It can be verified that standard regularity conditions; see for example Cox and Hinkley (1974) are fulfilled for the proposed model, whenever the parameters are within the parameter space. Then, the ML estimator $\widehat{\boldsymbol{\xi}}$ is consistent and follows a normal asymptotic joint distribution with asymptotic mean $\boldsymbol{\xi}$, and an asymptotic covariance matrix $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\xi}})$ that can be obtained from the corresponding expected Fisher information matrix. Thus, recalling that $\boldsymbol{\xi}=(\gamma,\kappa,\delta,\boldsymbol{\varphi}^\top)^\top$, we have, as $n\to\infty$,

$$\sqrt{n}(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi}) \quad \xrightarrow{D} \quad \mathrm{N}_{p+4}(\mathbf{0}_{(p+4)\times 1},\boldsymbol{\Sigma}(\widehat{\boldsymbol{\xi}})=\mathscr{J}(\boldsymbol{\xi})^{-1}), \qquad (4.7)$$

where $\mathbf{0}_{(p+4)\times 1}$ is a $(p+4)\times 1$ vector of zeros, $\mathscr{J}(\boldsymbol{\xi})=\lim_{n\to\infty}(1/n)\mathscr{I}(\boldsymbol{\xi})$, with $\mathscr{I}(\boldsymbol{\xi})$ being the corresponding expected Fisher information matrix. Note that $\widehat{\mathscr{I}}(\boldsymbol{\xi})^{-1}$ is a consistent estimator of the variance-covariance matrix of $\widehat{\boldsymbol{\xi}}$, $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\xi}})$ namely. In practice, one may approximate the expected Fisher information matrix by its observed version, whereas the diagonal elements of its inverse matrix can be used to approximate the corresponding standard errors (SEs); see Efron and Hinkley (1978) for the use of observed versus expected Fisher information matrices. Besides estimation, hypothesis testing is another important topic to be addressed. Let $\boldsymbol{\xi}^*$ be a proper disjoint subset of $\boldsymbol{\xi}$. We aim to test H_0: $\boldsymbol{\xi}^*=\boldsymbol{\xi}_0^*$ versus H_1: $\boldsymbol{\xi}^*\neq\boldsymbol{\xi}_0^*$. Also, let $\widehat{\boldsymbol{\xi}}_0^*$ maximize $\ell(\boldsymbol{\xi}^*)$ given in (4.4) constrained to H_0. Then, the corresponding likelihood ratio (LR) statistic is $\mathrm{LR}=2\log(\ell(\widehat{\boldsymbol{\xi}}^*)/\ell(\widehat{\boldsymbol{\xi}}_0^*))$. Under H_0 and some regularity conditions, that is, conditions needed for the asymptotic theory of ML estimators to hold Serfling (1980), the distribution of the LR statistic converges to the $\chi^2(\varpi)$ distribution, with $\varpi=\dim(\boldsymbol{\xi}^*)$ being the dimension of the vector $\boldsymbol{\xi}^*$.

4.2.2 A simulation study

In this subsection we conduct a simulation study to assess the performance of the ML estimators of the proposed model. Moreover, we evaluate the cost of estimating φ. In the simulation study, we consider four different simple sizes $n\in\{50,100,200,400\}$, 5000 MC replications, values of true parameter $\delta\in\{0.25,0.50,1.00\}$, and without of loss of generality we assume $\kappa=0.5$ and $\gamma=1.00$ in all cases. The random numbers of the covariate are generated

from the Binomial distribution with $p = 0.5$, and the regression coefficient is $\varphi = \log(2)$. In this case, the percentage of censored observations were fixed at 30%. The random numbers from the BS frailty regression model can be obtained by Algorithm 2.

Algorithm 2 – Generator of random numbers from the BS frailty regression model.

1: Obtain a random number x_i from $X \sim \text{Bernoulli}(1, 0.5)$.
2: Set a value for $\boldsymbol{\xi} = (\gamma, \kappa, \delta, \boldsymbol{\varphi}^\top)^\top$.
3: Calculate $\eta_i = x_i\varphi$ and determine expressions for $\Delta^*(y_i;\boldsymbol{\xi})$ and $\Delta(y_i;\boldsymbol{\xi})$ as a function of y_i based on the formulas given in (4.3).
4: Generate a random number m_i from $M \sim \text{U}(0,1)$.
5: Equate m_i to the SF defined in (4.3) and get the time to event of interest y_i by solving numerically the equation

$$m_i = \frac{\exp\left(\frac{\delta}{2}\left(1-\Delta^*(y_i;\boldsymbol{\xi})/\sqrt{\delta+1}\right)\right)\left(\Delta^*(y_i;\boldsymbol{\xi})+\sqrt{\delta+1}\right)}{2\Delta^*(y_i;\boldsymbol{\xi})}.$$

6: Establish the censoring time v_i from $V \sim \text{U}(a,b)$, for $a > 0$ and $b > 0$ fixed.
7: Find a random number $t_i = \min\{y_i, v_i\}$, that is,
7.1: If $y_i < v_i$, then $\varsigma_i = 1$ and $t_i = y_i$;
7.2: Else, $\varsigma_i = 0$ and $t_i = v_i$.
8: Repeat Steps 1 to 7 until the amount of n random numbers to be completed.

In Table 7 we report the empirical bias and MSE of the ML estimators. From this table we note that, as the sample size increases, the bias decreases, indicating that the ML estimators become more efficient. From this table we can also note that, increases in the values of δ tend to increase the bias of $\widehat{\delta}$, $B(\widehat{\delta})$. A similar behavior from the original parameterization of the BS distribution occurs; see for example Lemonte et al. (2007).

Estimation cost φ

When frailty models are considered, a question that may arise is the estimation cost of certain parameters with the inclusion of random effect, namely, considering the frailty BS model, one may want to know what is the cost (variance inflation) to estimate the covariate coefficient φ. Several authors have been studied the estimation cost of an additional parameter; see for example Bickel and Doksum (1981), Siqueira and Taylor (1999), Calsavara et al. (2016). To conduct this simulation study we consider $n \in \{50, 100, 200, 400\}$, 5000 MC replications, values of true parameter $\delta \in \{0.25, 0.50, 0.75, 1.00\}$, and without of loss of generality we assume $\kappa = 0.5$ and $\gamma = 1.00$ in all cases. The random numbers of the covariate from Binomial distribution with $p = 0.5$, with regression coefficient $\varphi_1 = \log(2)$. Let $\boldsymbol{\xi} = (\delta, \kappa, \gamma, \boldsymbol{\varphi}^\top)^\top$ parameters vector of the BS frailty regression model and the ML estimates $\widehat{\boldsymbol{\xi}} = (\widehat{\delta}, \widehat{\kappa}, \widehat{\gamma}, \widehat{\boldsymbol{\varphi}}^\top)^\top$ with observed information matrix denoted by $\widehat{\mathscr{J}}(\widehat{\boldsymbol{\xi}})$. Let $\widehat{\boldsymbol{\xi}}^* = (\widehat{\kappa}^*, \widehat{\gamma}^*, \widehat{\varphi}^*)^\top$ the ML estimates with δ known (or fixed), the observed matrix is denoted by $\widehat{\mathscr{J}}_*(\widehat{\boldsymbol{\xi}}^*)$. Thus, we can obtain the estimation cost $\widehat{\varphi}$ in presence of δ, which represents the increase (or decrease) variance of φ when δ is present in the model.

Table 7 – Empirical bias (with MSEs in parentheses) of the ML estimators of γ, κ, δ and φ_1 from the BS frailty regression model.

		BS frailty regression ($\boldsymbol{\xi}$)			
n	δ	B($\widehat{\delta}$)	B($\widehat{\kappa}$)	B($\widehat{\gamma}$)	B($\widehat{\beta}_1$)
50	0.25	0.0093 (0.0036)	0.0201 (0.0023)	−0.0854 (0.0430)	−0.1425 (0.0607)
	0.50	0.0207 (0.0145)	0.0177 (0.0024)	−0.0724 (0.0369)	−0.1322 (0.0565)
	1.00	−0.0852 (0.0411)	0.0242 (0.0026)	0.0421 (0.0223)	−0.1224 (0.0521)
100	0.25	0.0060 (0.0032)	0.0151 (0.0018)	−0.0521 (0.0304)	−0.1152 (0.0501)
	0.50	0.0202 (0.0123)	0.0138 (0.0018)	−0.0362 (0.0243)	−0.1020 (0.0424)
	1.00	−0.0608 (0.0329)	0.0201 (0.0018)	0.0331 (0.0181)	−0.0757 (0.0322)
200	0.25	0.0026 (0.0028)	0.0107 (0.0013)	−0.0107 (0.0190)	−0.0935 (0.0372)
	0.50	0.0124 (0.0111)	0.0070 (0.0012)	0.0063 (0.0157)	−0.0774 (0.0303)
	1.00	−0.0562 (0.0302)	0.0181 (0.0014)	−0.0188 (0.0128)	−0.0467 (0.0217)
400	0.25	−0.0008 (0.0024)	0.0039 (0.0008)	−0.0017 (0.0158)	−0.0462 (0.0206)
	0.50	0.0121 (0.0102)	0.0045 (0.0009)	0.0016 (0.0139)	−0.0427 (0.0191)
	1.00	−0.0487 (0.0253)	0.0140 (0.0009)	0.0041 (0.0115)	−0.0172 (0.0141)

A measure of this cost is given by

$$\mathrm{VR}(\widehat{\varphi}) = \frac{\widehat{\mathscr{J}}_{44}^{-1}(\widehat{\boldsymbol{\xi}})}{\widehat{\mathscr{J}}_{*33}^{-1}(\widehat{\boldsymbol{\xi}}^{*})},$$

where VR is the variance ratio, $\mathscr{J}_{ii}$ is the ith diagonal element of the observed information matrix $\mathscr{J}$, evaluated at $\widehat{\boldsymbol{\xi}}$ and $\mathscr{J}_{*ii}$ is the ith diagonal element of the observed information matrix $\mathscr{J}$ evaluated in $\widehat{\boldsymbol{\xi}}^{*}$. Table 8 provides the empirical mean of the VR. From this table, we note that the estimation cost decreases when the sample size increases for all values of δ, as expected. Another point that we can note is that as the population becomes more homogeneous (large values of δ), the VR takes larger values, thus it indicates that there is an estimation cost associated with φ.

Table 8 – Empirical mean of variance ratio.

	n			
δ	50	100	200	400
0.25	1.135	1.087	1.060	1.052
0.50	1.200	1.164	1.129	1.118
0.75	1.224	1.215	1.189	1.185
1.50	1.249	1.232	1.218	1.198

4.3 Influence diagnostics and residual analysis

In this section, we introduce global of local influence techniques and two types of residuals for the BS frailty regression model.

4.3.1 Global influence

Global influence is related to case deletion, that is, it is a technique to study the effect of dropping a case from the data set. Consider a version for case deletion from the expressions given in (4.3), with the subscript "(i)" meaning the set of patients has the case i deleted. Consequently, the corresponding log-likelihood function defined in (4.4) is now denoted by $\ell_{(i)}$. Let $\widehat{\boldsymbol{\xi}}_{(i)} = (\widehat{\gamma}_{(i)}, \widehat{\kappa}_{(i)}, \widehat{\delta}_{(i)}, \widehat{\boldsymbol{\varphi}}_{(i)}^{\top})^{\top}$ be the ML estimate of $\boldsymbol{\xi}$ from $\ell_{(i)}$. To assess the influence of the case i on the ML estimate $\widehat{\boldsymbol{\xi}} = (\widehat{\gamma}, \widehat{\kappa}, \widehat{\delta}, \widehat{\boldsymbol{\varphi}}^{\top})^{\top}$, the basic idea is to compare the difference between $\widehat{\boldsymbol{\xi}}_{(i)}$ and $\widehat{\boldsymbol{\xi}}$ in terms of $\ell_{(i)}$ and ℓ, respectively. If deletion of a case seriously influences the estimates, more attention should be paid to that case. Hence, if $\widehat{\boldsymbol{\xi}}_{(i)}$ is far from $\widehat{\boldsymbol{\xi}}$, the case i is regarded as potentially influential. A first measure of global influence is defined as the standardized norm of $\widehat{\boldsymbol{\xi}}_{(i)} - \widehat{\boldsymbol{\xi}}$, known as the generalized Cook distance (GCD), given by $\mathrm{GCD}_i(\boldsymbol{\xi}) = (\widehat{\boldsymbol{\xi}}_{(i)} - \widehat{\boldsymbol{\xi}})^{\top}(\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\xi}}))^{-1}(\widehat{\boldsymbol{\xi}}_{(i)} - \widehat{\boldsymbol{\xi}})$, where $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\xi}})$ is an estimate of $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\xi}})$ obtained from (4.7). An alternative way is to assess $\mathrm{GCD}_i(\gamma)$, $\mathrm{GCD}_i(\kappa)$, $\mathrm{GCD}_i(\delta)$ and $\mathrm{GCD}_i(\boldsymbol{\varphi})$, whose values reveal the impact of the case i on the estimates of γ, κ, δ and $\boldsymbol{\varphi}$, respectively. Also, $\widehat{\boldsymbol{\xi}}_{(i)}$ and $\widehat{\boldsymbol{\xi}}$ can be compared by their likelihood distance (LD) defined as $\mathrm{LD}_i(\boldsymbol{\xi}) = 2(\ell(\widehat{\boldsymbol{\xi}}) - \ell(\widehat{\boldsymbol{\xi}}_{(i)}))$, for $i = 1, \ldots, n$.

4.3.2 Local influence

Local influence is based on the curvature of the plane of the log-likelihood function. Consider the BS frailty regression model given in (4.3), recall $\boldsymbol{\xi} = (\gamma, \kappa, \delta, \boldsymbol{\varphi}^{\top})^{\top}$ and let $\ell(\boldsymbol{\xi}; \boldsymbol{\omega})$ be the log-likelihood function corresponding to this model defined in (4.4) but now perturbed by $\boldsymbol{\omega}$. The vector of perturbations $\boldsymbol{\omega}$ belongs to a subset $\Omega \in \mathbb{R}^n$ and $\boldsymbol{\omega}_0$ is a non-perturbed $n \times 1$ vector, such that $\ell(\boldsymbol{\xi}; \boldsymbol{\omega}_0) = \ell(\boldsymbol{\xi})$, for all $\boldsymbol{\xi}$. In this case, the LD is $\mathrm{LD}(\boldsymbol{\xi}) = 2(\ell(\widehat{\boldsymbol{\xi}}) - \ell(\widehat{\boldsymbol{\xi}}_{\omega}))$, where $\widehat{\boldsymbol{\xi}}_{\omega}$ denotes the ML estimate of $\boldsymbol{\xi}$ upon the perturbed BS frailty regression model, that is, $\widehat{\boldsymbol{\xi}}_{\omega}$ is obtained from $\ell(\boldsymbol{\xi}; \boldsymbol{\omega})$. Note that $\ell(\boldsymbol{\xi}; \boldsymbol{\omega})$ can be used to assess the influence of the perturbation on the ML estimate. Cook (1987) showed that the normal curvature for $\widehat{\boldsymbol{\xi}}$ in the direction $\boldsymbol{d}$, with $||\boldsymbol{d}|| = 1$, is expressed as $C_{\boldsymbol{d}}(\widehat{\boldsymbol{\xi}}) = 2|\boldsymbol{d}^{\top}\boldsymbol{\nabla}^{\top}\boldsymbol{\Sigma}(\widehat{\boldsymbol{\xi}})^{-1}\boldsymbol{\nabla}\boldsymbol{d}|$, where $\boldsymbol{\nabla}$ is a $(p+4) \times n$ matrix of perturbations with elements $\nabla_{ji} = \partial^2\ell(\xi; \boldsymbol{\omega})/\partial\xi_j\partial\omega_i$, evaluated at $\boldsymbol{\xi} = \widehat{\boldsymbol{\xi}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, for $j = 1, \ldots, p+4$ and $i = 1, \ldots, n$. A local influence diagnostic is generally based on index plots. For example, the index graph of the eigenvector $\boldsymbol{d}_{\max}$ corresponding to the maximum eigenvalue of $\boldsymbol{B}(\boldsymbol{\xi}) = -\boldsymbol{\nabla}^{\top}\boldsymbol{\Sigma}(\boldsymbol{\xi})^{-1}\boldsymbol{\nabla}$, say $C_{d_{\max}}(\boldsymbol{\xi})$, evaluated at $\boldsymbol{\xi} = \widehat{\boldsymbol{\xi}}$, can detect those cases that, under small perturbations, exercise a high influence on $\mathrm{LD}(\boldsymbol{\xi})$.

Another important direction of interest is $\boldsymbol{d}_i = \boldsymbol{e}_{in}$, which corresponds to the direction of the case i, where $\boldsymbol{e}_{in}$ is an $n \times 1$ vector of zeros with a value equal to one at the ith position, that is, $\{\boldsymbol{e}_{in}, 1 \leq i \leq n\}$ is the canonical basis of $\mathbb{R}^n$. In this case, the normal curvature is $C_i(\boldsymbol{\xi}) = 2|b_{ii}|$, where b_{ii} is the ith diagonal element of $\boldsymbol{B}(\boldsymbol{\xi})$ given above, for $i = 1, \ldots, n$, evaluated at $\boldsymbol{\xi} = \widehat{\boldsymbol{\xi}}$. If $C_i(\widehat{\boldsymbol{\xi}}) > 2\overline{C}(\widehat{\boldsymbol{\xi}})$, where $\overline{C}(\widehat{\boldsymbol{\xi}}) = \sum_{i=1}^{n} C_i(\widehat{\boldsymbol{\xi}})/n$, it indicates the case i as potentially influential. This procedure is called total local influence of the case i and can be carried for $\boldsymbol{\xi}$ or for γ, κ, δ or $\boldsymbol{\varphi}$, which are denoted as $C_i(\boldsymbol{\xi})$, $C_i(\gamma)$, $C_i(\kappa)$, $C_i(\delta)$ and $C_i(\boldsymbol{\varphi})$, respectively.

We consider the model defined in (4.3) and its log-likelihood function given by (4.4). The elements $\nabla(\gamma)$, $\nabla(\kappa)$, $\nabla(\delta)$, and $\boldsymbol{\nabla}(\boldsymbol{\varphi}^\top)$ of the matrix $\boldsymbol{\nabla}$ for each perturbation scheme detailed below were obtained numerically.

Case-weight perturbation

Under this perturbation scheme, we evaluate whether the contributions of the cases with different weights affect the ML estimate of $\boldsymbol{\xi}$. The log-likelihood function of the perturbed BS frailty model is $\ell(\boldsymbol{\xi}; \boldsymbol{\omega}) = \sum_{i=1}^{n} \omega_i \ell_i(\boldsymbol{\xi})$, where $0 \leq \omega_i \leq 1$, $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$ and $\ell_i(\boldsymbol{\xi})$ given in (4.4).

Response perturbation

We here assume an additive perturbation on the response variable (lifetime) for the case i such that $t_i(\omega_i) = t_i + \omega_i s_T$, where $s_T = (1/\widehat{\phi})^{1/2}$ is a scale factor and $\omega_i \in \mathbb{R}$, for $i = 1, \ldots, n$. Then, the log-likelihood function is $\ell(\boldsymbol{\xi}; \boldsymbol{\omega}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\xi}; \omega_i)$, where, for $\boldsymbol{\omega}_0 = (0, \ldots, 0)^\top$, we have

$$\begin{aligned}
\ell_i(\boldsymbol{\xi}; \omega_i) \;=\; & \varsigma_i(\log(\kappa) + \log(\gamma) + (\kappa - 1)\log(t_i(\omega_i)) + \eta) + (1 - \Delta^*(t_i(\omega_i); \boldsymbol{\xi})/\sqrt{\delta + 1}) \\
& + \log(\Delta^*(t_i(\omega_i); \boldsymbol{\xi}) + \sqrt{\delta + 1}) - \log(2\Delta^*(t_i(\omega_i); \boldsymbol{\xi})) \\
& + \varsigma_i \log(\delta(\delta + \Delta(t_i(\omega_i); \boldsymbol{\xi}) + 4\gamma(t_i(\omega_i))^\kappa \exp(\eta) + 3) + 2) \\
& - 2\varsigma_i \log(\Delta^*(t_i(\omega_i); \boldsymbol{\xi})(\delta + \Delta(t_i(\omega_i); \boldsymbol{\xi}) + 1)).
\end{aligned}$$

Covariate perturbation

We consider now an additive perturbation on a specific continuous covariate, X_k say, for $k = 1, \ldots, p$, by setting $x_{ik}(\omega_i) = x_{ik} + \omega_i s_X$, where s_X is a scale factor here assumed to be the standard deviation (SD) of X_k, and $\omega_i \in \mathbb{R}$, for $i = 1, \ldots, n$. Then, the log-likelihood function is $\ell(\boldsymbol{\xi}; \boldsymbol{\omega}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\xi}; \omega_i)$, where, for $\boldsymbol{\omega}_0 = (0, \ldots, 0)^\top$ and $\eta_i(\omega_i) = \boldsymbol{x}_i^\top(\omega_i)\boldsymbol{\varphi}$,

$$\begin{aligned}
\ell_i(\boldsymbol{\xi}; \omega_i) \;=\; & \varsigma_i(\log(\kappa) + \log(\gamma) + (\kappa - 1)\log(t_i) + \eta_i(\omega_i)) + (1 - \Delta^*(t_i; \gamma, \kappa, \delta, \eta_i(\omega_i))/\sqrt{\delta + 1}) \\
& + \log(\Delta^*(t_i; \gamma, \kappa, \delta, \eta_i(\omega_i)) + \sqrt{\delta + 1}) - \log(2\Delta^*(t_i; \gamma, \kappa, \delta, \eta_i(\omega_i))) \\
& + \varsigma_i \log(\delta(\delta + \Delta(t_i; \gamma, \kappa, \delta, \eta_i(\omega_i)) + 4\gamma t_i^\kappa \exp(\eta_i(\omega_i)) + 3) + 2) \\
& - 2\varsigma_i \log(\Delta^*(t_i; \gamma, \kappa, \delta, \eta_i(\omega_i))(\delta + \Delta(t_i; \gamma, \kappa, \delta, \eta_i(\omega_i)) + 1)).
\end{aligned}$$

4.3.3 Residual analysis

In order to check the goodness-of-fit of the BS frailty regression model, we propose two types of residuals for our model. These are the GCS and RQ residuals are given respectively by

$$\begin{aligned} r_i^{GCS} &= -\log(\widehat{S}_T(t_i;\boldsymbol{x},\boldsymbol{\xi})), \\ r_i^{RQ} &= \Phi^{-1}(\widehat{S}_T(t_i;\boldsymbol{x},\boldsymbol{\xi})), \quad i=1,\ldots,n, \end{aligned} \tag{4.8}$$

where Φ^{-1} is the inverse function of the N(0, 1) CDF and $\widehat{S}_T(t_i;\boldsymbol{x})$ is the estimated SF and evaluated at the lifetime t_i, that is,

$$\widehat{S}_T(t_i;\boldsymbol{x},\boldsymbol{\xi}) = \frac{\exp\left(\frac{\widehat{\delta}}{2}\left(1-\widehat{\Delta}(t_i;\boldsymbol{\xi})\sqrt{\widehat{\delta}+1}\right)\right)\left(\widehat{\Delta}(t_i;\boldsymbol{\xi})+\sqrt{\widehat{\delta}+1}\right)}{(2\widehat{\Delta}(t_i;\boldsymbol{\xi}))},$$

with

$$\widehat{\Delta}(t_i;\boldsymbol{\xi}) = \sqrt{\widehat{\delta}+4\widehat{\gamma}t_i^{\widehat{\kappa}}\exp(\widehat{\eta})+1}.$$

If the frailty model is correctly specified, then the GCS residual has an EXP(1) distribution, regardless of the frailty model specification, whereas the RQ residual has a N(0, 1) distribution.

4.4 Applications to medical data sets

In this section, we summarize the proposed methodology by an algorithm and then apply it to two real-world medical data sets. We illustrate the methodology proposed by reanalysing the two real-world data sets used in Chapter 3, but in this section we will consider the covariates from these datas. We compare the proposed BS frailty regression model, in terms of model fitting, with the Weibull regression model and GA and IG frailty regression models, both of them having a Weibull baseline. To make sure that the GA and IG models are identifiable, we consider $U \sim \text{Gamma}(1/\zeta, 1/\zeta)$ and $U \sim \text{IG}(1,\sigma^2)$; see Wienke (2011).

4.4.1 Summary of the proposed methodology

The proposed methodology is summarized by Algorithm 3.

Algorithm 3 – Methodology based on a frailty regression model.

1: Collect n data of a response (usually lifetime), $t_1,\ldots,t_n$ say, and the values of p covariates $(x_{1i},\ldots,x_{pi})$ for the patient i associated with this response. Data can be censored or not.
2: Carry out an exploratory data analysis for identifying possible candidate models to be considered.
3: Propose a suitable frailty distribution to capture covariates which cannot be observed or measured.
4: Formulate a frailty regression model to estimate the survival probability of a patient according to the general model defined in (2.19). The formulated frailty regression model must include the response and observed covariates under analysis, as well as the frailty term by its corresponding distribution.
5: Estimate the parameters of the frailty regression model defined in Step 4 and assess the statistical significance of these parameters, as well as the presence of frailty or not by evaluating its variance.
6: Check the frailty regression model estimated in Step 5 by using quantile versus quantile (QQ) plots and GCS and RQ residuals.
7: Compare the model checking in Step 6 to other models (with frailty or not, with covariates or not, nested or not) by using the Akaike (AIC) and Bayesian (BIC) information criteria.
8: Select the best model that describes the data among the compared models in Step 7.
9: Conduct global and local diagnostic studies for the best model selected in Step 8 that describes the data to identify possible influential cases. If no influential cases are detected,
9.1: Then consider as final model to estimate the survival probability of a patient the model selected in Step 8;
9.2: Else, compute the relative change (RC) in the ML estimates of the model parameters and evaluate whether inferential changes are produced or not. If no inferential changes are detected,
9.2.1: Then consider as final model to estimate the survival probability of a patient the model selected in Step 8;
9.2.2: Else, remove the influential cases or propose a robust procedure to estimate the model parameters.

4.4.2 Application 1: leukemia data

The data set corresponds to the survival times of 33 patients, who died from acute myelogenous leukemia (a kind of cancer that often starts in the bone marrow). The measures of the patients about white blood cell count at the time of diagnosis were also recorded; see Feigl and Zelen (1965). The patients were separated into 2 groups depending on the presence or absence of a morphological characteristic of white blood cells. At the time of diagnosis, those patients with the presence of significant granulation of the leukemic cells in the bone marrow were termed as AG positive. The following variables were associated with each studied patient, for $i = 1,\ldots,33$: (i) T_i is the time from diagnosis to death (in weeks); (ii) X_{i1} is the logarithm of the white blood cell count at the time of diagnosis; and (iii) X_{i2} is the group to which they belong (1: presence – Group 1 – or 0: absence – Group 2 – of a morphological characteristic).

Estimation and model checking

We consider the BS frailty regression model with the structure: $\eta_i = \varphi_0 + \varphi_1 x_{i1} + \varphi_2 x_{i2}$. Table 9 provides the estimation and hypothesis testing results for the BS frailty regression model analyzing leukemia data. Results of the Weibull regression model without frailty and the GA

and IG frailty regression models also are detailed in this table, as well as their AIC and BIC values. It is worth to highlight that, when the models are not nested, such as our case, the AIC and/or BIC should be used to make a decision for the best-fitting model; see Wienke (2011). From Table 9, observe that the BS frailty regression model provides a better fit compared to the other models based on the values of AIC and BIC. Note that, for $\widehat{\delta} = 0.015 < 0.5$, a look at the log-BS distribution reveals bimodality, a behavior not captured by the others models; Subsection 2.4. This confirms the flexibility of the proposed model. Figure 11 shows QQ-plots with simulated envelope for both GCS and RQ residuals defined in (4.8) based on the BS frailty regression model. Figure 11 indicates that, in general, the GCS and RQ residuals present a good agreement with the EXP and N(0, 1) distributions, respectively. Moreover, the fitted SFs confirms graphically the good fit of the BS frailty regression model.

Table 9 – ML estimates (with estimated asymptotic SEs in parentheses) and model selection measures for the fit to leukemia data with Weibull baseline HR, and respective p-values in brackets.

Parameter	Weibull	BS frailty	GA frailty	IG frailty
φ_0	−1.903 (1.314)	−17.269 (3.673)	−16.793 (4.711)	−10.978 (2.991)
φ_1	1.158 (0.350)	2.975 (0.775)	3.245 (1.047)	1.945 (0.650)
p-value	[<0.0001]	[0.0001]	[0.0020]	[0.0028]
φ_2	−0.956 (0.316)	−1.829 (1.050)	−2.108 (0.827)	−1.900 (0.680)
p-value	[<0.0001]	[0.0815]	[0.0108]	[0.0052]
κ	1.001 (0.139)	3.587 (0.695)	1.743 (0.421)	1.704 (0.292)
δ	-	0.015 (0.014)	-	-
ζ	-	-	1.392 (0.827)	-
σ^2	-	-	-	8.312 (8.438)
log-likelihood	−145.100	−142.109	−144.718	−144.373
AIC	298.144	294.219	299.437	298.747
BIC	304.130	301.701	306.920	306.229

Note that the variance in the BS case increases when δ is close to zero, that is, a quite small value of δ indicates the presence of high unobserved heterogeneity. Based on Table 9 and expressions given in (3.11), we compute the corresponding frailty variances. The estimated frailty variances for the BS, GA and IG frailty regression models based on the leukemia data are 4.881, 1.392 and 8.312, respectively. This indicates the presence of unobserved heterogeneity. Thus, according to the estimated frailty variances, we conclude that the frailty models considered in this study capture the unobserved heterogeneity in the data.

Diagnostic analysis

Next, we carry out our diagnostic analysis based on global and local influence. First, Figure 12 presents the case-deletion measures $\mathrm{GCD}_i(\boldsymbol{\xi})$ and $\mathrm{LD}_i(\boldsymbol{\xi})$ discussed in Section 4.3. From this figure, on the one hand, the $\mathrm{GCD}_i(\boldsymbol{\xi})$ statistic indicates that the case #16 ($t_{16} = 5.0$, $x_{16,1} = 4.716$, $x_{16,2} = 1.0$) is potentially influential. Note that this case corresponds to a patient with a lifetime of five days (within the lowest values), a logarithm of the number of white blood cells of 4.716 and that belongs to the Group 1, that is, with presence of morphological

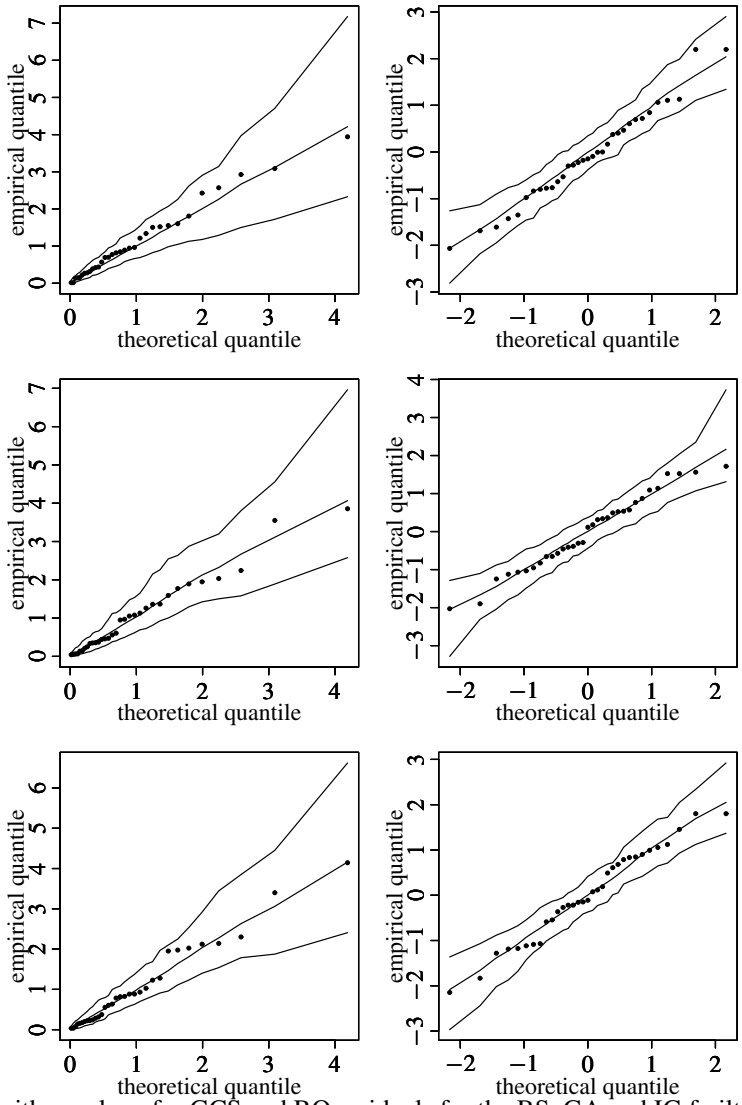

Figure 11 – QQ plot with envelope for GCS and RQ residuals for the BS, GA and IG frailty models, respectively, with leukemia data.

characteristics. On the other hand, $\mathrm{LD}_i(\boldsymbol{\xi})$ statistic does not suggest any case as potentially influential.

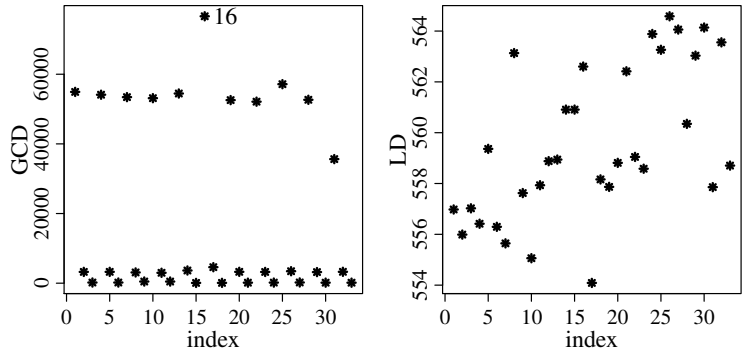

Figure 12 – Generalized Cook (left) and likelihood (right) distances for leukemia data.

Index plots of C_i for δ under the case-weight, response and covariate perturbation schemes are displayed in Figure 13 (plots corresponding to γ, κ and $\boldsymbol{\varphi}$ look very similar to that

for δ and then they are omitted here). Note that the cases #14 and #15 are detected as potentially influential on $\widehat{\delta}$, $\widehat{\kappa}$ and $\widehat{\boldsymbol{\varphi}}$ under both the case-weight and response perturbation schemes. The cases #14 and #15 correspond to the minimum values of the lifetime of patients (one week in both cases) and the maxima values of the logarithm of the number of white blood cells ($x_{14,1} = 5.0$ and $x_{15,1} = 5.0$). In addition, both of them are in the Group 1, that is, with presence of morphological characteristics. When the perturbation of the covariate X_{i1} is analyzed, observe that the case #17 ($t_{17} = 65.0$, $x_{17,1} = 5.0$; belonging to the Group 1) and the case #31 ($t_{31} = 30.0$, $x_{31,1} = 4.898$; belonging to the Group 2) are detected as potentially influential on $\widehat{\delta}$, $\widehat{\kappa}$ and $\widehat{\boldsymbol{\varphi}}$. It is important to stress that $t_{14} = 1.0$ and $t_{15} = 1.0$ represent the minimum lifetimes of the patients observed in the study and they also present the maximum values for the logarithm of the number of white blood cells ($x_{14,1} = x_{15,1} = 5.0$). In addition, the lifetimes $t_{17} = 65.0$ and $t_{31} = 30.0$ are the maxima values for the logarithm of the number of white blood cells ($x_{17,1} = 5.000, x_{31,1} = 4.897$).

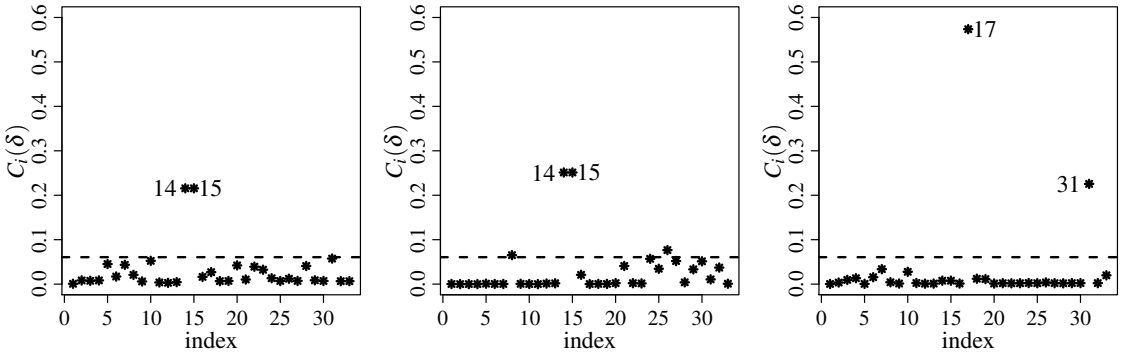

Figure 13 – Index plots of C_i for δ under the case-weight (left), response (center) and covariate (right) perturbation schemes with leukemia data.

In order to check the impact on the model inference of the detected influential cases, we implement the RC. It is computed by removing influential cases and reestimating the parameters as well as their corresponding SEs through the expressions

$$\mathrm{RC}(\xi_{j(i)}) = \left|\frac{\widehat{\xi}_j - \widehat{\xi}_{j(i)}}{\widehat{\xi}_j}\right| \times 100\%, \quad \mathrm{RC}(\mathrm{SE}(\xi_{j(i)})) = \left|\frac{\widehat{\mathrm{SE}}(\widehat{\xi}_j) - \widehat{\mathrm{SE}}(\widehat{\xi}_j)_{(i)}}{\widehat{\mathrm{SE}}(\widehat{\xi}_j)}\right| \times 100\%,$$

where $\widehat{\xi}_{j(i)}$ and $\widehat{\mathrm{SE}}(\widehat{\xi}_j)$ are the ML estimate of ξ_j and its corresponding SE, respectively, after dropping the case i, for $j = 1, \ldots, 5$ and $i = 1, \ldots, 33$, with $\xi_1 = \delta$, $\xi_2 = \kappa$, $\xi_3 = \varphi_0$, $\xi_4 = \varphi_1$ and $\xi_5 = \varphi_2$. Table 10 shows the RCs in the parameter estimates and their corresponding estimated SEs. In addition, p-values are shown for the regression coefficients. From this table, note that the largest RCs are associated with the cases #14, #15 and #31. Observe also that the significance of the parameter estimate of β_2 is altered after removing the cases #14 and #15 cases.

4.4.3 Application 2: lung cancer data

The data set corresponds to the survival times on 137 advanced lung cancer patients from a Veterans' Administration Lung Cancer trial; see Kalbfleisch and Prentice (2011). The

Table 10 – RCs (in %) in ML estimates and their corresponding SEs for the indicated parameter and dropped cases, and respective p-values in brackets with leukemia data.

Dropped case		$\widehat{\delta}$	$\widehat{\kappa}$	$\widehat{\varphi}_0$	$\widehat{\varphi}_1$	$\widehat{\varphi}_2$
{14}	RC($\xi_{j(i)}$)	29.16	13.07	0.11	5.71	49.95
	RC(SE($\xi_{j(i)}$))	(28.08)	(11.12)	(4.31)	(3.14)	(2.22)
	p-value	-	-		[<0.0001]	[<0.0001]
{15}	RC($\xi_{j(i)}$)	29.16	13.07	0.11	5.71	49.95
	RC(SE($\xi_{j(i)}$))	(28.08)	(11.12)	(4.31)	(3.14)	(2.22)
	p-value	-	-		[<0.0001]	[<0.0001]
{17}	RC($\xi_{j(i)}$)	4.02	2.22	25.43	33.80	20.10
	RC(SE($\xi_{j(i)}$))	(6.62)	(2.52)	(25.30)	(25.38)	(6.79)
	p-value	-	-		[<0.0001]	[<0.0001]
{31}	RC($\xi_{j(i)}$)	2038.38	41.40	0.49	13.28	57.08
	RC(SE($\xi_{j(i)}$))	(2756.32)	(12.01)	(37.94)	(44.12)	(6.13)
	p-value	-	-		[0.0604]	[0.0093]
{14,15}	RC($\xi_{j(i)}$)	16.49	3.02	2.06	1.70	1.65
	RC(SE($\xi_{j(i)}$))	(18.16)	(0.36)	(0.25)	(0.11)	(0.15)
	p-value	-	-		[<0.0001]	[0.0085]
{17,31}	RC($\xi_{j(i)}$)	378.91	3.77	106.86	153.38	110.68
	RC(SE($\xi_{j(i)}$))	(335.89)	(14.42)	(127.10)	(138.62)	(18.04)
	p-value	-	-		[<0.0001]	[0.0002]
{14,15,17,31}	RC($\xi_{j(i)}$)	12.44	22.59	42.75	49.96	133.95
	RC(SE($\xi_{j(i)}$))	(53.84)	(47.18)	(75.42)	(94.84)	(16.00)
	p-value	-	-		[0.0045]	[0.0005]

percentage of censored observations is 6.57%. The following variables were associated with each studied patient, for $i = 1,\ldots,137$: T_i is the lifetime (in days); X_{i1} is the Karnofsky performance score (100 = good); and the factor tumor type, that is, (X_{i2}) cell type 1 (1 = squamous, 0 = other), (X_{i3}) cell type 2 (1 = small, 0 = other), (X_{i4}) cell type 3 (1 = adeno, 0 = other) and cell type 4 (1 = large, 0 = other).

Estimation and model checking

In this case, the regression structure of the model is

$$\eta_i = \varphi_0 + \varphi_1 x_{i1} + \varphi_2 x_{i2} + \varphi_3 x_{i3} + \varphi_4 x_{i4}.$$

The ML estimates of the model parameters, AICs and BICs are reported in Table 11. The results of the information criteria indicate that the BS frailty regression model has the smallest AIC and BIC values, suggesting that it provides the best fit to this data set.

Table 11 – ML estimates (with estimated asymptotic SEs in parentheses) and model selection measures for the fit to lung cancer data, and respective p-values.

Parameter	Weibull	BS frailty	gamma frailty	IG frailty
φ_0	−1.329 (0.340)	−4.109 (0.616)	−3.625 (0.630)	−4.066(0.606)
φ_1	−0.031 (0.005)	−0.049 (0.011)	−0.057 (0.013)	−0.046 (0.009)
p-value	[<0.0001]	[<0.0001]	[<0.0001]	[<0.0001]
φ_2	0.755 (0.246)	1.016 (0.359)	0.613 (0.411)	0.995 (0.340)
p-value	[0.0020]	[0.0046]	[0.1364]	[0.0034]
φ_3	1.182 (0.285)	1.363 (0.416)	1.088 (0.435)	1.353 (0.394)
p-value	[<0.0001]	[0.0011]	[0.0123]	[0.0006]
φ_4	0.343 (0.269)	0.244 (0.405)	−0.470 (0.527)	0.275 (0.371)
p-value	[0.2010]	[0.5476]	[0.3724]	[0.4580]
κ	1.066 (0.066)	1.469 (0.228)	1.574 (0.243)	1.406 (0.178)
δ		2.043 (1.643)	-	-
ζ		-	0.886 (0.423)	-
σ^2		-	-	0.936 (0.767)
log-likelihood	−716.510	−712.930	−713.744	−713.248
AIC	1445.030	1439.861	1441.489	1440.497
BIC	1462.550	1460.301	1461.929	1460.937

The estimated frailty variances for the indicated model based on lung cancer data are 0.981, 0.886 and 0.936 for the BS, GA and IG frailty regression models, respectively. This indicates the presence of unobserved heterogeneity. Notice that a slight difference between estimated frailty variances is detected, being it in the BS model slightly greater than in the GA and IG models, indicating that the BS model captures the unobserved heterogeneity in the data in a better way. Figure 14 displays the QQ-plots with simulated envelopes for the GCS and RQ residuals and the fitted SFs based on the KM estimator, as well as on the BS, GA and IG frailty models without covariates. These graphical plots show the notorious agreement, in terms of fitting to the data, of the BS frailty regression model.

We now carry out our diagnostic analysis based on global and local influence. First, Figure 15 presents the case-deletion measures $\mathrm{GCD}_i(\boldsymbol{\xi})$ and $\mathrm{LD}_i(\boldsymbol{\xi})$. From this figure, note that the $\mathrm{GCD}_i(\boldsymbol{\xi})$ statistic indicates that the cases #74, #76, #77 and #78 are potentially influential. All these cases have type-1 cell and their lifetimes are equal to 242, 111, 1 and 587 days, respectively. Notice that $t_{77} = 1.0$ (one day), that is, it corresponds to the minimum lifetime observed value in the data set. The values of Karnofsky performance score from these cases are 50, 70, 20 and 60, respectively. From Figure 15 and the $\mathrm{LD}_i(\boldsymbol{\xi})$ statistic, observe that the cases #9, #43, #58, #64, #70, #106 and #132 are potentially influential. The cases #9 and #70 have type-1 cell, whereas the cases #43 and #106 have type-2 cell, and the other cases have type-4 cell. From theses cases, the values of the Karnofsky performance score are 50, 70, 90 and 30, respectively, where $x_{9,1} = x_{43,1} = 50$, $x_{64,1} = x_{70,1} = 90$ and $x_{106,1} = x_{132,1} = 30$. In addition, the case #78 ($t_{78} = 587$) is the maximum lifetime, whereas the case #77 ($t_{77} = 1$) is the minimum lifetime. The cases #64 as #70 have the largest values related to the Karnofsky performance

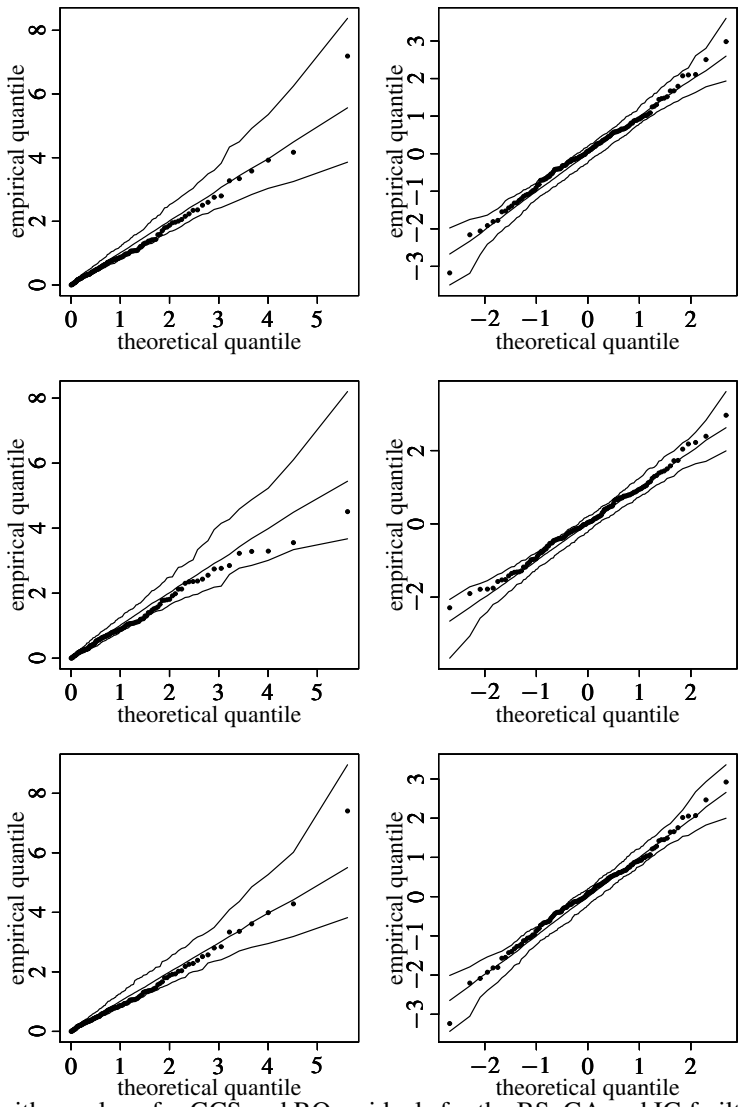

Figure 14 – QQ plot with envelope for GCS and RQ residuals for the BS, GA and IG frailty models, respectively, with lung cancer data.

score and the cases #77, #106 and #132 have the minimum value of this covariate.

Index plots of C_i for δ under the case-weight, response and covariate perturbation schemes are displayed in Figure 16 (such as in Application 1, plots corresponding to γ, κ and $\boldsymbol{\varphi}$ look very similar to that for δ and then they are omitted here). Observe that the case #44 ($t_{44} = 392$, $x_{44,1} = 40$, $x_{44,2} = 0$, $x_{44,3} = 1$, $x_{44,4} = 0$) is detected as potentially influential on $\widehat{\delta}$, $\widehat{\kappa}$ and $\widehat{\boldsymbol{\varphi}}$ under the case-weight, response and covariate perturbation schemes. This case corresponds to a patient with the maximum lifetime (three hundred ninety-two days), with almost half of the value considered a good Karnofsky score and small tumor. Regarding the response perturbation, note that the cases #77 ($t_{77} = 1.0$, $x_{77,1} = 20$, $x_{77,2} = 1$, $x_{77,3} = 0$, $x_{77,4} = 0$) and #95 ($t_{95} = 2.0$, $x_{95,1} = 40$, $x_{95,2} = 0$, $x_{95,3} = 1$, $x_{95,4} = 0$) are also detected as potentially influential on $\widehat{\delta}$, $\widehat{\kappa}$ and $\widehat{\boldsymbol{\varphi}}$. The cases #77 and #95 correspond to patients with the smallest lifetimes and present squamous and small tumors, respectively. However, the case #77 has a value of Karnofsky score situated within the smallest ones, whereas the case #95 has a value of Karnofsky score around

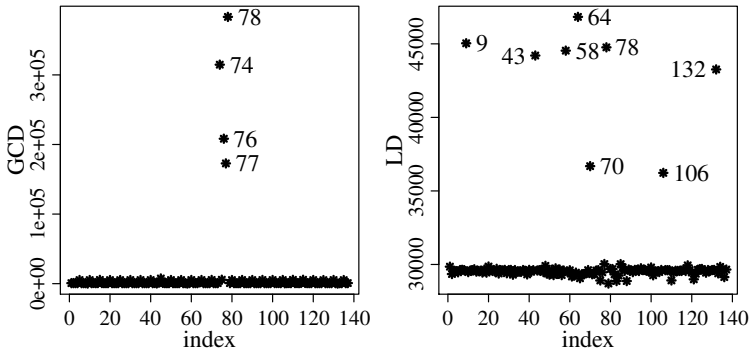

Figure 15 – Generalized Cook (left) and likelihood (right) distances.

the median value.

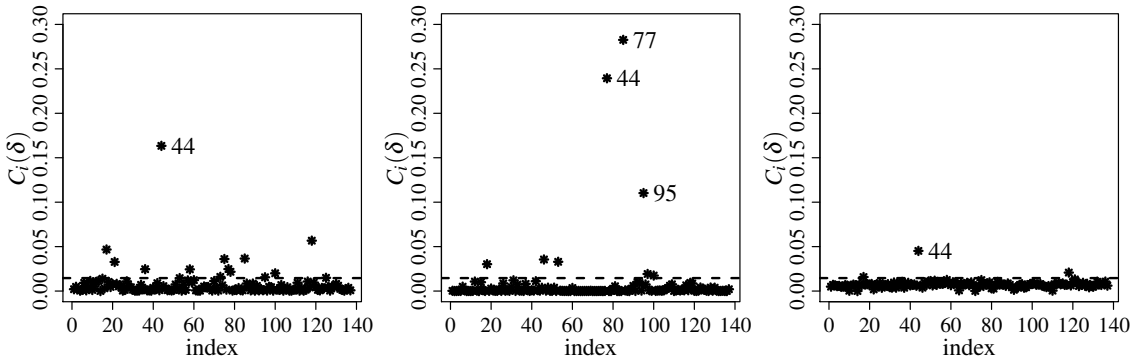

Figure 16 – Index plots of C_i for δ under the case-weight (left), response (center) and covariate (right) perturbation schemes with lung cancer data.

Table 12 shows the RCs in the parameter estimates and their corresponding estimated SEs. Also, p-values are shown for the regression coefficients. From this table, note that the largest RCs are associated with the set of cases {#44,#95} and {#44, #77, #95}. Observe also that the significance of the parameter φ_2 is altered after removing those cases.

4.5 Concluding remarks

In this chapter, we introduced a methodology based on a new regression model with Birnbaum-Saunders distribution for its frailty. In this methodology, a frailty parameter as well as covariates are included in the model to bring further information that may be useful in practice. The inclusion of covariates aims to account for differences in risk, whereas the inclusion of a frailty helps to capture unobserved heterogeneity that covariates may fail to fully account for. This may be due to a missing covariate in the model that can be explained by the frailty. The introduced methodology encompassed inference about the model parameters and influence diagnostics. Also, we proposed two types of residuals for the new frailty regression model. The methodology was summarized by an algorithm that allows a practitioner to understand it in a better form. We applied the proposed model to two real-world data sets concerning the survival times of patients who: (i) died due to acute myelogenous leukemia or (ii) had advanced lung

Table 12 – RCs (in %) in ML estimates and their corresponding SEs for the indicated parameter and dropped cases, and respective p-values in brackets with lung cancer data.

Dropped case		$\widehat{\delta}$	$\widehat{\kappa}$	$\widehat{\varphi}_0$	$\widehat{\varphi}_1$	$\widehat{\varphi}_2$	$\widehat{\varphi}_3$	$\widehat{\varphi}_4$
{9}	RC($\xi_{j(i)}$)	3.85	0.84	2.37	2.22	7.57	5.25	30.94
	RC(SE($\xi_{j(i)}$))	(16.17)	(9.41)	(2.09)	(1.21)	(1.18)	(3.91)	(9.11)
	p-value	-	-		[< 0.0001]	[< 0.0001]	[0.0002]	[0.0677]
{43}	RC($\xi_{j(i)}$)	3.57	0.21	0.42	0.49	0.71	0.21	2.04
	RC(SE($\xi_{j(i)}$))	(0.32)	(6.40)	(4.87	(3.94)	(0.57)	(0.33)	(4.05)
	p-value	-	-		[< 0.0001]	[0.0005]	[0.0001]	[0.0539]
{44}	RC($\xi_{j(i)}$)	25.17	0.98	2.10	3.11	18.10	3.19	20.06
	RC(SE($\xi_{j(i)}$))	(56.10)	(14.09)	(56.06)	(11.50)	(0.59)	(2.69)	(3.64)
	p-value	-	-		[< 0.0001]	[< 0.0001]	[< 0.0001]	[0.0438]
{58	RC($\xi_{j(i)}$)	7.96	0.80	0.69	1.74	0.29	0.08	42.12
	RC(SE($\xi_{j(i)}$))	(11.05)	(0.49)	(0.20)	(0.76)	(1.24)	(1.44)	(3.90)
	p-value	-	-		[< 0.00001]	[0.0004]	[0.0001]	[0.0374]
{64}	RC($\xi_{j(i)}$)	1.48	0.36	0.43	1.26	0.20	0.08	12.00
	RC(SE($\xi_{j(i)}$))	(14.53)	(14.33)	(1.86)	(6.42)	(1.90)	(1.76)	(4.24)
	p-value	-	-		[< 0.00001]	[0.0004]	[0.0001]	[0.0482]
{70}	RC($\xi_{j(i)}$)	0.67	0.22	0.39	2.15	4.86	3.97	27.43
	RC(SE($\xi_{j(i)}$))	(20.30)	(16.22)	(0.60)	(12.94)	(2.06)	(1.25)	(2.10)
	p-value	-	-		[< 0.0001]	[0.0006]	[0.0001]	[0.0669]
{74}	RC($\xi_{j(i)}$)	5.52	0.86	1.93	2.04	5.63	3.90	24.10
	RC(SE($\xi_{j(i)}$))	(3.81)	(3.63)	(0.53)	(2.68)	(1.52)	(0.73)	(0.35)
	p-value	-	-		[< 0.00001]	[0.0007]	[0.0001]	[0.0649]
{76}	RC($\xi_{j(i)}$)	3.84	0.91	0.10	0.77	1.67	1.06	10.53
	RC(SE($\xi_{j(i)}$))	(24.43)	(16.35)	(5.46)	(7.74)	(0.14)	(0.50)	(2.53)
	p-value	-	-		[< 0.0001]	[0.0004]	[0.0001]	[0.0495]
{77}	RC($\xi_{j(i)}$)	14.21	5.06	2.56	1.21	12.75	10.39	27.60
	RC(SE($\xi_{j(i)}$))	(15.97)	(7.52)	(10.10)	(1.39)	(4.83)	(6.43)	(3.22)
	p-value	-	-		[< 0.0001]	[< 0.0001]	[< 0.0001]	[0.0441]
{78}	RC($\xi_{j(i)}$)	0.44	0.63	1.82	0.77	9.46	6.75	40.12
	RC(SE($\xi_{j(i)}$))	(10.46)	(9.05)	(1.94)	(4.73)	(0.88)	(1.23)	(0.05)
	p-value	-	-		[< 0.0001]	[0.0010]	[0.0002]	[0.0719]
{95}	RC($\xi_{j(i)}$)	11.53	3.80	3.42	2.93	1.30	3.30	1.00
	RC(SE($\xi_{j(i)}$))	(5.39)	(13.08)	(4.30)	(9.09)	(0.07)	(1.09)	(0.30)
	p-value	-	-		[< 0.0001]	[< 0.0001]	[< 0.0001]	[0.0550]
{106}	RC($\xi_{j(i)}$)	4.13	0.43	1.11	1.89	2.01	0.25	1.16
	RC(SE($\xi_{j(i)}$))	(9.74)	(3.80)	(1.22)	(0.21)	(1.65)	(1.98)	(1.08)
	p-value	-	-		[< 0.0001]	[0.0004]	[0.0001]	[0.0548]
{132}	RC($\xi_{j(i)}$)	2.31	0.08	0.04	0.23	0.04	0.10	9.90
	RC(SE($\xi_{j(i)}$))	(10.44)	(7.35)	(0.79)	(2.17)	(1.62)	(1.77)	(1.63)
	p-value	-	-		[< 0.0001]	[0.0004]	[0.0001]	[0.0502]
{44,77}	RC($\xi_{j(i)}$)	0.04	7.39	1.18	6.49	32.26	14.01	45.48
	RC(SE($\xi_{j(i)}$))	(18.88)	(25.19)	(20.50)	(15.54)	(8.09)	(5.57)	(0.40)
	p-value	-	-		[< 0.0001]	[< 0.0001]	[< 0.0001]	[0.0367]
{44,95}	RC($\xi_{j(i)}$)	7.21	5.25	5.74	6.85	18.01	6.68	18.89
	RC(SE($\xi_{j(i)}$))	(30.76)	(20.78)	(0.94)	(20.13)	(1.63)	(0.11)	(0.70)
	p-value	-	-		[< 0.0001]	[< 0.0001]	[< 0.0001]	[0.0471]
{44,77,95}	RC($\xi_{j(i)}$)	23.38	14.65	16.23	13.05	34.60	20.31	44.09
	RC(SE($\xi_{j(i)}$))	(30.05	(13.88)	(12.52)	(9.74)	(7.91)	(8.25)	(0.03)
	p-value	-	-		[< 0.0001]	[< 0.0001]	[< 0.0001]	[0.0385]

cancer, considering uncensored and censored data, respectively. We also applied global and local influence diagnostic tools for the proposed model with both of these data sets.

The two applications illustrated the potential of the introduced methodology based on the Birnbaum-Saunders frailty regression model. From a medical point of view, it is important to adequately handle biological variation among individuals. In this sense, good parametric frailty regression models should be used more frequently in medical survival analysis. As a simple example of the applicability of the proposed methodology, one can think of medical doctors, researchers and/or practitioners estimating the survival time of a patient or group of patients in a clinical study. Moreover, the methodology introduced in this paper may be applied in a medical context to find surrogate measures (specific scores, for example, concerning the activity of daily living) or to detect frail individuals; see Wienke (2011). We implemented all functions developed in this paper in the R software. Then, the use of the introduced methodology becomes easier.

CHAPTER

5

A CURE RATE FRAILTY MODEL BASED ON A REPARAMETERIZED BIRNBAUM-SAUNDERS DISTRIBUTION

5.1 Introduction

In this chapter we introduce a cure rate BS frailty regression model and its inference based on ML methods, conduct a simulation study to evaluate the behavior of the cure fraction parameter estimator, derive influence diagnostics tools for this model and illustrate the potentiality of the proposed model with real-world data sets.

5.2 Cure rate BS frailty model

We introduce here the BS cure rate frailty model based on the cure rate model proposed by Rodrigues et al. (2009) and the BS frailty model presented in Section 3.2, where in this case we consider the Weibull distribution as baseline, which has $h_0(t;\gamma,\kappa) = \gamma\kappa t^{\kappa-1}$ and $H_0(t;\gamma,\kappa) = \gamma t^{\kappa}$, for $t > 0$, where $\kappa > 0$ and $\gamma > 0$ are shape and scale parameters, respectively.

Let the number of competing causes N follow a negative binomial distribution with parameter ϕ and θ, for $\theta > 0$ and $\phi\theta > -1$; see Piegorsch (1990). By considering (3.2) in (2.26), the long term SF of cured individuals is given by

$$S_p(t) = \left(1 + \phi\theta\left(1 - \frac{\exp\left(\frac{\delta}{2}\left(1 - \frac{\Delta(t|\boldsymbol{\xi})}{\sqrt{\delta+1}}\right)\right)\left(\Delta(t|\boldsymbol{\xi}) + \sqrt{\delta+1}\right)}{2\Delta(t|\boldsymbol{\xi})}\right)\right)^{-1/\phi}, \qquad (5.1)$$

where $\boldsymbol{\xi} = (\delta,\kappa,\gamma)^\top$ and $\Delta(t;\boldsymbol{\xi}) = \sqrt{(\delta+1)(\delta+4\gamma t^{\kappa}+1)}$. In this model, the BS frailty is used to quantify the amount of heterogeneity among non-cured individuals. The PDF and HR

obtained from (5.1) are respectively given by

$$f_p(t) = \theta h_T(t) S_T(t)\{1+\phi\theta[1-S_T(t)]\}^{-1/\phi-1}, \tag{5.2}$$

and

$$h_p(t) = \theta h_T(t) S_T(t)(1+\phi\theta(1-S_T(t)))^{-1}, \tag{5.3}$$

where $S_T(t)$ and $h_T(t)$ represent the unconditional SF and HR from the BS frailty model which has the same form given by (3.2) and (3.3). Hence, we name the model defined by (5.1), (5.2) and (5.3) as the BS cure rate negative binomial frailty (BSCrNBF) model. Particular cases of the BSCrNBF model are the BS cure rate Poisson frailty (BSCrPoF), BS cure rate Bernoulli frailty (BSCrBeF), and BS cure rate Geometric frailty (BSCrGeF) models; see Table 13.

Table 13 – survival function $S_p(t)$ and different cure rates for the distribution of hidden causes N.

BS cure rate frailty model	Parameter	Distribution (N)	$S_p(t)$
BSCrPoF	$\phi \to 0$	Poisson	$\exp\{-\theta[1-S_T(t)]\}$
BSCrBeF	$\phi = -1$	Bernoulli	$1-\theta+\theta S_T(t)$
BSCrGeF	$\phi = 1$	Geometric	$\{1+\theta S_T(t)\}^{-1}$

In order to estimate the model parameters, we utilize the ML method. Let us consider the same situation as in Subsection 2.6 when the time to event is not completely observed and is subject to right censoring. Let v_i denote the censoring time and y_i the time to event of interest. We observe $t_i = \min(y_i, v_i)$ and $\varsigma_i = \mathbb{I}(y_i \leq v_i)$, where $\varsigma_i = 1$ if $t_i = y_i$ is a time to event and $\varsigma_i = 0$, $t_i = v_i$ if it is right censored, for $i = 1,\ldots,n$. From n pairs of times and censoring indicators $(t_1,\varsigma_1),\ldots,(t_n,\varsigma_n)$, the corresponding likelihood function; see Cancho et al. (2011), under uninformative censoring, can be expressed as

$$L(\boldsymbol{t};\boldsymbol{\vartheta}) = \prod_{i=1}^{n}[f_p(t_i;\boldsymbol{\vartheta})]^{\varsigma_i}[S_p(t_i;\boldsymbol{\vartheta})]^{1-\varsigma_i}, \tag{5.4}$$

where $\boldsymbol{\vartheta} = (\phi,\theta,\boldsymbol{\xi})^\top$, $S_p(t_i;\boldsymbol{\vartheta})$ and $f_p(t_i;\boldsymbol{\vartheta})$ are given in (5.1) and (5.2), respectively. Following Castro et al. (2009), we consider the Fisher's parametrization of the NB distribution Ross and Preece (1985), and for $\phi \geq -1$, we define $\theta = (p_0^{-\phi}-1)/\phi$, if $\phi \neq 0$, and $\theta = -\log(p_0)$, if $\phi = 0$. We incorporate covariates only for the parametric cure rate model (2.27) through the cure parameter p_0. When covariates are included, we have a different cure rate parameter p_{0i} for each subject, $i = 1,\ldots,n$. To model the effects of the covariates on cure rate, we can use different functions connections; see Peng and Zhang (2008a). Defining $\boldsymbol{b}$ as the vector of parameters to be estimated for covariates associated with the fraction healing and logit link function, we have the logistic regression model

$$\log\left(\frac{p_{0i}}{1-p_{0i}}\right) = \boldsymbol{z}_i^\top\boldsymbol{b} \text{ or } p_{0i} = \frac{\exp(\boldsymbol{z}_i^\top\boldsymbol{b})}{1+\exp(\boldsymbol{z}_i^\top\boldsymbol{b})}, \tag{5.5}$$

for $i=1,\ldots,n$, where $\boldsymbol{b}=(b_0,b_1,\ldots,b_p)^\top$ stands for the vector of regression coefficients. From the variance in (2.25), we obtain $\mathrm{Var}(N_i)=\mathbb{E}(M_i)p_{i0}^{-\phi}$. Thus, extra variability in the number of competing causes due to omitted covariates is governed by the dispersion parameter ϕ. Under this relation, the improper functions in (2.27) and (2.30) can be expressed as

$$S_p(t_i;\boldsymbol{\vartheta},\boldsymbol{b})=\begin{cases}[1-(p_{0i}^{-\phi}-1)(1-S_T(t_i))]^{-1/\phi}, & \text{if } \phi\neq 0;\\ p_{0i}^{1-S_T(t_i)}, & \text{if } \phi=0.\end{cases}, \tag{5.6}$$

and

$$f_p(t_i;\boldsymbol{\vartheta},\boldsymbol{b})=\begin{cases}[1-(p_{0i}^{-\phi}-1)(1-S_T(t_i))]^{-1/\phi-1}\left(\dfrac{p_{0i}^{-\phi}-1}{\phi}\right)h_T(t_i)S_T(t_i), & \text{if } \phi\neq 0;\\ -\log(p_{0i})p_{0i}^{1-S_T(t_i)}h_T(t_i)S_T(t_i), & \text{if } \phi=0.\end{cases} \tag{5.7}$$

Let $T_1,\ldots,T_n$ be a random sample from the BSCrNBF model and $t_1,\ldots,t_n$ its observations (data). Then, the corresponding likelihood function for n individuals is defined as

$$L(\boldsymbol{\Theta};\boldsymbol{D})=\begin{cases}\prod_{i=1}^n\left[\left(\dfrac{p_{0i}^{-\phi}-1}{\phi}\right)h_T(t_i)S_T(t_i)\right]^{\varsigma_i}\left[1-(p_{0i}^{-\phi}-1)(1-S_T(t_i))\right]^{-\varsigma_i-\frac{1}{\phi}}, & \text{if } \phi\neq 0;\\ \prod_{i=1}^n\left[-\log(p_{0i})h_T(t_i)S_T(t_i)\right]^{\varsigma_i}p_{0i}^{1-S_T(t_i)}, & \text{if } \phi=0.\end{cases} \tag{5.8}$$

where $\boldsymbol{\Theta}=(\boldsymbol{\vartheta},\boldsymbol{b})^\top$ and $\boldsymbol{D}=(\boldsymbol{t},\boldsymbol{\varsigma},\boldsymbol{Z})$ and $\boldsymbol{Z}=(\boldsymbol{z}_1^\top,\ldots,\boldsymbol{z}_n^\top)$. The parameter vector $\boldsymbol{\Theta}$ may be estimated by numerical maximization of the corresponding log-likelihood function $\ell(\boldsymbol{\Theta};\boldsymbol{D})=\log(L(\boldsymbol{\Theta};\boldsymbol{D}))$, where $L(\boldsymbol{\Theta};\boldsymbol{D})$ is given in (5.8). The `R` software by its functions `optim` and `optimx` can be used to proceed with the numerical maximization; see <www.R-project.org> and R Core Team (2016).

Under some standard regularity conditions; see Cox and Hinkley (1974), the ML estimator $\widehat{\boldsymbol{\Theta}}$ is consistent and follows a normal joint asymptotic distribution with asymptotic mean $\boldsymbol{\Theta}$, and an asymptotic covariance matrix $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\Theta}})$, which can be obtained from the corresponding expected Fisher information matrix. Thus, as $n\to\infty$, we have

$$\sqrt{n}(\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta})\quad\xrightarrow{D}\quad \mathrm{N}_{p+5}(\boldsymbol{0}_{(p+5)\times 1},\boldsymbol{\Sigma}(\widehat{\boldsymbol{\Theta}})=\mathscr{J}(\boldsymbol{\Theta})^{-1}), \tag{5.9}$$

where $\boldsymbol{0}_{(p+5)\times 1}$ is a $(p+5)\times 1$ vector of zeros and $\mathscr{J}(\boldsymbol{\Theta})=\lim_{n\to\infty}(1/n)\mathscr{I}(\boldsymbol{\Theta})$, with $\mathscr{I}(\boldsymbol{\Theta})$ being the corresponding expected Fisher information matrix. In practice, the expected Fisher information matrix can be approximated by its observed version, and the square root of the diagonal elements of the inverse of this matrix can be used to approximate the corresponding standard errors; see Efron and Hinkley (1978).

5.3 Local influence

Important tools to assess the sensitivity of the ML estimators are influence methods. Local influence is based on the curvature of the plane of the log-likelihood function. Local influence calculation can be carried out for model (5.1), (5.2) and (5.3). Recall that $\boldsymbol{\Theta} = (\boldsymbol{\vartheta}, \boldsymbol{b})^\top$ and let the vector of perturbations $\boldsymbol{\omega}$ belong to a subset $\Omega \in \mathbb{R}^n$ and $\boldsymbol{\omega}_0$ be a non-perturbed n-vector, such that the log-likelihood function corresponding to the model perturbed by $\boldsymbol{\omega}$ is $\ell(\boldsymbol{\Theta}; \boldsymbol{\omega}_0) = \ell(\boldsymbol{\Theta})$, for all $\boldsymbol{\Theta}$. In this case, the likelihood displacement LD is given by $\mathrm{LD}(\boldsymbol{\Theta}) = 2(\ell(\widehat{\boldsymbol{\Theta}}) - \ell(\widehat{\boldsymbol{\Theta}}_\omega))$, where $\widehat{\boldsymbol{\Theta}}_\omega$ denotes the ML estimate of $\boldsymbol{\Theta}$ upon the perturbed model, which can be used to assess the influence of the perturbation on the ML estimate. Cook (1987) showed that the normal curvature for $\widehat{\boldsymbol{\Theta}}$ in the direction $\boldsymbol{d}$, with $||\boldsymbol{d}|| = 1$, is expressed as $C_l(\widehat{\boldsymbol{\Theta}}) = 2|\boldsymbol{d}^\top \boldsymbol{\nabla}^\top \ddot{\Sigma}(\widehat{\boldsymbol{\Theta}})^{-1} \boldsymbol{\nabla} \boldsymbol{d}|$, where $\boldsymbol{\nabla}$ is a $(3+q) \times n$ matrix of perturbations with elements $\nabla_{ji} = \partial^2 \ell(\Theta|\boldsymbol{\omega}) / \partial \Theta_j \partial \omega_i$ evaluated at $\boldsymbol{\Theta} = \widehat{\boldsymbol{\Theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, for $j = 1, \ldots, (3+q)$ and $i = 1, \ldots, n$. A local influence diagnostic is generally based on index plots. For example, the index graph of the eigenvector $\boldsymbol{d}_{\max}$ corresponding to the maximum eigenvalue of $\boldsymbol{B}(\boldsymbol{\Theta}) = -\boldsymbol{\nabla}^\top \boldsymbol{\Sigma}(\boldsymbol{\Theta})^{-1} \boldsymbol{\nabla}$, say $C_{d_{\max}}(\boldsymbol{\Theta})$, evaluated at $\boldsymbol{\Theta} = \widehat{\boldsymbol{\Theta}}$, can revel those observations that under small perturbations exercise a great influence on $\mathrm{LD}(\boldsymbol{\Theta})$. Another important direction of interest is $\boldsymbol{d}_i = \boldsymbol{e}_{in}$, which corresponds to the direction of the ith observation, where $\boldsymbol{e}_{in}$ is an $n \times 1$ vector of zeros with a value equal to one at the ith position, that is, $\{\boldsymbol{e}_{in}, 1 \leq i \leq n\}$ is the canonical basis of $\mathbb{R}^n$. In this case, the normal curvature is given by $C_i(\boldsymbol{\Theta}) = 2|b_{ii}|$, where b_{ii} is the ith diagonal element of $\boldsymbol{B}(\boldsymbol{\Theta})$ given earlier, for $i = 1, \ldots, n$, evaluated at $\boldsymbol{\Theta} = \widehat{\boldsymbol{\Theta}}$. Those cases when $C_i(\widehat{\boldsymbol{\Theta}}) > 2\overline{C}(\widehat{\boldsymbol{\Theta}})$, where $\overline{C}(\widehat{\boldsymbol{\Theta}}) = \sum_{i=1}^n C_i(\widehat{\boldsymbol{\Theta}})/n$, are considered as potentially influential. This procedure is called total local influence of the ith case; see Lesaffre and Verbeke (1998). Thus, the perturbation matrix is

$$\boldsymbol{\nabla} = (\boldsymbol{\nabla}_{vi})_{((3+q)\times n)} = \left(\frac{\partial^2 \ell(\boldsymbol{\Theta}|\boldsymbol{\omega})}{\partial \xi_v \partial \omega_i} \right)_{((3+q)\times n)}, v = 1, \ldots, 3+q, i = 1, \ldots, n.$$

We consider the model defined in (5.1)-(5.3) and its log-likelihood function given by (5.8). The elements of the matrix $\boldsymbol{\nabla}$ associated with each perturbation scheme were obtained numerically.

Case-weight perturbation

Under the case-weight perturbation scheme, we want to evaluate whether the contributions of the observations with different weights affect the ML estimate of $\boldsymbol{\Theta}$. The log-likelihood function of the perturbed model is

$$\ell(\boldsymbol{\Theta}|\boldsymbol{\omega}) = \begin{cases} \sum_{i=1}^n \omega_i \varsigma_i \log \left[\left(\dfrac{p_{0i}^{-\phi} - 1}{\phi} \right) h_T(t_i) S_T(t_i) \right] \\ -\sum_{i=1}^n \omega_i (\varsigma_i + 1/\phi) \log \left[1 - (p_{0i}^{-\phi} - 1)(1 - S_T(t_i)) \right], & \text{if } \phi \neq 0; \\ \sum_{i=1}^n \omega_i \varsigma_i \log \left[-\log(p_{0i}) h_T(t_i) S_T(t_i) \right] + \sum_{i=1}^n \omega_i (1 - S_T(t_i)) \log(p_{0i}), & \text{if } \phi = 0, \end{cases}$$

where $0 \leq \omega_i \leq 1$ and $\boldsymbol{\omega}_0 = (1,\ldots,1)^\top$. Then, considering its derivative with respect to $\boldsymbol{\omega}^\top$, we obtain $\boldsymbol{\nabla} = (\nabla_\Theta, \nabla_{\boldsymbol{b}_j})^\top$, where the elements of ∇_Θ and $\nabla_{\boldsymbol{b}_j}$ are, respectively, given by $\nabla_\Theta^{(i)} = d_\Theta^{(i)}$, and $\nabla_{\boldsymbol{b}_j}^{(i)} = d_{\boldsymbol{b}_j}^{(i)}$, for $i = 1,\ldots,n$, which must be evaluated numerically at $\boldsymbol{\Theta} = \widehat{\boldsymbol{\Theta}}$.

Response perturbation

We here assume an additive perturbation on the response i such that $t_i(\omega_i) = t_i + \omega_i s(t_i)$, where $s(t_i) = (1/\widehat{\phi})^{1/2}$ is a scale factor and $\omega_i \in \mathbb{R}$, for $i = 1,\ldots,n$. Then, the log-likelihood function is $\ell(\boldsymbol{\Theta}|\boldsymbol{\omega}) = \sum_{i=1}^n \ell_i(\boldsymbol{\Theta}|\omega_i)$, where

$$\ell(\boldsymbol{\Theta}|\boldsymbol{\omega}) = \begin{cases} \sum_{i=1}^n \varsigma_i \log\left[\left(\dfrac{p_{0i}^{-\phi}-1}{\phi}\right) h_T(t_i(\omega_i)) S_T(t_i(\omega_i))\right] \\ -\sum_{i=1}^n (\varsigma_i + 1/\phi)\log\left[1-(p_{0i}^{-\phi}-1)(1-S_T(t_i(\omega_i)))\right], & \text{if } \phi \neq 0; \\ \sum_{i=1}^n \varsigma_i \log\left[-\log(p_{0i}) h_T(t_i(\omega_i)) S_T(t_i(\omega_i))\right] & \text{if } \phi = 0 \\ +\sum_{i=1}^n (1-S_T(t_i(\omega_i)))\log(p_{0i}) & , \end{cases}$$

where $\boldsymbol{\omega}_0 = (0,\ldots,0)^\top$.

Covariate perturbation

We consider now an additive perturbation on a specific continuous covariate, z_t say, for $t = 1,\ldots,p$, by setting $z_{it}(\omega_i) = z_{it} + \omega_i s_z$, where s_z is a scale factor here assumed to be the standard deviation of z_t, and $\omega_i \in \mathbb{R}$, for $i = 1,\ldots,n$. Then, the log-likelihood function is $\ell(\boldsymbol{\Theta};\boldsymbol{\omega}) = \sum_{i=1}^n \ell_i(\boldsymbol{\Theta};\omega_i)$, where

$$\ell(\boldsymbol{\Theta};\boldsymbol{\omega}) = \begin{cases} \sum_{i=1}^n \varsigma_i \log\left[\left(\dfrac{p_{0i}^{*-\phi}-1}{\phi}\right) h_T(t_i) S_T(t_i)\right] \\ -\sum_{i=1}^n (\varsigma_i + 1/\phi)\log\left[1-(p_{0i}^{*-\phi}-1)(1-S_T(t_i))\right], & \text{if } \phi \neq 0; \\ \sum_{i=1}^n \varsigma_i \log\left[-\log(p_{0i}^*) h_T(t_i) S_T(t_i)\right] + \sum_{i=1}^n (1-S_T(t_i))\log(p_{0i}^*), & \text{if } \phi = 0, \end{cases}$$

where $p_{0i}^* = \dfrac{\exp(\boldsymbol{z}_i^{*\top}\boldsymbol{b})}{1+\exp(\boldsymbol{z}_i^{*\top}\boldsymbol{b})}$, $\boldsymbol{z}_i^{*\top}\boldsymbol{b} = b_1 z_{i1} + \cdots + b_t(z_{it} + \omega_i s_z) + \cdots + b_q z_{iq}$ and $\boldsymbol{\omega}_0 = (0,\ldots,0)^\top$.

5.4 Numerical evaluation

In this section, we present a Monte Carlo simulation study and an application to real medical data.

5.4.1 Simulation study

We consider the BSCrBeF model, with an exponential baseline, namely, $H_0(t) = \gamma t$ and $h_0(t) = \gamma$. The simulation scenario considers the following: sample sizes $n \in \{50, 100, 200, 400\}$, values of the parameters are $\gamma = 1.00$ and $\delta \in \{0.25, 0.50, 1.00\}$ and 5,000 MC replications. The cured fraction is $p_{0i} = \exp(b_0 + b_1 z_i)/(1+\exp(b_0 + b_1 z_i))$. We consider a binary covariate Z

with values generated from a Bernoulli distribution with parameter 0.5, $b_0 = 0.5$ and $b_1 = -1.0$, such that the cured fraction for the two levels of Z are $p_{00} = 0.62$ and $p_{01} = 0.38$, respectively. The censoring times were generated from the uniform distribution, $\mathrm{U}(a,b)$, where $a,b > 0$ are set to control the proportion of censored cases. In our study, the proportion of censored cases is on the average and approximately equal to 55.6%. Algorithm 4 is used to generate the observed times and censoring indicators.

Algorithm 4 – Generator of random numbers from the BSCrBeF model.

1: Define the baseline distribution and fix the parameters values.
2: Obtain a random number z_i from the covariate $Z \sim \text{Bernoulli}(1, 1/2)$.
3: Generate a random number v_i from $V \sim \mathrm{U}(p_0, 1)$.
4: Compute the lifetime y_i by solving numerically the equation

$$p(z_i) + (1 - p(z_i))S_T(y_i) = v_i.$$

5: Extract a random number c_i from the censoring time $C \sim \mathrm{U}(a,b)$, for $a,b > 0$ fixed.
6: Compute $t_i = \min\{y_i, c_i\}$.
7: If $y_i < c_i$, then $\varsigma_i = 1$, otherwise $\varsigma_i = 0$.
8: Repeat Steps 1 to 7 until a number m of data is completed.

For each value of the parameter and sample size, we report the empirical mean (EM) and empirical mean squared error (EMSE) of the ML estimators in Table 14. From this table, note that, as the sample size increases, the ML estimators become more efficient. In general, all of these results show the good performance of the ML estimators of the proposed model. The `R` software by its functions `optim` and `optimx` was used to estimate the parameters numerically; see <www.R-project.org>.

Table 14 – EM and EMSE of the ML estimators of δ and γ and cure fractions for simulated data from the BSCrBeF model.

		50		100		200		400	
δ	Parameters	EM	EMSE	EM	EMSE	EM	EMSE	EM	EMSE
0.25	p_{00}	0.5476	0.0066	0.5376	0.0077	0.5433	0.0068	0.6099	0.0008
	p_{01}	0.2767	0.0196	0.3083	0.0135	0.4007	0.0069	0.4098	0.0056
	γ	0.7730	0.1644	0.8373	0.1129	0.9131	0.0686	0.9707	0.0465
	δ	0.3026	0.0238	0.2683	0.0115	0.2452	0.0045	0.2494	0.0026
0.50	p_{00}	0.5430	0.0071	0.5359	0.0079	0.5432	0.0068	0.6080	0.0010
	p_{01}	0.2783	0.0189	0.3069	0.0138	0.3998	0.0071	0.4141	0.0057
	γ	0.8926	0.1050	0.9217	0.0707	0.9693	0.0506	0.9847	0.0395
	δ	0.4326	0.0255	0.4505	0.0180	0.4630	0.0131	0.4768	0.0092
1.00	p_{00}	0.5390	0.0075	0.5535	0.0067	0.5530	0.0058	0.6189	0.0008
	p_{01}	0.2792	0.0198	0.2973	0.0151	0.3860	0.0076	0.4044	0.0049
	γ	0.9764	0.0830	1.0643	0.0587	0.9815	0.0232	1.0555	0.0140
	δ	0.4930	0.2773	0.5511	0.2116	0.5946	0.1710	0.6284	0.1419

5.4.2 Illustrative example

Now, we illustrate the methodology proposed by applying it to a real-world medical data set, which corresponds to the lifetimes (ranging from 0.0274 to 15.25 years) until the patient's death or the censoring time. It contains 205 patients observed after operation for removal of malignant melanoma between 1962 and 1977. This data set is available in the `timereg` package in `R`. The percentage of censored observations was 72%. The explanatory variables recorded when the patient is taken on study include: tumor thickness, z_{1i}, (in mm, mean = 2.92 and standard deviation = 2.96); ulceration, z_{2i} (absent, $n = 115$; present, $n = 90$); and sex, z_{3i} (male, $n = 79$; female, $n = 126$).

Exploratory data analysis

Table 15 provides a descriptive summary of the observed times (in years) including MD, mean ($\bar{y}$), SD, CV, CS, CK, and minimum ($y_{(1)}$) and maximum ($y_{(n)}$) values. From this table, we observe the positively skewed nature and moderate kurtosis level of the data distribution. The skewed nature is confirmed by the histogram of Figure 17(left). Figure 17(center) suggests an increasing HR for the observed times. Therefore, the Weibull distribution is a good candidate as a baseline HR, since it allows us to model constant, increasing and decreasing HRs. Also, the Weibull distribution is one of the most used models in survival and reliability analysis due to its good properties and flexibility in data modeling.

Table 15 – Descriptive statistics for the observed lifetime.

MD	$\bar{t}$	SD	CV	CS	CK	$t_{(1)}$	$t_{(n)}$	n
5.49	5.90	3.07	52.12	0.33	−0.28	0.03	15.25	205

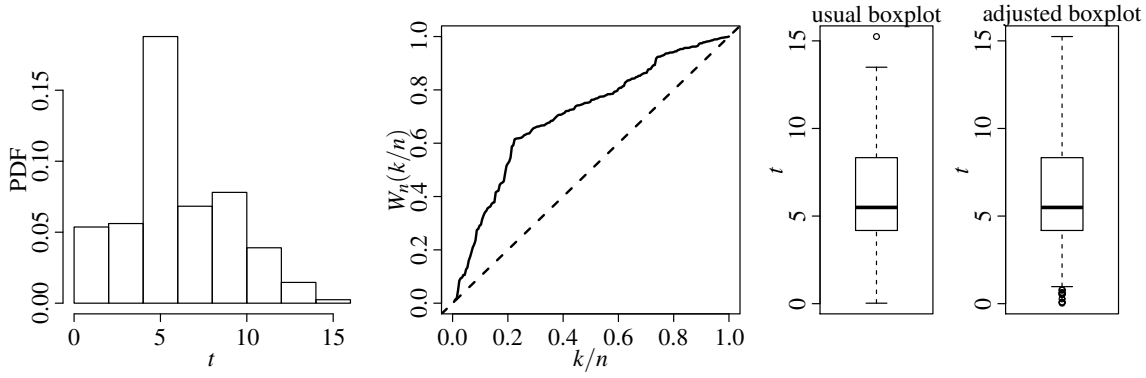

Figure 17 – Histogram, TTT plot and boxplots for the melanoma data.

Estimation and verification of assumptions

From (5.5), we assume the following regression structure

$$\boldsymbol{z}_i^\top \boldsymbol{b} = b_0 + b_1 z_{1i} + b_2 z_{2i} + b_3 z_{3i}, \quad i = 1, 2, \ldots, n.$$

In Table 14, results are given for the BSCrNBF model with Weibull baseline analyzing melanoma data. For comparison, the results of the BSCrPoF and BSCrBeF models with the same baseline, as well as the BS negative binomial, BS Poisson and BS Bernoulli models without frailty are compared as well. Note that the BSCrNBF model nests all the remaining models. From Table 16, observe that the BSCrNBF model has the smallest values for the Akaike (AIC) and Bayesian (BIC) information criteria, suggesting that this model provides the best fit to these data. Moreover, the statistic $-2\max \ell(\widehat{\boldsymbol{\Theta}})$ supports this result. The ML estimates (with estimated asymptotic SEs in parentheses) of the parameters for the BSCrNBF model are reported in Table 17.

Table 16 – Statistics from the fitted models.

Model	Statistics		
	$-2\max \ell(\boldsymbol{\Theta})$	AIC	BIC
BS Negative binomial frailty (BSCrNBF)	401.675	415.675	438.936
BS Poisson frailty (BSCrPoF)	412.723	426.723	449.984
BS Bernoulli frailty (BSCrBeF)	417.104	426.723	454.365
BS Negative binomial	413.020	425.020	444.950
BS Poisson	415.070	427.070	447.010
BS Bernoulli	417.020	429.020	448.960

Table 17 – ML estimates of the parameters for the BSCrNBF model.

Parameter	Estimates	SE	p-value
ϕ	3.3493	1.4692	–
κ	2.6783	0.4686	–
γ	7.0570	2.0131	–
δ	1.3821	2.6103	–
b_0	1.6897	0.5502	0.0012
b_1	−0.1270	0.0416	0.0022
b_2	−1.2909	0.3338	< 0.001
b_3	−0.5365	0.2875	0.0588

We now carry out a diagnostic analysis based on local influence. We present the results for each perturbation scheme.

Case-weight perturbation

Index plots of C_i under the case-weight perturbation are displayed in Figure 18. This figure detects observations #5, #6, #7 and #10 as potential influential cases.

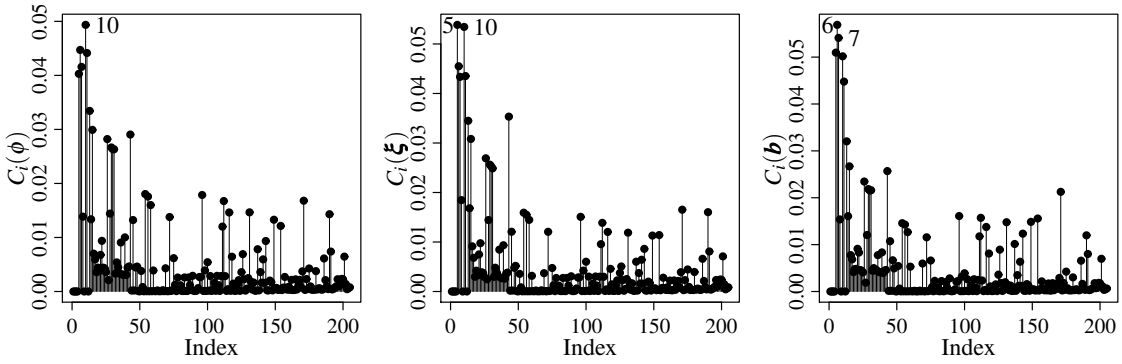

Figure 18 – Index plots of C_i for α, $\boldsymbol{\xi}$ and $\boldsymbol{b}$ under the case-weight perturbation scheme.

Response perturbation

Figure 19 displays index plots of C_i under the response variable perturbation. We note that the observations #5, #6 and #7 are detected as potential influential cases.

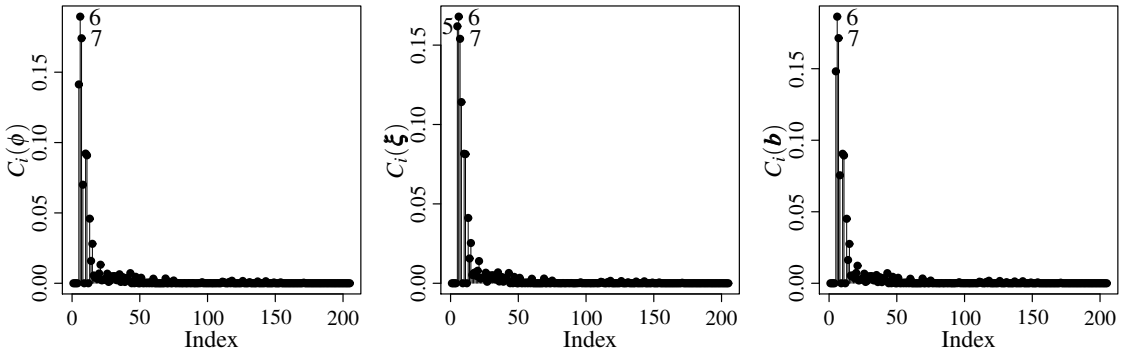

Figure 19 – Index plots of C_i for ϕ, $\boldsymbol{\xi}$ and $\boldsymbol{b}$ under the response perturbation scheme.

Covariate perturbation

The perturbation of the explanatory variable tumor thickness (x_{1i}) is assessed here. Figure 20 shows index plots of C_i after a perturbation of this explanatory variable. From this figure, we observe that the case # 8 is detected as a common potential influential case.

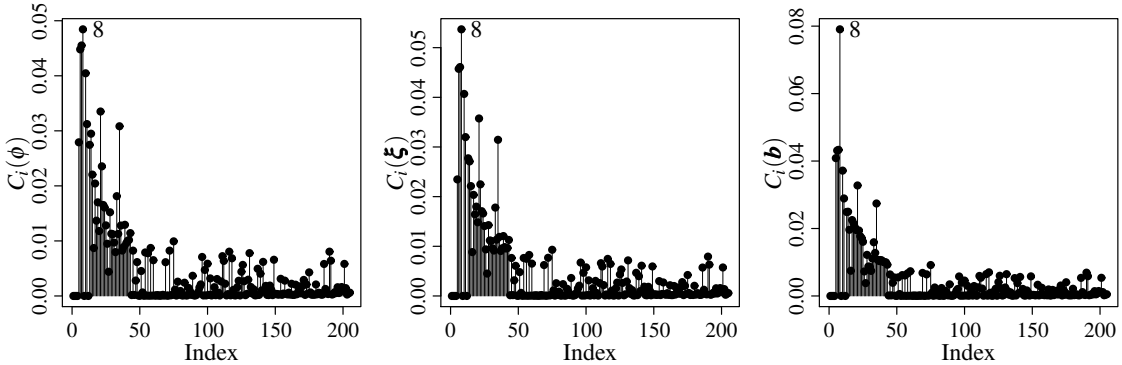

Figure 20 – Index plots of C_i for ϕ, $\boldsymbol{\xi}$ and $\boldsymbol{b}$ under the regressor perturbation scheme.

Similar to the Section 4.4 we check the impact of the detected influential observations on the model inference where we compute the RC. Table 18 shows the RCs in the parameter estimates and their corresponding estimated SEs. Also, p-values are shown for the regression coefficients. From this table, we note that the largest RCs are associated with observations {#6} and {#7}. Note also that no inferential changes (at the level of 1%) are found; namely, the diagnostic measures identify potentially influential observations, but these do not alter the inference of the model.

Table 18 – RCs (in %) in ML estimates and their corresponding SEs for the indicated parameter and dropped cases, and respective p-values.

Dropped case		$\widehat{\phi}$	$\widehat{\kappa}$	$\widehat{\gamma}$	$\widehat{\delta}$	$\widehat{b}_0$	$\widehat{b}_1$	$\widehat{b}_2$	$\widehat{b}_3$
{5}	$\mathrm{RC}_{\xi_{j(i)}}$	12.8	3.16	1.65	6.1	2.74	8.23	1.78	0.68
	$\mathrm{RC}_{\mathrm{SE}(\xi_{j(i)})}$	(1.75)	(2.17)	(9.34)	(0.53)	(8.24)	(4.91)	(2.82)	(0.61)
	p-value	-	-	-		[0.005]	[0.007]	[<0.001]	[0.065]
{6}	$\mathrm{RC}_{\xi_{j(i)}}$	22.1	3.97	2.77	14.94	1.34	5.92	3.03	4.41
	$\mathrm{RC}_{\mathrm{SE}(\xi_{j(i)})}$	(13.58)	(1.11)	(0.52)	(22.47)	(24.63)	(4.39)	(3.66)	(2.23)
	p-value	-	-	-		[0.015]	[0.001]	[0.001]	[0.056]
{7}	$\mathrm{RC}_{\xi_{j(i)}}$	25.12	4.98	2.84	16.23	0.96	7.89	2.97	4.46
	$\mathrm{RC}_{\mathrm{SE}(\xi_{j(i)})}$	(18.08)	(2.59)	(6.62)	(4.71)	(21.37)	(6.78)	(3.34)	(2.58)
	p-value	-	-	-		[0.012]	[0.002]	[<0.001]	[0.057]
{8}	$\mathrm{RC}_{\xi_{j(i)}}$	16.73	5.62	1.2	5.51	2.77	6.94	2.76	3.56
	$\mathrm{RC}_{\mathrm{SE}(\xi_{j(i)})}$	(0.48)	(4.79)	(5.79)	(2.06)	(6.58)	(4.91)	(2.70)	(0.81)
	p-value	-	-	-		[0.005]	[0.006]	[<0.001]	[0.055]
{10}	$\mathrm{RC}_{\xi_{j(i)}}$	12.44	1.33	1.55	8.00	2.02	0.78	0.79	22.44
	$\mathrm{RC}_{\mathrm{SE}(\xi_{j(i)})}$	(14.47)	(1.49)	(0.14)	(6.71)	(5.48)	(1.54)	(0.38)	(0.69)
	p-value	-	-	-		[0.002]	[0.002]	[<0.001]	[0.021]
{6,7}	$\mathrm{RC}_{\xi_{j(i)}}$	22.28	0.38	3.78	18.31	0.36	13.33	0.53	7.17
	$\mathrm{RC}_{\mathrm{SE}(\xi_{j(i)})}$	(17.89)	(7.07)	(21.43)	(4.22)	(30.03)	(11.10)	(4.91)	(3.64)
	p-value	-	-	-		[0.018]	[0.001]	[<0.001]	[0.094]
{5,10}	$\mathrm{RC}_{\xi_{j(i)}}$	24.99	3.58	3.27	15.51	0.81	5.34	1.93	18.76
	$\mathrm{RC}_{\mathrm{SE}(\xi_{j(i)})}$	(13.96)	(0.03)	(18.64)	(5.15)	(22.95)	(11.21)	(4.43)	(2.18)
	p-value	-	-	-		[0.013]	[0.009]	[<0.001]	[0.030]

5.5 Concluding remarks

In this chapter, we introduced a cure rate frailty model based on a Birnbaum-Saunders distribution. In this new approach, a frailty parameter is included in the model to bring additional information that may be useful in practice. The proposed methodology encompassed estimation and inference about the model parameters and influence diagnostics. We carried out a Monte Carlo simulation study which showed that the estimates based on the maximum likelihood estimators of the model parameters tend to their true values, when the sample size increased, as expected. We applied the proposed methodology to real-world data concerning to the survival times of patients who died from acute myelogenous leukemia. The results showed the potentiality of this methodology.

CHAPTER

6

ON A BIVARIATE BIRNBAUM-SAUNDERS DISTRIBUTION PARAMETERIZED BY ITS MEANS

6.1 Introduction

In this chapter we introduce a bivarite reparameterized BS distribution by its mean. We use the ML and MM methods for estimating the model parameters. We introduce a regression model and its inference by ML methods, conduct a simulation study to evaluate the behavior of the parameter estimators and illustrate the proposed model with real data sets.

6.2 Reparameterized bivariate Birnbaum-Saunders distribution

Chapter 2 presented a review of the univariate and bivariate BS distributions in terms of the orginal parameterization as well as the univariate case of the BS reparameterized by its mean. Here, from (2.14) we obtain the reparameterized bivariate mean-based Birnbaum-Saunders (BBSM) distribution. Let $\boldsymbol{T} = (T_1, T_2)^\top$ be a bivariate random vector following a bivariate Birnbaum-Saunders distribution parameterized by its mean with parameters μ_1, δ_1, μ_2, δ_2, ρ. Then, the joint CDF of T_1 and T_2 can be expressed as

$$P(T_1 \leq t_1, T_2 \leq t_2) \quad = \Phi_2\left(\sqrt{\tfrac{\delta_1}{2}}\,(a_1 - b_1), \sqrt{\tfrac{\delta_2}{2}}\,(a_2 - b_2); \rho\right), \tag{6.1}$$

where $a_1=\sqrt{\frac{(\delta_1+1)t_1}{\delta_1\mu_1}}$, $b_1=\sqrt{\frac{\delta_1\mu_1}{(\delta_1+1)t_1}}$, $a_2=\sqrt{\frac{(\delta_2+1)t_2}{\delta_2\mu_2}}$, $b_2=\sqrt{\frac{\delta_2\mu_2}{(\delta_2+1)t_2}}$ and $t_1>0, t_2>0$, $\mu_1>0$, $\delta_1>0$, $\mu_2>0$, $\delta_2>0$, $-1<\rho<1$, and $\Phi_2(u,v;\rho)$ is the standard bivariate normal CDF with correlation coefficient ρ. Then, the joint PDF associated with (6.1) is given by

$$f_{T_1,T_2}(t_1,t_2) = \phi_2\left(\sqrt{\tfrac{\delta_1}{2}}(a_1-b_1),\sqrt{\tfrac{\delta_2}{2}}(a_2-b_2);\rho\right)\frac{\sqrt{\delta_1}}{2\sqrt{2}t_1}(a_1+b_1)\frac{\sqrt{\delta_2}}{2\sqrt{2}t_2}(a_2+b_2), \quad (6.2)$$

where $\phi_2(\cdot,\cdot;\rho)$ is a normal joint PDF given by $\phi_2(u,v;\rho)=\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left(\frac{1}{(1-\rho^2)}(u^2+v^2-2\rho uv)\right)$. From now on a bivariate distribution with PDF (6.2) will be denoted by BBSM($\boldsymbol{\psi}$), where $\boldsymbol{\psi}=(\mu_1,\delta_1,\mu_2,\delta_2,\rho)^\top$ is the vector parameter.

Theorem 1 If $\boldsymbol{T}=(T_1,T_2)^\top\sim$ BBSM($\boldsymbol{\psi}$), then

a) $\boldsymbol{T}^{-1}=(T_1^{-1},T_2^{-1})^\top\sim$ BBSM$\left(\frac{(\delta_1+1)^2}{\mu_1\delta_1^2},\delta_1,\frac{(\delta_2+1)^2}{\mu_2\delta_2^2},\delta_2,\rho\right)$;

b) $\boldsymbol{T}_1^{-1}=(T_1^{-1},T_2)^\top\sim$ BBSM$\left(\frac{(\delta_1+1)^2}{\mu_1\delta_1^2},\delta_1,\mu_2,\delta_2,-\rho\right)$;

c) $\boldsymbol{T}_2^{-1}=(T_1,T_2^{-1})^\top\sim$ BBSM$\left(\mu_1,\delta_1,\frac{(\delta_2+1)^2}{\mu_2\delta_2^2},\delta_2,-\rho\right)$.

Proof. By using the PDF in (6.2) and making suitable transformations.

6.2.1 Maximum likelihood estimation

Let $\{(t_{1i},t_{2i}),i=1,\ldots,n\}$ be a bivariate random sample of size n from the BBSM($\boldsymbol{\psi}$) distribution with PDF as given in (6.2). Then, the log-likelihood function, without the additive constant, is given by

$$\begin{aligned}\ell(\boldsymbol{\psi}) &= -\frac{n}{2}\log(1-\rho^2)-\frac{1}{4(1-\rho^2)}\sum_{i=1}^{n}\Big\{-2\sqrt{\delta_2}\sqrt{\delta_1}\rho\,(a_{1i}-b_{1i})(a_{2i}-b_{2i})\\ &\quad+\frac{\delta_1^2\mu_1}{(\delta_1+1)t_{1i}}-2\delta_1a_{1i}b_{1i}+\frac{\delta_2^2\mu_2}{(\delta_2+1)t_{2i}}-2\delta_2a_{2i}b_{2i}+\frac{(\delta_1+1)t_{1i}}{\mu_1}+\frac{(\delta_2+1)t_{2i}}{\mu_2}\Big\}\\ &\quad+\frac{n}{2}\log(\delta_1)+\sum_{i=1}^{n}\log(a_{1i}+b_{1i})+\frac{n}{2}\log(\delta_2)+\sum_{i=1}^{n}\log(a_{2i}+b_{2i}), \quad (6.3)\end{aligned}$$

where $a_{1i}=\sqrt{\frac{(\delta_1+1)t_{1i}}{\delta_1\mu_1}}$, $b_{1i}=\sqrt{\frac{\delta_1\mu_1}{(\delta_1+1)t_{1i}}}$, $a_{2i}=\sqrt{\frac{(\delta_2+1)t_{2i}}{\delta_2\mu_2}}$ and $b_{2i}=\sqrt{\frac{\delta_2\mu_2}{(\delta_2+1)t_{2i}}}$. Note that $\left\{\sqrt{\frac{\delta_1}{2}}[a_1-b_1],\sqrt{\frac{\delta_2}{2}}[a_2-b_2]\right\}$ is bivariate normal distributed with mean vector $(0,0)^\top$ and covariance matrix $\begin{pmatrix}1&\rho\\ \rho&1\end{pmatrix}$. From this result, we see that, for given μ_1, δ_1, μ_2, δ_2, the ML estimate of ρ is

$$\widehat{\rho}(\mu_1,\delta_1,\mu_2,\delta_2)=\frac{\sum_{i=1}^{n}\sqrt{\frac{\delta_1}{2}}[a_{1i}-b_{1i}]\sqrt{\frac{\delta_2}{2}}[a_{2i}-b_{2i}]}{\sqrt{\sum_{i=1}^{n}\left\{\sqrt{\frac{\delta_1}{2}}[a_{1i}-b_{1i}]\right\}^2}\sqrt{\sum_{i=1}^{n}\left\{\sqrt{\frac{\delta_2}{2}}[a_{1i2}-b_{2i}]\right\}^2}}. \quad (6.4)$$

Therefore, when the parameters μ_1, δ_1, μ_2 and δ_2 are unknown, the ML estimates of μ_1, δ_1, μ_2 and δ_2 can be obtained by maximizing the profile log-likelihood function

$$\begin{aligned}\ell_p(\boldsymbol{\eta}) &= -\frac{n}{2}\log(1-\widehat{\rho}(\boldsymbol{\eta})^2) - \frac{1}{4(1-\widehat{\rho}(\boldsymbol{\eta})^2)}\sum_{i=1}^{n}\Big\{-2\sqrt{\delta_2}\sqrt{\delta_1}\widehat{\rho}(\boldsymbol{\eta}) \\ &\quad \times (a_{1i}-b_{1i})(a_{2i}-b_{2i}) + \frac{\delta_1^2\mu_1}{(\delta_1+1)t_{1i}} - 2\delta_1 a_{1i}b_{1i} + \frac{\delta_2^2\mu_2}{(\delta_2+1)t_{2i}} \\ &\quad -2\delta_2 a_{2i}b_{2i} + \frac{(\delta_1+1)t_{1i}}{\mu_1} + \frac{(\delta_2+1)t_{2i}}{\mu_2}\Big\} + \frac{n}{2}\log(\delta_1) \\ &\quad + \sum_{i=1}^{n}\log(a_{1i}+b_{1i}) + \frac{n}{2}\log(\delta_2) + \sum_{i=1}^{n}\log(a_{2i}+b_{2i}),\end{aligned} \tag{6.5}$$

where $\boldsymbol{\eta} = (\mu_1, \delta_1, \mu_2, \delta_2)^\top$. In order to maximize function (6.5) with respect to μ_1, δ_1, μ_2, δ_2, one may use the Newton-Raphson algorithm or some other optimization algorithm. Once $\widehat{\mu}_1$, $\widehat{\delta}_1$, $\widehat{\mu}_2$ and $\widehat{\delta}_2$ are obtained, the ML estimates of ρ, say $\widehat{\rho}$, is computed from (6.4). Under some regularity conditions Cox and Hinkley (1974), the asymptotic distribution of $\widehat{\boldsymbol{\psi}} = (\widehat{\mu}_1, \widehat{\delta}_1, \widehat{\mu}_2, \widehat{\delta}_2, \widehat{\rho})$, as $n \to \infty$, is given by

$$\sqrt{n}(\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi}) \xrightarrow{D} \mathrm{N}_5\left(\mathbf{0}, \boldsymbol{J}^{-1}\right), \tag{6.6}$$

where $\xrightarrow{D}$ denotes convergence in distribution and $\mathrm{N}_5\left(\mathbf{0}, \boldsymbol{J}^{-1}\right)$ denotes a 5-variate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{J}^{-1}$. For the sake of space we omit the elements of the matrix $\boldsymbol{J}$.

6.2.2 Modified moment estimation

Let $\{(t_{1i}, t_{2i}), i = 1, \ldots, n\}$ be a bivariate random sample of size n from $T \sim \mathrm{BBSM}(\boldsymbol{\psi})$. Also, let the sample arithmetic and harmonic means be defined as

$$s_k = \frac{1}{n}\sum_{i=1}^{n} t_{ki} \quad \text{and} \quad r_k = \left[\frac{1}{n}\sum_{i=1}^{n} t_{ki}^{-1}\right]^{-1}, \quad k = 1, 2,$$

respectively. Then, the modified moment estimators of μ_1, δ_1, μ_2 and δ_2 are obtained by equating $\mathrm{E}[T_1]$, $\mathrm{E}[T_1^{-1}]$, $\mathrm{E}[T_2]$ and $\mathrm{E}[T_2^{-1}]$ to the corresponding sample estimates, that is,

$$\mathrm{E}[T_1] = s_1, \quad \mathrm{E}[T_1^{-1}] = r_1^{-1}, \quad \mathrm{E}[T_2] = s_2 \quad \text{and} \quad \mathrm{E}[T_2^{-1}] = r_2^{-1}. \tag{6.7}$$

Thus, we have

$$s_1 = \widetilde{\mu}_1, \quad r_1^{-1} = \frac{[\widetilde{\delta}_1 + 1]^2}{\widetilde{\mu}_1\widetilde{\delta}_1^2}, \quad s_2 = \widetilde{\mu}_2 \quad \text{and} \quad r_2^{-1} = \frac{[\widetilde{\delta}_2 + 1]^2}{\widetilde{\mu}_2\widetilde{\delta}_2^2}. \tag{6.8}$$

Solving (6.8) for μ_1, δ_1, μ_2 and δ_2, we obtain the modified moment estimators of these parameters, denoted by $\widetilde{\mu}_1$, $\widetilde{\delta}_1$, $\widetilde{\mu}_2$ and $\widetilde{\delta}_2$, namely,

$$\widetilde{\mu}_1 = s_1, \quad \widetilde{\delta}_1 = \left[\sqrt{\frac{s_1}{r_1}} - 1\right]^{-1}, \quad \widetilde{\mu}_2 = s_2, \quad \text{and} \quad \widetilde{\delta}_2 = \left[\sqrt{\frac{s_2}{r_2}} - 1\right]^{-1}. \tag{6.9}$$

Theorem 2 The asymptotic distributions of $\widetilde{\mu}_k$ and $\widetilde{\delta}_k$, for $k = 1,2$, are given by

$$\sqrt{n}(\widetilde{\mu}_k - \mu_k) \sim \mathrm{N}\left(0, \frac{\mu_k^2\{2\delta_k+5\}}{\{\delta_k+1\}^2}\right), \quad \sqrt{n}(\widetilde{\delta}_k - \delta_k) \sim \mathrm{N}\left(0, 2\delta_k^2\right). \tag{6.10}$$

Proof.

Let $\boldsymbol{T} = (T_1, T_2)^\top$ follow a BBSM($\boldsymbol{\psi}$) distribution, then

$$\mathrm{Var}[T_k] = \frac{\mu_k^2[2\delta_k+5]}{[\delta_k+1]^2}, \quad \mathrm{Var}[T_k^{-1}] = \frac{[2\delta_k+5][\delta_k+1]^2}{\mu_k^2\delta_k^4} \quad \text{and} \quad \mathrm{Cov}[T_k] = 1 - \frac{[\delta_k+1]^2}{\delta_k^2}, \quad k = 1,2.$$

Now, let $\{(t_{1i}, t_{2i}), i = 1,\ldots,n\}$ be a bivariate random sample from the BBSM($\boldsymbol{\psi}$) distribution. Then the MM estimates are given by

$$\widetilde{\mu}_k = s_k \quad \text{and} \quad \widetilde{\delta}_k = \left[\sqrt{\frac{s_k}{r_k}} - 1\right]^{-1}, \quad k = 1,2.$$

Consider $S_k = \frac{1}{n}\sum_{i=1}^n T_{kj}$ and $R_k^* = R_k^{-1} = \sum_{i=1}^n \frac{1}{T_{ki}}$, with $k = 1,2$, which implies that the vector $(S_k, R_k^{-1})^\top$ is bivariate normal distributed, that is,

$$\sqrt{n}\begin{pmatrix} S_k - \mathrm{E}[T_k] \\ R_k^* - \mathrm{E}[T_k^{-1}] \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathrm{Var}[T_k], 1 - \mathrm{E}[T_k]\mathrm{E}[T_k^{-1}] \\ 1 - \mathrm{E}[T_k]\mathrm{E}[T_k^{-1}], \mathrm{Var}[T_k] \end{pmatrix}\right].$$

However, we need to find the asymptotic joint distribution of $(\widetilde{\mu}_k, \widetilde{\delta}_k)^\top$. Consider $\widetilde{\mu}_k = f_k(S_k, R_k^*)$ and $\widetilde{\delta}_k = f_k(S_k, R_k^*)$, such that $f_k(x,y) = x$ and $f_2(x,y) = [\sqrt{xy} - 1]^{-1}$. By using the delta method, we readily have

$$\sqrt{n}\begin{pmatrix} \widetilde{\mu}_k - \mu_k \\ \widetilde{\delta}_k - \delta_k \end{pmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_k\right), \quad k = 1,2,$$

where

$$\boldsymbol{\Sigma}_k = \begin{pmatrix} \frac{\mu_k^2[2\delta_k+5]}{[\delta_k+1]^2} & -\frac{2\mu_k\delta_k}{\delta_k+1} \\ -\frac{2\mu_k\delta_k}{\delta_k+1} & 2\delta_k^2 \end{pmatrix}.$$

6.3 BBSM regression model

Let $\boldsymbol{T} = (T_1, T_2)^\top$ follow a BBSM($\boldsymbol{\psi}$) distribution. Assume that there are p and q covariates, say $\boldsymbol{X}^{(1)} = (x_1^{(1)}, x_2^{(1)}, \ldots, x_p^{(1)})^\top$ and $\boldsymbol{X}^{(2)} = (x_1^{(2)}, x_2^{(2)}, \ldots, x_q^{(2)})^\top$, associated with the random variables T_1 and T_2, respectively. Then, from (6.1) we have

$$g(\mu_k) = \boldsymbol{\varphi}_k^\top \boldsymbol{X}^{(k)} = \varphi_{k0} + \varphi_{k1}x_1^{(k)} + \varphi_{k2}x_2^{(k)} + \cdots + \varphi_{kl}x_l^{(k)}, \tag{6.11}$$

where $\boldsymbol{\varphi}_k = (\varphi_{j1}, \varphi_{j2}, \ldots, \varphi_{jl})$ is a vector of l unknown parameters, for $k = 1,2$ and $l = p,q$, and g is a link function such that $\mu_k = g^{-1}(\boldsymbol{\varphi}_k^\top \boldsymbol{X}^{(k)})$, with g^{-1} being the inverse function of $g(\cdot)$. In this work, $g(\mu_k) = \log(\mu_k)$. Then,

$$\mu_k = \exp(\boldsymbol{\varphi}_k^\top \boldsymbol{X}^{(k)}) = \exp(\varphi_{k0} + \varphi_{k1}x_1^{(k)} + \varphi_{k2}x_2^{(k)} + \cdots + \varphi_{kl}x_l^{(k)}). \tag{6.12}$$

Note that the precision parameter δ_k is independent of the covariates $\boldsymbol{X}^{(k)}$. In addition, since $\text{Var}[T_k] = \mu_k/\phi_k$, where $\phi_k = [\delta_k+1]^2/[2\delta_k+5]$, is a function of μ_k and, consequently, of the covariates, we can analyze situations where a non-constant variance is present by using the model in (6.12).

6.3.1 Maximum likelihood estimation

Let $\{(t_{1i}, t_{2i}), i = 1,\ldots,n\}$ be a bivariate random sample of size n with the covariates corresponding to t_{1i} as $\boldsymbol{X}_i^{(1)} = (x_{1i}^{(1)}, x_{2i}^{(1)}, \ldots, x_{pi}^{(1)})$ and t_{2i} as $\boldsymbol{X}_i^{(2)} = (x_{1i}^{(2)}, x_{2i}^{(2)}, \ldots, x_{qi}^{(2)})$. The problem of interest is to estimate the unknown parameters δ_1, δ_2, $\boldsymbol{\varphi}_1$, $\boldsymbol{\varphi}_2$ and ρ. The log-likelihood function, without the additive constant, can be written as follows:

$$\begin{aligned}
\ell(\boldsymbol{\vartheta}) &= -\frac{n}{2}\log(1-\rho^2) - \frac{1}{4(1-\rho^2)}\sum_{i=1}^{n}\left\{-2\sqrt{\delta_2}\sqrt{\delta_1}\rho\,(a_{1i}-b_{1i})(a_{2i}-b_{2i})\right. \\
&\quad \left. + \frac{\delta_1^2\mu_{1i}}{(\delta_1+1)t_{1i}} - 2\delta_1 a_{1i}b_{1i} + \frac{\delta_2^2\mu_{2i}}{(\delta_2+1)t_{2i}} - 2\delta_2 a_{2i}b_{2i} + \frac{(\delta_1+1)t_{1i}}{\mu_{1i}} + \frac{(\delta_2+1)t_{2i}}{\mu_{2i}}\right\} \\
&\quad + \frac{n}{2}\log(\delta_1) + \sum_{i=1}^{n}\log(a_{1i}+b_{1i}) + \frac{n}{2}\log(\delta_2) + \sum_{i=1}^{n}\log(a_{2i}+b_{2i}), \qquad (6.13)
\end{aligned}$$

where $\boldsymbol{\vartheta} = (\delta_1, \delta_2, \boldsymbol{\varphi}_1^\top, \boldsymbol{\varphi}_2^\top, \rho)^\top$, $a_{1i} = \sqrt{\frac{(\delta_1+1)t_{1i}}{\delta_1\mu_{1i}}}$, $b_{1i} = \sqrt{\frac{\delta_1\mu_{1i}}{(\delta_1+1)t_{1i}}}$, $a_{2i} = \sqrt{\frac{(\delta_2+1)t_{2i}}{\delta_2\mu_{2i}}}$ and $b_{2i} = \sqrt{\frac{\delta_2\mu_{2i}}{(\delta_2+1)t_{2i}}}$, with $\mu_{ji} = \exp(\varphi_{j0} + \varphi_{j1}x_{1i}^{(k)} + \varphi_{j2}x_{2i}^{(k)} + \cdots + \varphi_{jl}x_{li}^{(k)})$, for $k = 1,2$, $i = 1,2,\ldots,n$ and $l = p,q$.

Similarly to the no covariate case, it can be seen that for fixed δ_1, δ_2, $\boldsymbol{\varphi}_1$ and $\boldsymbol{\varphi}_2$, the ML estimates of ρ can be obtained as

$$\widehat{\rho}(\delta_1, \boldsymbol{\varphi}_1, \delta_2, \boldsymbol{\varphi}_2) = \frac{\sum_{i=1}^{n}\sqrt{\frac{\delta_1}{2}}(a_{1i}-b_{1i})\sqrt{\frac{\delta_2}{2}}[a_{2i}-b_{2i}]}{\sqrt{\sum_{i=1}^{n}\left\{\sqrt{\frac{\delta_1}{2}}(a_{1i}-b_{1i})\right\}^2}\sqrt{\sum_{i=1}^{n}\left\{\sqrt{\frac{\delta_2}{2}}(a_{1i2}-b_{2i})\right\}^2}}. \qquad (6.14)$$

Therefore, the ML estimates of μ_1, $\boldsymbol{\varphi}_1$, δ_2 and $\boldsymbol{\varphi}_2$ can be obtained by maximizing the profile log-likelihood function

$$\begin{aligned}
\ell_p(\boldsymbol{\zeta}) &= -\frac{n}{2}\log(1-\widehat{\rho}(\boldsymbol{\zeta})^2) - \frac{1}{4(1-\widehat{\rho}(\boldsymbol{\zeta})^2)}\sum_{i=1}^{n}\left\{-2\sqrt{\delta_2}\sqrt{\delta_1}\widehat{\rho}(\boldsymbol{\zeta})\right. \\
&\quad \left. \times (a_{1i}-b_{1i})(a_{2i}-b_{2i}) + \frac{\delta_1^2\mu_1}{(\delta_1+1)t_{1i}} - 2\delta_1 a_{1i}b_{1i} + \frac{\delta_2^2\mu_2}{(\delta_2+1)t_{2i}} - 2\delta_2 a_{2i}b_{2i}\right. \\
&\quad \left. + \frac{(\delta_1+1)t_{1i}}{\mu_1} + \frac{(\delta_2+1)t_{2i}}{\mu_2}\right\} + \frac{n}{2}\log(\delta_1) + \sum_{i=1}^{n}\log(a_{1i}+b_{1i}) \\
&\quad + \frac{n}{2}\log(\delta_2) + \sum_{i=1}^{n}\log(a_{2i}+b_{2i}), \qquad (6.15)
\end{aligned}$$

where $\boldsymbol{\zeta} = (\delta_1, \boldsymbol{\varphi}_1, \delta_2, \boldsymbol{\varphi}_2)^\top$. The estimation of the parameters can be performed by using a numerical optimization method. The asymptotic distribution of $\widehat{\boldsymbol{\zeta}} = (\widehat{\delta}_1, \widehat{\delta}_2, \widehat{\boldsymbol{\varphi}}_1, \widehat{\boldsymbol{\varphi}}_2, \widehat{\rho})$, as $n \to \infty$, under some regularity conditions Cox and Hinkley (1974), is given by

$$\sqrt{n}(\widehat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}) \xrightarrow{D} \mathrm{N}_{p+q+5}\left(\mathbf{0}, \boldsymbol{R}^{-1}_{p+q+5}\right), \tag{6.16}$$

where $\mathrm{N}_{p+q+5}\left(\mathbf{0}, \boldsymbol{R}^{-1}_{p+q+5}\right)$ denotes a $(p+q+5)$-variate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{R}^{-1}_{p+q+5}$.

Initial values for the regression coefficients can be obtained by the least squares estimator (LSE) of $\boldsymbol{\varphi}_k$, $k = 1, 2$, that is, by minimizing the sum of squares

$$S(\boldsymbol{\varphi}_k) = \sum_{i=1}^{n}\left(\log(t_{ki}) - \varphi_{j0} - \varphi_{j1}x_{1i}^{(k)} - \varphi_{j2}x_{2i}^{(k)} + \cdots + \varphi_{jl}x_{li}^{(k)}\right), \tag{6.17}$$

for $k = 1, 2$, $i = 1, 2, \ldots, n$ and $l = p, q$. Then, the initial value for the ML estimates of $\boldsymbol{\varphi}_k$, based on the LSE, say $\boldsymbol{\varphi}_k^{(0)}$, is given by

$$\boldsymbol{\varphi}_k^{(0)} = \left(\boldsymbol{X}^{(k)\top}\boldsymbol{X}^{(k)}\right)^{-1}\boldsymbol{X}^{(k)\top}\log(\boldsymbol{t}_k), \tag{6.18}$$

where

$$\boldsymbol{X}^{(k)} = \begin{pmatrix} 1 & x_{11}^{(k)} & x_{21}^{(k)} & \ldots & x_{l1}^{(k)} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & x_{1n}^{(k)} & x_{2n}^{(k)} & \ldots & x_{ln}^{(k)} \end{pmatrix} \quad \text{and} \quad \log(\boldsymbol{t}_k) = \begin{pmatrix} \log(t_{k1}) \\ \vdots \\ \log(t_{kn}) \end{pmatrix}$$

.

6.4 Numerical applications

In this section, we carry out a simulation study to evaluate the performance of the ML estimators of the BBSM model parameters. Then, we illustrate the BBSM distribution and its corresponding regression model by using two real data sets, respectively. The first data set corresponds to two different measurements of stiffness, whereas the second data set represents bone mineral density data with 5 covariates associated with each dependent variable.

6.4.1 A simulation study

The simulation scenario considered the following: the sample sizes $n \in \{10, 50, 100\}$; the values of the shape and scale parameters as $\delta_k \in \{0.25, 2.0\}$ and $\mu_k = 2.0$, for $k = 1, 2$, respectively; the values of ρ are $0.00, 0.25, 0.50$ and 0.95 (the results for negative ρ are quite similar so are omitted here); and $5,000$ MC replications. The values of δ_k cover low and high skewness. We also present the 90% and 95% probability coverages of confidence intervals for the BBSM model.

6.4.2 BBSM simulation results

Tables 19-20 report the empirical values of the biases and mean square errors of the ML and MM estimates, for the BBSM distribution. From these tables, we observe that, as n increases, the bias and MSE of all the estimators decrease, tending to be unbiased, as expected. In terms of MSE, the performances of the three methods are quite similar. From Tables 19–20, it is also worth noting that the ML and MM estimates are quite similar in terms of bias and MSE. Furthermore, we note that, as the values of the shape parameters δ_k increase, the performances of the estimators of μ_k, the scale parameters, deteriorate. In general, the results do not seem to depend on ρ.

Table 19 – Simulated values of biases and MSEs (within parentheses) of the MM in comparison with those of ML estimates ($\delta_k = 0.25$, $\mu_k = 2.0$, for $k = 1,2$), for the BBS distribution.

n		ML estimates				
	ρ	B($\widehat{\delta}_1$)	B($\widehat{\delta}_2$)	B($\widehat{\mu}_1$)	B($\widehat{\mu}_2$)	B($\widehat{\rho}$)
10	0.00	0.0470 (0.0687)	0.0566 (0.0650)	−0.0250 (1.4560)	−0.0537 (1.3561)	−0.0119 (0.1181)
	0.25	0.0492 (0.0557)	0.0549 (0.0669)	−0.0884 (1.4203)	−0.0899 (1.3267)	−0.0057 (0.1092)
	0.50	0.0487 (0.0653)	0.0507 (0.0587)	−0.0430 (1.3532)	−0.0218 (1.4600)	−0.0082 (0.0680)
	0.95	0.0480 (0.0638)	0.0467 (0.0766)	−0.0374 (1.2801)	−0.0265 (1.2389)	−0.0077 (0.0028)
50	0.00	0.0073 (0.0031)	0.0105 (0.0034)	−0.0157 (0.2621)	−0.0209 (0.2768)	−0.0071 (0.0218)
	0.25	0.0115 (0.0035)	0.0099 (0.0032)	−0.0224 (0.2560)	−0.0122 (0.2622)	−0.0037 (0.0178)
	0.50	0.0093 (0.0035)	0.0100 (0.0035)	−0.0178 (0.2748)	−0.0100 (0.2743)	−0.0015 (0.0118)
	0.95	0.0058 (0.0033)	0.0048 (0.0031)	−0.0035 (0.2255)	−0.0042 (0.2320)	−0.0003 (0.0002)
100	0.00	0.0040 (0.0015)	0.0057 (0.0016)	−0.0121 (0.1385)	−0.0048 (0.1360)	−0.0007 (0.0105)
	0.25	0.0026 (0.0014)	0.0053 (0.0014)	−0.0067 (0.1466)	−0.0220 (0.1257)	−0.0004 (0.0101)
	0.50	0.0053 (0.0014)	0.0016 (0.0014)	−0.0159 (0.1256)	−0.0100 (0.1297)	−0.0019 (0.0059)
	0.95	0.0038 (0.0015)	0.0036 (0.0015)	−0.0043 (0.1091)	−0.0067 (0.1118)	−0.0001 (0.0001)
n		MM estimates				
	ρ	B($\widetilde{\delta}_1$)	B($\widetilde{\delta}_2$)	B($\widetilde{\mu}_1$)	B($\widetilde{\mu}_2$)	B($\widetilde{\rho}$)
10	0.00	0.0517 (0.0695)	0.0612 (0.0658)	−0.0386 (1.3952)	−0.0633 (1.3487)	−0.0119 (0.1103)
	0.25	0.0541 (0.0565)	0.0598 (0.0679)	−0.1187 (1.3182)	−0.1224 (1.2675)	−0.0021 (0.1038)
	0.50	0.0554 (0.0670)	0.0575 (0.0597)	−0.0435 (1.3913)	−0.0602 (1.5602)	−0.0228 (0.0669)
	0.95	0.0595 (0.0673)	0.0583 (0.0797)	−0.0422 (1.3808)	−0.0230 (1.3971)	−0.0122 (0.0032)
50	0.00	0.0080 (0.0031)	0.0112 (0.0034)	−0.0164 (0.2723)	−0.0299 (0.2884)	−0.0068 (0.0213)
	0.25	0.0123 (0.0035)	0.0107 (0.0032)	−0.0230 (0.2692)	−0.0123 (0.2724)	−0.0019 (0.0175)
	0.50	0.0104 (0.0035)	0.0111 (0.0035)	−0.0130 (0.2971)	−0.0077 (0.3001)	−0.0052 (0.0118)
	0.95	0.0083 (0.0034)	0.0073 (0.0032)	−0.0058 (0.2657)	−0.0132 (0.2684)	−0.0006 (0.0002)
100	0.00	0.0043 (0.0015)	0.0061 (0.0016)	−0.0095 (0.1463)	−0.0027 (0.1460)	−0.0007 (0.0104)
	0.25	0.0030 (0.0014)	0.0057 (0.0014)	−0.0084 (0.1547)	−0.0184 (0.1346)	−0.0016 (0.0100)
	0.50	0.0059 (0.0014)	0.0022 (0.0014)	−0.0205 (0.1415)	−0.0027 (0.1439)	−0.0038 (0.0059)
	0.95	0.0051 (0.0015)	0.0049 (0.0015)	−0.0061 (0.1333)	−0.0079 (0.1364)	−0.0004 (0.0001)

Table 20 – Simulated values of biases and MSEs (within parentheses) of the ML in comparison with those of MM estimates ($\delta_k = 2.0$, $\mu_k = 2.0$, for $k = 1,2$), for the BBSM distribution.

n		ML estimates				
	ρ	B($\widehat{\delta}_1$)	B($\widehat{\delta}_2$)	B($\widehat{\mu}_1$)	B($\widehat{\mu}_2$)	B($\widehat{\rho}$)
10	0.00	0.4288 (4.0119)	0.4007 (3.8598)	−0.0241 (0.4009)	−0.0393 (0.4058)	−0.0037 (0.1058)
	0.25	0.3783 (3.0608)	0.3574 (4.2999)	−0.0361 (0.4112)	−0.0219 (0.3867)	−0.0356 (0.1023)
	0.50	0.3771 (3.4396)	0.3888 (2.7985)	−0.0334 (0.3855)	−0.0621 (0.3855)	−0.0222 (0.0678)
	0.95	0.3996 (4.0946)	0.3819 (4.2157)	−0.0383 (0.3981)	−0.0334 (0.3869)	−0.0064 (0.0025)
50	0.00	0.0726 (0.2149)	0.0605 (0.2057)	−0.0064 (0.0763)	−0.0101 (0.0767)	−0.0030 (0.0196)
	0.25	0.0375 (0.1969)	0.0732 (0.2061)	−0.0076 (0.0801)	−0.0174 (0.0822)	−0.0074 (0.0170)
	0.50	0.0755 (0.2147)	0.0873 (0.2179)	−0.0169 (0.0803)	−0.0020 (0.0775)	−0.0015 (0.0117)
	0.95	0.0548 (0.2031)	0.0623 (0.2043)	−0.0021 (0.0782)	−0.0026 (0.0797)	−0.0007 (0.0002)
100	0.00	0.0316 (0.0847)	0.0382 (0.0940)	−0.0131 (0.0369)	−0.0065 (0.0383)	−0.0069 (0.0106)
	0.25	0.0386 (0.0927)	0.0378 (0.0991)	−0.0017 (0.0429)	−0.0015 (0.0425)	−0.0026 (0.0090)
	0.50	0.0441 (0.0959)	0.0356 (0.0955)	−0.0036 (0.0376)	−0.0066 (0.0375)	−0.0062 (0.0061)
	0.95	0.0293 (0.0837)	0.0344 (0.0824)	−0.0014 (0.0372)	−0.0028 (0.0366)	−0.0003 (0.0001)
n		MM estimates				
	ρ	B($\widetilde{\delta}_1$)	B($\widetilde{\delta}_2$)	B($\widetilde{\mu}_1$)	B($\widetilde{\mu}_2$)	B($\widetilde{\rho}$)
10	0.00	0.4310 (4.0159)	0.4030 (3.8628)	−0.0244 (0.3965)	−0.0387 (0.4119)	−0.0038 (0.1046)
	0.25	0.3809 (3.0624)	0.3598 (4.3061)	−0.0371 (0.4085)	−0.0219 (0.3885)	−0.0368 (0.1015)
	0.50	0.3819 (3.4418)	0.3935 (2.8012)	−0.0368 (0.3752)	−0.0653 (0.3854)	−0.0242 (0.0674)
	0.95	0.4124 (4.0979)	0.3946 (4.2191)	−0.0382 (0.4099)	−0.0319 (0.4012)	−0.0070 (0.0025)
50	0.00	0.0728 (0.2149)	0.0607 (0.2057)	−0.0055 (0.0768)	−0.0003 (0.0763)	−0.0030 (0.0195)
	0.25	0.0379 (0.1969)	0.0736 (0.2061)	−0.0070 (0.0796)	−0.0175 (0.0823)	−0.0070 (0.0169)
	0.50	0.0764 (0.2147)	0.0882 (0.2180)	−0.0172 (0.0814)	−0.0018 (0.0783)	−0.0021 (0.0117)
	0.95	0.0573 (0.2032)	0.0648 (0.2044)	−0.0001 (0.0797)	−0.0048 (0.0818)	−0.0008 (0.0002)
100	0.00	0.0317 (0.0847)	0.0382 (0.0940)	−0.0129 (0.0370)	−0.0065 (0.0385)	−0.0069 (0.0105)
	0.25	0.0387 (0.0927)	0.0380 (0.0991)	−0.0026 (0.0430)	−0.0024 (0.0425)	−0.0028 (0.0090)
	0.50	0.0445 (0.0960)	0.0360 (0.0955)	−0.0027 (0.0380)	−0.0056 (0.0377)	−0.0066 (0.0061)
	0.95	0.0305 (0.0838)	0.0356 (0.0824)	−0.0001 (0.0380)	−0.0030 (0.0376)	−0.0003 (0.0001)

6.4.3 Probability coverage simulation results

We compute the 90% and 95% probability coverages of confidence intervals for the BBSM model using the asymptotic distributions given earlier, with $\alpha_k = 0.5$, $\beta_k = 1.0$, for $k = 1,2$. The $100(1-\varpi)\%$ confidence intervals for θ_l, $l = 1,\ldots,5$, based on the ML estimates can be obtained from

$$\left[\left(\widehat{\psi}_l + \frac{z_{\frac{\varpi}{2}}}{\sqrt{\boldsymbol{J}_{ll}(\widehat{\boldsymbol{\psi}})}}\right), \left(\widehat{\psi}_l + \frac{z_{1-\frac{\varpi}{2}}}{\sqrt{\boldsymbol{J}_{ll}(\widehat{\boldsymbol{\psi}})}}\right)\right],$$

respectively, where $\widehat{\boldsymbol{\psi}} = (\widehat{\psi}_1, \widehat{\psi}_2, \widehat{\psi}_3, \widehat{\psi}_4, \widehat{\psi}_5)^\top = (\widehat{\mu}_1, \widehat{\delta}_1, \widehat{\mu}_2, \widehat{\delta}_2, \widehat{\rho})^\top$ and z_r is the $100r$th percentile of the standard normal distribution. The corresponding $100(1-\varpi)\%$ confidence intervals for μ_k and δ_k, $k = 1,2$, based on the MM estimates are given by

$$\left[\widetilde{\mu}_k\left(1 + z_{\frac{\varpi}{2}}\sqrt{\frac{h(\widetilde{\delta}_k)}{n}}\right)^{-1}, \widetilde{\mu}_k\left(1 + z_{1-\frac{\varpi}{2}}\sqrt{\frac{h(\widetilde{\delta}_k)}{n}}\right)^{-1}\right],$$

and

$$\left[\widetilde{\delta}_k\left(1+z_{\frac{\varpi}{2}}\sqrt{\frac{2}{n}}\right)^{-1},\widetilde{\delta}_k\left(1+z_{1-\frac{\varpi}{2}}\sqrt{\frac{2}{n}}\right)^{-1}\right],$$

where $h(x)=\frac{2x+5}{[x+1]^2}$. To obtain $100(1-\varpi)\%$ confidence interval for ρ based on the MM estimator from $\widetilde{\rho}_k$ we can make use of the Fisher's z-transformation Fisher (1915) and the generalized confidence interval proposed by Krishnamoorthy and Xia (2007). The latter method is suggested by Kazemi and Jafari (2015) as one of the best approaches to construct confidence interval for the correlation coefficient in a bivariate normal distribution.

First, note that

$$\begin{aligned} X_1 &= \sqrt{\frac{\delta_1}{2}}\left(\sqrt{\frac{(\delta_1+1)T_1}{\mu_1\delta_1}}-\sqrt{\frac{\mu_1\delta_1}{(\delta_1+1)T_1}}\right)\sim \mathrm{N}(0,1) \quad \text{and} \\ X_2 &= \sqrt{\frac{\delta_2}{2}}\left(\sqrt{\frac{(\delta_2+1)T_2}{\mu_2\delta_2}}-\sqrt{\frac{\mu_2\delta_2}{(\delta_2+1)T_2}}\right)\sim \mathrm{N}(0,1). \end{aligned}$$

Note also that the we can express $\widetilde{\rho}$ as

$$\widetilde{\rho}=\frac{\sum_{i=1}^n x_{1i}x_{2i}}{\sqrt{\sum_{i=1}^n x_{1i}^2}\sqrt{\sum_{i=1}^n x_{2i}^2}},$$

where $x_{1i}=\sqrt{\frac{\delta_1}{2}}\left(\sqrt{\frac{(\delta_1+1)t_{1i}}{\mu_1\delta_1}}-\sqrt{\frac{\mu_1\delta_1}{(\delta_1+1)t_{1i}}}\right)$ and $x_{2i}=\sqrt{\frac{\delta_2}{2}}\left(\sqrt{\frac{(\delta_2+1)t_{2i}}{\mu_2\delta_2}}-\sqrt{\frac{\mu_2\delta_2}{(\delta_2+1)t_{2i}}}\right)$. The pairs (x_{1i},x_{2i}) for $i=1,\ldots,n$ can be thought of as realizations of the pair (X_1,X_2). Then, $\widetilde{\rho}$ is an estimator of the correlation coefficient of a standard bivariate normal distribution. Below, we detail the two methods to compute the confidence interval.

Fisher's z-transformation (FI)

Based on the Fisher's z-transformation Fisher (1915), we readily have

$$z=\frac{1}{2}\log\left(\frac{1+\widetilde{\rho}}{1-\widetilde{\rho}}\right)=\tanh^{-1}(\widetilde{\rho}),$$

which has an asymptotic normal distribution with mean $\frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right)=\tanh^{-1}(\rho)$ and variance $1/(n-3)$. Then, we can obtain an approximate $100(1-\varpi)\%$ confidence interval for ρ by

$$\left[\tanh\left(\widetilde{\rho}+\frac{z_{\frac{\varpi}{2}}}{\sqrt{n-3}}\right),\tanh\left(\widetilde{\rho}+\frac{z_{1-\frac{\varpi}{2}}}{\sqrt{n-3}}\right)\right].$$

Krishnamoorthy and Xia's method (KX)

Based on Krishnamoorthy and Xia (2007) we can construct an approximate $100(1-\varpi)\%$ confidence interval for ρ from the following Algorithm 5

Algorithm 5 – Confidence interval for ρ from KX method

1: Compute $\overline{\rho} = \frac{\widetilde{\rho}}{\sqrt{1-\widetilde{\rho}^2}}$ for a given n and $\widetilde{\rho}$.

2: For $i = 1$ to m (1,000,000 say), generate $U_1 \sim \chi^2_{n-1}$, $U_2 \sim \chi^2_{n-2}$ and $Z_0 \sim N(0,1)$ and compute

$$Q_i = \frac{\overline{\rho}\sqrt{U_2} - Z_0}{\sqrt{(\overline{\rho}\sqrt{U_2} - Z_0)^2 + U_1}}.$$

The upper and lower limits for ρ are the $100(\varpi)$th and $100(1-\varpi)$th percentiles of the Q_i's. Table 21 presents the 90% and 95% probability coverages of confidence intervals. The results show that the asymptotic confidence intervals do no provide good results for δ_k and μ_k when the sample size is small ($n = 10$), since the coverage probabilities are much lower than the corresponding nominal values. The scenario changes when $n = 50$ and 100 with satisfactory results for both δ_k and μ_k. Overall, the coverages for ρ associated with the MM estimates have quite good performances, whereas the coverages based on the ML estimates have poor performances.

Table 21 – Probability coverages of 90% and 95% confidence intervals for the BBSM model ($\mu_k = 1.0$, $\delta_k = 0.5$, for $k = 1, 2$).

		ML									
n		90%					95%				
	ρ	δ_1	δ_2	μ_1	μ_2	ρ	δ_1	δ_2	μ_1	μ_2	ρ
10	0.00	78.90	79.32	82.74	81.76	76.98	84.24	84.18	88.00	87.96	84.18
	0.25	79.52	78.44	81.08	82.12	77.44	84.34	83.96	86.28	87.66	83.16
	0.50	78.88	79.68	75.58	75.44	77.42	84.08	83.62	83.30	84.06	83.98
	0.95	78.64	78.82	37.58	36.26	79.24	83.63	82.68	41.75	44.40	83.42
50	0.00	87.66	88.58	87.64	89.30	88.56	92.24	92.62	93.32	93.06	93.20
	0.25	87.16	88.18	87.14	86.56	87.28	92.44	92.78	93.06	92.18	93.54
	0.50	87.34	87.86	83.02	83.08	88.12	93.38	93.18	90.42	90.14	93.54
	0.95	88.54	88.26	38.54	37.82	88.28	93.32	92.86	44.20	44.74	93.32
100	0.00	89.00	89.70	89.30	89.10	87.50	94.00	95.10	94.10	94.40	92.90
	0.25	89.10	88.90	89.20	88.60	90.00	95.10	92.30	93.60	92.10	93.80
	0.50	89.30	89.80	83.90	84.40	90.50	92.60	94.10	90.50	89.80	93.10
	0.95	90.10	89.90	38.44	38.44	88.20	93.00	94.30	48.00	48.50	93.80

		MM											
n		90%						95%					
	ρ	δ_1	δ_2	μ_1	μ_2	ρ (FI)	ρ (KX)	δ_1	δ_2	μ_1	μ_2	ρ (FI)	ρ (KX)
10	0.00	78.90	79.34	84.94	83.96	90.42	89.46	84.22	84.16	89.82	89.52	94.54	95.78
	0.25	79.50	78.48	84.64	84.88	90.08	90.28	84.32	83.96	89.28	90.24	95.48	95.14
	0.50	78.78	79.62	83.98	84.06	90.14	89.44	84.04	83.58	89.98	90.00	94.94	95.18
	0.95	78.58	78.86	84.64	83.94	90.64	90.28	83.53	82.68	90.81	89.65	95.62	94.61
50	0.00	87.66	88.58	88.14	89.52	90.56	90.60	92.24	92.62	93.58	93.46	95.02	94.82
	0.25	87.16	88.16	88.78	88.40	89.96	89.68	92.44	92.74	94.02	93.34	95.16	95.42
	0.50	87.32	87.86	89.04	89.38	90.08	90.10	93.40	93.18	94.32	94.46	95.32	95.72
	0.95	88.48	88.24	88.24	88.12	90.50	90.46	93.24	92.92	93.64	93.64	95.36	95.14
100	0.00	89.00	89.70	90.00	89.70	90.30	88.30	94.03	95.10	94.20	94.40	94.60	94.20
	0.25	89.10	88.90	90.70	89.50	89.30	89.90	95.12	92.30	94.70	93.70	95.60	94.90
	0.50	89.40	89.80	89.50	90.40	88.30	90.90	92.65	94.19	95.10	94.60	96.50	94.20
	0.95	90.00	89.80	89.90	89.30	90.10	89.80	93.00	94.38	95.00	95.00	95.10	95.30

6.4.4 Example 1

In this example, the data set corresponds to two different measurements of stiffness, namely, shock and vibration of each of $n = 30$ boards. The former concerns the emission of shock wave down the board, while the latter is obtained during the vibration of the board; see Johnson and Wichern (2007). We consider probability versus probability (PP) plots with acceptance bands based on the marginal distributions of T_1 and T_2 to support the BBSM model; see Figure 21(top). We also consider the TTT plots in Figure 21(bottom) to have an idea about the shape of the failure rate of the marginals; see Aarset (1987) and Azevedo et al. (2012). The failure rate of a random variable X is defined by $h(x) = f(x)/[1 - F(x)]$, where $f(\cdot)$ and $F(\cdot)$ are the PDF and CDF of X, respectively. From Figure 21, we observe that the PP plots support the BBSM model and the TTT plots suggest that both marginals have unimodal failure rates.

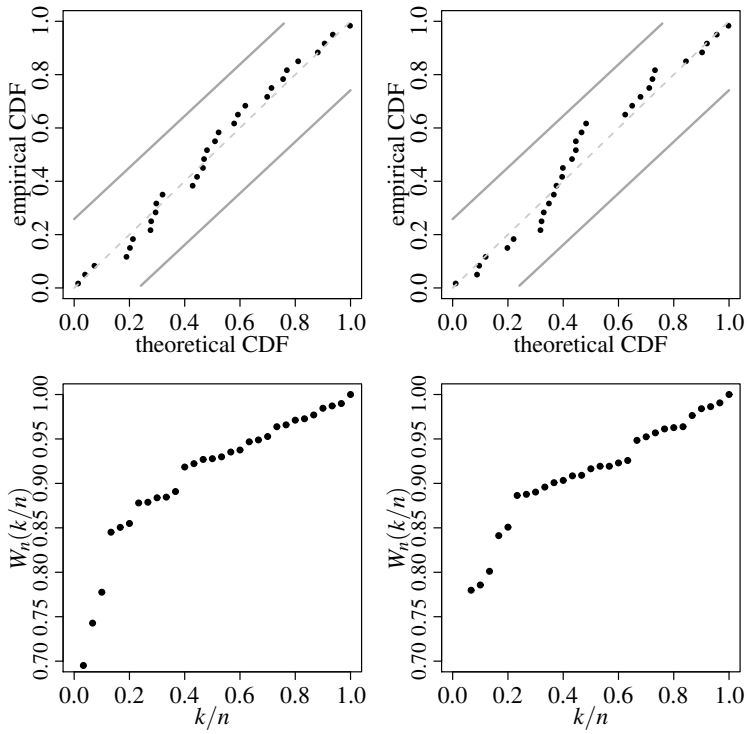

Figure 21 – PP plots with acceptance bands (top) and scaled TTT plots (bottom) for the two different measurements of stiffness.

We now fit the BBSM distribution to the stiffness data set. From the observations, we obtain $s_1 = 1906.1$, $r_1 = 1857.55$, $s_2 = 1749.53$ and $r_2 = 1699.99$. Table 22 presents the ML and MM estimates along with their corresponding 95% CIs, as well as the log-likelihood values. We note that across the models the log-likelihood values are quite similar. The Kolmogorov-Smirnov (KS) distances with the corresponding p-values (within brackets) between the empirical marginals and the fitted marginals for T_1 and T_2 are $0.0893(0.9531)$ and $0.1505(0.505)$, respectively. These results support the assumption of a BBSM model.

Table 22 – Estimates of the parameter for the indicated estimator.

Parameter	ML estimates	95% CIs	MM estimates	95% CIs
μ_1	1906.100	(1885.766,1926.435)	1906.100	(1795.857,2016.343)
δ_1	77.030	(69.885,76.279)	77.030	(38.048,108.116)
μ_2	1749.533	(1729.819,1769.247)	1749.533	(1642.688,1856.378)
δ_2	69.134	(62.720,75.548)	69.134	(34.148,104.121)
ρ	0.908	(0.897,0.920)	0.908	(0.814,0.956)
$\ell(\boldsymbol{\psi})$	−400.648		−400.648	

6.4.5 Example 2

Here, the data set corresponds to the bone mineral density (BMD) measured in g/cm^2 for 24 individuals included in a experimental study; see Johnson and Wichern (2007). The objective of the study was to determine whether exercise or dietary supplements would slow bone loss. The data represent the BMD of the bone dominant radius (t_1 and t_2). The explanatory variables associated are: radius ($x_1^{(k)}$), dominant humerus ($x_2^{(k)}$), humerus ($x_3^{(k)}$), dominant ulna ($x_4^{(k)}$) and ulna ($x_5^{(k)}$), for $k = 1,2$. The PP and TTT plots showed in Figure 22 based on the marginal distributions of T_1 and T_2 support the assumed BBSM model and suggest that the marginals have unimodal hazard rates, respectively. In Table (23), we report the ML estimates, the LSEs utilized as initial values to determine the ML estimates of the regression coefficients, the corresponding p-values, and the 95% CIs. The KS distance between the empirical distribution function and the fitted distribution function and the p-value (within brackets) for T_1 and T_2 are $0.1541(0.5666)$ and $0.1649(0.4812)$, respectively. Therefore, taking into account the KS distances, the BBSM distribution is indeed a good model for the BMD data.

Table 23 – Estimates of the parameter for the indicated estimator.

Parameter	MLE	LSE	p-value	95% CIs
δ_1	311.358	-		(251.676,371.040)
δ_2	166.890	-		(140.226,193.554)
φ_{10}	−1.220	−1.299	<0.001	(−1.273,−1.167)
φ_{11}	0.650	1.082	0.014	(0.543,0.756)
φ_{12}	0.163	0.482	0.281	(0.102,0.223)
φ_{13}	−0.069	−0.509	0.755	(−0.157,0.020)
φ_{14}	0.680	0.263	0.005	(0.582,0.778)
φ_{15}	−0.197	0.104	0.366	(−0.284,−0.110)
φ_{20}	−1.227	−1.196	<0.001	(−1.297,−1.157)
φ_{21}	0.533	1.125	0.130	(0.392,0.674)
φ_{22}	0.166	0.412	0.268	(0.106,0.227)
φ_{23}	0.011	−0.321	0.957	(−0.072,0.094)
φ_{24}	0.444	−0.156	0.182	(0.311,0.577)
φ_{25}	−0.023	0.041	0.911	(−0.106,0.060)
ρ	0.890	-		(0.853,0.926)

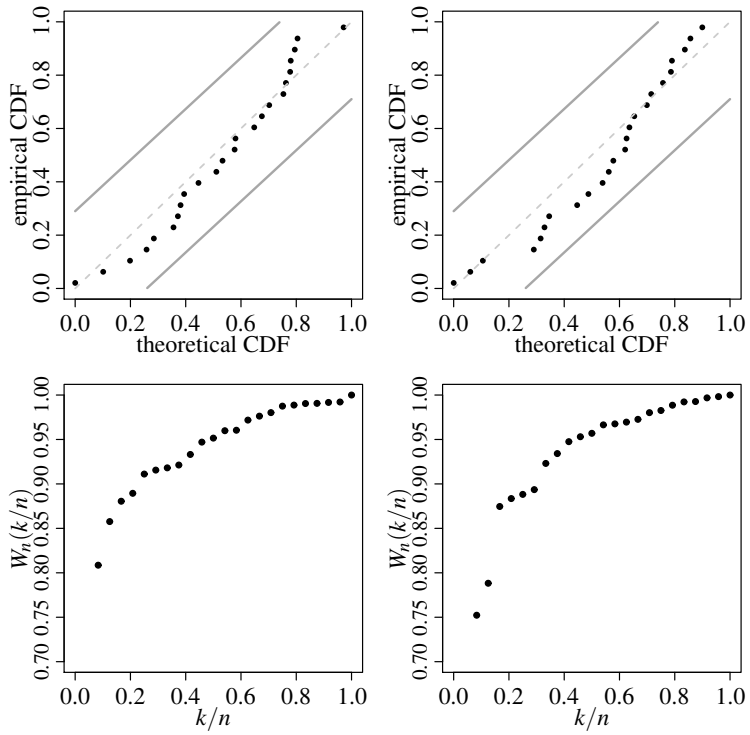

Figure 22 – PP plots with acceptance bands (top) and scaled TTT plots (bottom) for the BMD data.

6.5 Concluding remarks

In this chapter a bivariate Birnbaum-Saunders which is characterized by two mean parameters is introduced. We have discussed maximum likelihood estimation and modified moment estimation of the five parameters. We also have proposed a new bivariate Birnbaum-Saunders regression model, which allows us to describe the means of the bivariate data in their original scale. A simulation study is carried out to evaluate the performance of the maximum likelihood estimators. The probability coverages of confidence intervals are also discussed. Finally, two examples are performed to illustrate all the methods proposed here.

CHAPTER

7

DISCUSSION, CONCLUSIONS AND FUTURE RESEARCH

The inclusion of a frailty parameter in the medical data modeling brings additional information that may be useful in practice. We have considered a source of unobserved heterogeneity that is not captured by covariables. This introduced a frailty component in the hazard rate structure. Then, the effects of omitted covariables can be captured. No measurable biological variations among patients are detected, which justifies the presence of heterogeneity. For instance, some patients may have a genetic disposition with respect to certain disease, having an increasing risk of developing it compared to others. The heterogeneity therefore affects the observed survival times.

In this work, we have focused on the use of the reparameterized Birnbaum-Saunders distribution to propose frailty, regression frailty and cure rate frailty models applied to medical data. Influence diagnostic tools were derived in the regression case to evaluate the effect of atypical observations on the model. In addition, we have studied the bivariate Birnbaum-Saunders distribution based on its mean.

The present study leaves some open topics to be addressed in the future. As part of a future research, we leave open the following issues. Economou and Caroni (2008) introduced a manner to construct diagnostic plots and to verify the correct choice of the frailty distribution. This was conducted in the case of proportional hazard models and for exponential family members, which include the gamma and inverse Gaussian models. These diagnostic plots are based on the closure property which holds for the exponential family and says that the distribution among survivors belongs to the same family of distributions; see Hougaard (1984). The Birnbaum-Saunders model does not hold this property. However, as mentioned in Subsection 2.4, it is closed under reciprocation, which allows us to construct a graphical procedure, based on the TTT plot, to verify the correct choice of the frailty distribution in a proportional hazard model; see Athayde (2016). The construction of this diagnostic plot is beyond the scope of our study, so that we will

consider it as part of a future work. In addition, the proposed methodology can be extended into a model with a nonparametric baseline hazard rate. One can also consider a Bayesian approach to estimate the model parameters. Furthermore, it might be of interest to consider the likelihood ratio method for testing homogeneity of the frailty term; see Economou and Stehlik (2015). The proposed methodology can also be extended to the correlated frailty case, where the nature of the heterogeneity and the dependence are explicitly specified, and are of main importance; see Petersen (1998). Finally, the inclusion of multivariate aspects in frailty models, as well as spatial components, can also be considered; see Garcia-Papani et al. (2016) and Marchant et al. (2016a), Marchant et al. (2016b). In addition from the bivariate Birnbaum-Saunders it would be of interest to develop likelihood inferential methods by considering censored data. Also, the extension of the proposed bivariate regression model to the multivariate case would be of practical relevance.

BIBLIOGRAPHY

AALEN, O.; BORGAN, O.; GJESSING, H. **Survival and event history analysis: a process point of view**. [S.l.]: Springer Science & Business Media, 2008. Cited 2 times on pages 28 and 44.

AALEN, O. O.; TRETLI, S. Analyzing incidence of testis cancer by means of a frailty model. **Cancer Causes & Control**, Springer, v. 10, n. 4, p. 285–292, 1999. Cited on page 26.

AARSET, M. How to identify a bathtub hazard rate. **IEEE Transactions on Reliability**, v. 36, p. 106–108, 1987. Cited on page 99.

ANDROULAKIS, E.; KOUKOUVINOS, C.; VONTA, F. On the uniform frailty model with penalized likelihood and clustered data. **Journal of Reliability and Statistical Studies**, v. 5, p. 98, 2012. Cited on page 26.

ATHAYDE, E. A characterization of the Birnbaum-Saunders distribution. **REVSTAT Statistical Journal**, p. in press available at <https://www.ine.pt/revstat/inicio.html>, 2016. Cited on page 103.

AZEVEDO, C.; LEIVA, V.; ATHAYDE, E.; BALAKRISHNAN, N. Shape and change point analyses of the birnbaum–saunders-t hazard rate and associated estimation. **Computational Statistics & Data Analysis**, Elsevier, v. 56, n. 12, p. 3887–3897, 2012. Cited 2 times on pages 53 and 99.

BALAKRISHNAN, N.; PENG, Y. Generalized gamma frailty model. **Statistics in medicine**, Wiley Online Library, v. 25, n. 16, p. 2797–2816, 2006. Cited on page 27.

BARKER, P.; HENDERSON, R. Small sample bias in the gamma frailty model for univariate survival. **Lifetime data analysis**, Springer, v. 11, n. 2, p. 265–284, 2005. Cited on page 26.

BERKSON, J.; GAGE, R. P. Survival curve for cancer patients following treatment. **Journal of the American Statistical Association**, v. 47, p. 501–515, 1952. Cited on page 28.

BICKEL, P. J.; DOKSUM, K. A. An analysis of transformations revisited. **Journal of the american statistical association**, Taylor & Francis Group, v. 76, n. 374, p. 296–311, 1981. Cited on page 65.

BIRNBAUM, Z.; SAUNDERS, S. C. A new family of life distributions. **Journal of applied probability**, JSTOR, p. 319–327, 1969. Cited 7 times on pages 27, 33, 34, 35, 36, 37, and 38.

BOAG, J. W. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. **Journal of the Royal Statistical Society B)**, v. 11, p. 15–53, 1949. Cited on page 28.

BRENT, R. P. **Algorithms for minimization without derivatives**. [S.l.]: Courier Corporation, 2013. Cited on page 50.

CAI, B. Bayesian semiparametric frailty selection in multivariate event time data. **Biometrical Journal**, v. 52, p. 171–185, 2010. Cited on page 26.

CALSAVARA, V. F.; RODRIGUES, A. S.; TOMAZELLA, V. L. D.; CASTRO, M. de. Frailty models power variance function (pvf) with cure fraction and latent risk factors negative binomial. **Communications in Statistics-Theory and Methods**, Taylor & Francis, n. just-accepted, 2016. Cited 2 times on pages 29 and 65.

CANCHO, V. G.; LOUZADA, F.; BARRIGA, G. D. C. The geometric Birnbaum-Saunders regression model with cure rate. **Journal of Statistical Planning and Inference**, v. 142, p. 993–1000, 2012. Cited on page 28.

CANCHO, V. G.; RODRIGUES, J.; CASTRO, M. de. A flexible model for survival data with a cure rate: a Bayesian approach. **Journal of Applied Statistics**, v. 38, p. 57–70, 2011. Cited 3 times on pages 28, 29, and 80.

CASTRO, M. de; CANCHO, V. G.; RODRIGUES, J. A Bayesian long-term survival model parametrized in the cured fraction. **Biometrical Journal**, v. 51, p. 443–455, 2009. Cited 2 times on pages 28 and 80.

____. A hands-on approach for fitting long-term survival models under the GAMLSS framework. **Computer Methods and Programs in Biomedicine**, v. 97, p. 168–177, 2010. Cited on page 28.

CHAN, R. Z. K. C. G. A marginalizable frailty model for correlated right-censored data. **arXiv preprint arXiv:1403.6744**, 2014. Cited on page 26.

CHEN, M.; IBRAHIM, J. G.; SINHA, D. A new Bayesian model for survival data with a surviving fraction. **Journal of the American Statistical Association**, v. 94, p. 909–919, 1999. Cited 2 times on pages 28 and 45.

CLAYTON, D. G. A monte carlo method for bayesian inference in frailty models. **Biometrics**, JSTOR, p. 467–485, 1991. Cited on page 43.

COLLETT, D. **Modelling Survival Data in Medical Research**. Boca Raton, US: Chapman and Hall/CRC, 2015. Cited on page 62.

COOK, R. D. Influence assessment. **Journal of Applied Statistics**, v. 14, p. 117–131, 1987. Cited 3 times on pages 61, 66, and 82.

CORDEIRO, G. M.; CANCHO, V. G.; ORTEGA, E. M. M.; BARRIGA, G. D. C. A model with long-term survivors: Negative binomial birnbaum-saunders. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 45, n. 5, p. 1370–1387, 2016. Cited on page 28.

COX, D.; HINKLEY, D. **Theoretical Statistics**. London, UK: Chapman and Hall, 1974. Cited 4 times on pages 63, 81, 91, and 94.

COX, D. R. Regression models and life-tables. **Journal of the Royal Statistical Society. Series B (Methodological)**, JSTOR, p. 187–220, 1972. Cited 2 times on pages 25 and 26.

COX, D. R.; HINKLEY, D. V. **Theoretical statistics**. [S.l.]: CRC Press, 1979. Cited on page 49.

CRAMÉR, H. **Mathematical methods of statistics**. [S.l.]: JSTOR, 1947. Cited on page 36.

CROW, E. L.; SHIMIZU, K. **Lognormal Distributions: Theory and Applications**. New York, US: Dekker, 1988. Cited on page 40.

DESMOND, A. Stochastic models of failure in random environments. **The Canadian Journal of Statistics**, JSTOR, p. 171–183, 1985. Cited 4 times on pages 27, 36, 38, and 39.

DUCHATEAU, L.; JANSSEN, P. **The Frailty Model**. New York, US: Springer, 2008. Cited on page 26.

DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals. **Journal of Computational and Graphical Statistics**, Taylor & Francis, v. 5, n. 3, p. 236–244, 1996. Cited on page 50.

DíAZ-GARCIA, J.; LEIVA, V. A new family of life distributions based on elliptically contoured distributions. **Journal of Statistical Planning and Inference**, v. 128, p. 445–457, 2005. Cited 2 times on pages 30 and 41.

ECONOMOU, P.; CARONI, C. Closure properties and diagnostic plots for the frailty distribution in proportional hazards models. In: VONTA, F.; NIKULIN, M. S.; LIMNIOS, N.; HUBER-CAROL, C. (Ed.). **Statistical Models and Methods for Biomedical and Technical Systems**. Boston, US: Birkhäuser, 2008. chap. 4, p. 43–53. Cited on page 103.

ECONOMOU, P.; STEHLIK, M. On small samples testing for frailty through homogeneity test. **Communications in Statistics: Simulation and Computation**, v. 44, p. 40–65, 2015. Cited on page 104.

EFRON, B.; HINKLEY, D. Assessing the accuracy of the maximum likelihood estimator: Observed vs. expected Fisher information. **Biometrika**, v. 65, p. 457–487, 1978. Cited 2 times on pages 63 and 81.

ELBERS, C.; RIDDER, G. True and spurious duration dependence: The identifiability of the proportional hazard model. **The Review of Economic Studies**, Oxford University Press, v. 49, n. 3, p. 403–409, 1982. Cited on page 47.

ENKI, D. G.; NOUFAILY, A.; FARRINGTON, C. P. et al. A time-varying shared frailty model with application to infectious diseases. **The Annals of Applied Statistics**, Institute of Mathematical Statistics, v. 8, n. 1, p. 430–447, 2014. Cited on page 26.

ESPINHEIRA, P.; FERRARI, S.; CRIBARI-NETO, F. Influence diagnostics in beta regression. **Computational Statistics and Data Analysis**, v. 52, p. 4417–4431, 2008. Cited on page 61.

EUDES, A. M.; TOMAZELLA, V. L. D.; CALSAVARA, V. F. Modelagem de sobrevivência com fração de cura para dados de tempo de vida weibull modificada. **Rev. Bras. Biom**, v. 30, n. 3, p. 326–342, 2013. Cited on page 28.

FAN, J.; LI, R. Variable selection for cox's proportional hazards model and frailty model. **Annals of Statistics**, JSTOR, p. 74–99, 2002. Cited on page 26.

FEIGL, P.; ZELEN, M. Estimation of exponential survival probabilities with concomitant information. **Biometrics**, JSTOR, p. 826–838, 1965. Cited 2 times on pages 53 and 69.

FISHER, R. A. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. **Biometrika**, JSTOR, v. 10, n. 4, p. 507–521, 1915. Cited on page 97.

FROST, N.; DUGDALE, D. The propagation of fatigue cracks in sheet specimens. **Journal of the Mechanics and Physics of Solids**, Elsevier, v. 6, n. 2, p. 92–110, 1958. Cited on page 37.

GARCIA-PAPANI, F.; URIBE-OPAZO, M.; LEIVA, V.; AYKROYD, R. Birnbaum-Saunders spatial modelling and diagnostics applied to agricultural engineering data. **Stochastic Environmental Research and Risk Assessment**, p. in press available at <http://dx.doi.org/10.1007/s00477--015--1204--4>, 2016. Cited on page 104.

GONZALES, J. F. B.; TOMAZELLA, V. L. D.; TACONELLI, J. P. Estimação paramétrica do modelo de mistura com fragilidade gamma na presenç de covariáveis. **Rev. Bras. Biom**, v. 31, n. 2, p. 233–247, 2013. Cited on page 29.

GREENWOOD, M.; YULE, G. U. An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. **Journal of the Royal statistical society**, JSTOR, p. 255–279, 1920. Cited on page 26.

HOUGAARD, P. Life table methods for heterogeneous populations: distributions describing the heterogeneity. **Biometrika**, Biometrika Trust, v. 71, n. 1, p. 75–83, 1984. Cited 2 times on pages 44 and 103.

_____. Frailty models for survival data. **Lifetime data analysis**, Springer, v. 1, n. 3, p. 255–273, 1995. Cited on page 27.

_____. **Analysis of Multivariate Survival Data**. New York, US: Springer, 2000. Cited on page 26.

HUBERT, M.; VANDERVIEREN, E. An adjusted boxplot for skewed distributions. **Computational statistics & data analysis**, Elsevier, v. 52, n. 12, p. 5186–5201, 2008. Cited on page 54.

IBRAHIM, J. G.; CHEN, M.; SINHA, D. **Bayesian Survival Analysis**. New York, US: Wiley, 2005. Cited on page 28.

JAMALIZADEH, A.; KUNDU, D. A multivariate birnbaum-saunders distribution based on the multivariate skew normal distribution. **Journal of the Japan Statistical Society**, v. 45, n. 1, p. 1–20, 2015. Cited on page 41.

JOHNSON, N.; KOTZ, S.; BALAKRISHNAN, N. **Continuous Univariate Distributions**. New York, US: Wiley, 1995. Cited on page 40.

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. New Yersey, US: Prentice-Hall, 2007. Cited 2 times on pages 99 and 100.

JONES, M. C. On reciprocal symmetry. **Journal of Statistical Planning and Inference**, v. 138, p. 3039–3043, 2008. Cited on page 39.

KALBFLEISCH, J. D.; PRENTICE, R. L. **The statistical analysis of failure time data**. [S.l.]: John Wiley & Sons, 2011. 32 – 34 p. Cited 3 times on pages 53, 56, and 72.

KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. **Journal of the American statistical association**, Taylor & Francis, v. 53, n. 282, p. 457–481, 1958. Cited on page 25.

KAZEMI, M. R.; JAFARI, A. A. Comparing seventeen interval estimates for a bivariate normal correlation coefficient. **Journal of Statistics Applications & Probability Letters**, v. 2, p. 1–13, 2015. Cited on page 97.

KHOSRAVI, M.; KUNDU, D.; JAMALIZADEH, A. On Bivariate and Mixture of Bivariate Birnbaum-Saunders Distributions. **Statistical Methodology**, v. 23, p. 1–17, 2015. Cited 2 times on pages 30 and 41.

KLEIN, J. P.; MOESCHBERGER, M. L. **Survival analysis: techniques for censored and truncated data**. [S.l.]: Springer Science & Business Media, 2003. Cited on page 44.

KOCH, E.; NAVEAU, P. A frailty-contagion model for multi-site hourly precipitation driven by atmospheric covariates. **Advances in Water Resources**, Elsevier, v. 78, p. 145–154, 2015. Cited on page 27.

KOCHERLAKOTA, S. The bivariate inverse Gaussian distribution: an introduction. **Communications in Statistics: Theory and Methods**, v. 15, p. 1081–1112, 1986. Cited 2 times on pages 30 and 41.

KOTI, K. M. Gamma failure-time mixture models: Yet another way to establish efficacy. **Pharmaceutical Statistics**, v. 2, p. 133–144, 2003. Cited on page 28.

KOTZ, S.; LEIVA, V.; SANHUEZA, A. Two new mixture models related to the inverse Gaussian distribution. **Methodology and Computing in Applied Probability**, v. 12, p. 199–212, 2010. Cited on page 27.

____. Two new mixture models related to the inverse gaussian distribution. **Methodology and Computing in Applied Probability**, Springer, v. 12, n. 1, p. 199–212, 2010. Cited on page 40.

KRISHNAMOORTHY, K.; XIA, Y. Inferences on correlation coefficients: one-sample, independent and correlated cases. **Journal of Statistical Planning and Inference**, Elsevier, v. 137, n. 7, p. 2362–2379, 2007. Cited on page 97.

KUK, A. Y. C.; CHEN, C. A mixture model combining logistic regression with proportional hazards regression. **Biometrika**, v. 79, p. 531–541, 1992. Cited on page 28.

KUNDU, D. Bivariate log birnbaum–saunders distribution. **Statistics**, Taylor & Francis, v. 49, n. 4, p. 900–917, 2015. Cited on page 41.

KUNDU, D.; BALAKRISHNAN, N.; JAMALIZADEH, A. Bivariate Birnbaum-Saunders distribution and associated inference. **Journal of Multivariate Analysis**, v. 101, p. 113–125, 2010. Cited 3 times on pages 30, 41, and 42.

____. Generalized multivariate Birnbaum-Saunders distributions and related inferential issues. **Journal of Multivariate Analysis**, v. 116, p. 230–244, 2013. Cited 2 times on pages 30 and 41.

KUNDU, D. et al. Bivariate sinh-normal distribution and a related model. **Brazilian Journal of Probability and Statistics**, Brazilian Statistical Association, v. 29, n. 3, p. 590–607, 2015. Cited on page 41.

LANGE, K. **Numerical analysis for statisticians**. [S.l.]: Springer Science & Business Media, 2010. Cited on page 49.

LAWLESS, J. F. **Statistical models and methods for lifetime data**. [S.l.]: John Wiley & Sons, 2011. Cited 3 times on pages 26, 48, and 62.

LEHMANN, E. L.; CASELLA, G. **Theory of point estimation**. [S.l.]: Springer Science & Business Media, 2006. Cited on page 49.

LEIVA, V. **The Birnbaum-Saunders Distribution**. New York, US: Academic Press, 2016. Cited 2 times on pages 33 and 40.

LEIVA, V.; MARCHANT, C.; RUGGERI, F.; SAULO, H. A criterion for environmental assessment using birnbaum–saunders attribute control charts. **Environmetrics**, Wiley Online Library, v. 26, n. 7, p. 463–476, 2015. Cited on page 27.

LEIVA, V.; ROJAS, E.; GALEA, M.; SANHUEZA, A. Diagnostics in birnbaum–saunders accelerated life models with an application to fatigue data. **Applied Stochastic Models in Business and Industry**, Wiley Online Library, v. 30, n. 2, p. 115–131, 2014. Cited on page 61.

LEIVA, V.; RUGGERI, F.; SAULO, H.; VIVANCO, J. F. A methodology based on the Birnbaum-Saunders distribution for reliability analysis applied to nano-materials. **Reliability Engineering and System Safety**, v. 157, p. 192–201, 2017. Cited on page 27.

LEIVA, V.; SANTOS-NETO, M.; CYSNEIROS, F. J. A.; BARROS, M. Birnbaum–saunders statistical modelling: a new approach. **Statistical Modelling**, SAGE Publications, v. 14, n. 1, p. 21–48, 2014. Cited on page 39.

____. Birnbaum-Saunders statistical modelling: A new approach. **Statistical Modelling**, v. 14, p. 21–48, 2014. Cited on page 27.

LEIVA, V.; SAULO, H.; aO, J. L.; MARCHANT, C. A family of autoregressive conditional duration models applied to financial data. **Computational Statistics and Data Analysis**, v. 79, p. 175–191, 2014. Cited on page 27.

LEIVA, V.; TEJO, M.; GUIRAUD, P.; SCHMACHTENBERG, O.; ORIO, P.; MARMOLEJO, F. Modeling neural activity with cumulative damage distributions. **Biological Cybernetics**, v. 109, p. 421–433, 2015. Cited on page 27.

LEMONTE, A. J.; CRIBARI-NETO, F.; VASCONCELLOS, K. L. Improved statistical inference for the two-parameter birnbaum–saunders distribution. **Computational Statistics & Data Analysis**, Elsevier, v. 51, n. 9, p. 4656–4681, 2007. Cited on page 64.

LESAFFRE, E.; VERBEKE, G. Local influence in linear mixed models. **Biometrics**, v. 54, p. 570–582, 1998. Cited on page 82.

LIU, B.; LU, W.; ZHANG, J. Accelerated intensity frailty model for recurrent events data. **Biometrics**, v. 70, p. 579–587, 2014. Cited on page 26.

LONGINI, I. M.; HALLORAN, M. E. A frailty mixture model for estimating vaccine efficacy. **Applied Statistics**, v. 79, p. 165–173, 1996. Cited on page 29.

LOUZADA, F.; CASTRO, M. de; TOMAZELLA, V.; GONZALES, J. F. Modeling categorical covariates for lifetime data in the presence of cure fraction by bayesian partition structures. **Journal of Applied Statistics**, Taylor & Francis, v. 41, n. 3, p. 622–634, 2014. Cited on page 28.

MALLER, R.; ZHOU, X. **Survival Analysis with Long-term Survivors**. New York, US: Wiley, 1996. Cited on page 28.

MALLICK, M.; RAVISHANKER, N. Additive positive stable frailty models. **Methodology and Computing in Applied Probability**, Springer, v. 8, n. 4, p. 541–558, 2006. Cited on page 26.

MALLICK, M.; RAVISHANKER, N.; KANNAN, N. Bivariate positive stable frailty models. **Statistics & Probability Letters**, Elsevier, v. 78, n. 15, p. 2371–2377, 2008. Cited on page 26.

MANN, N. R.; SCHAFER, R. E.; SINGPURWALLA, N. D. **Methods for statistical analysis of reliability and life data**. [S.l.]: Wiley New York, 1974. Cited 2 times on pages 33 and 34.

MARCHANT, C.; LEIVA, V.; CYSNEIROS, F. A multivariate log-linear model for Birnbaum-Saunders distributions. **IEEE Transactions on Reliability**, v. 65, p. 816–827, 2016. Cited on page 104.

MARCHANT, C.; LEIVA, V.; CYSNEIROS, F.; VIVANCO, J. Diagnostics in multivariate generalized Birnbaum-Saunders regression models. **Journal of Applied Statistics**, p. in press available at <http://dx.doi.org/10.1080/02664763.2016.1148671>, 2016. Cited on page 104.

MARTINUSSEN, T.; SCHEIKE, T. H.; ZUCKER, D. M. The aalen additive gamma frailty hazards model. **Biometrika**, Biometrika Trust, p. asr049, 2011. Cited on page 26.

MAZROUI, Y.; MATHOULIN-PELISSIER, S.; MACGROGAN, G.; BROUSTE, V.; G., R. Multivariate frailty models for two types of recurrent events with a dependent terminal event: Application to breast cancer data. **Biometrical Journal**, v. 55, p. 866–884, 2013. Cited on page 26.

MIGON, H. S.; GAMERMAN, D.; LOUZADA, F. **Statistical inference: an integrated approach**. [S.l.]: CRC press, 2014. Cited on page 49.

NOCEDAL, J.; WRIGHT, S. **Numerical optimization**. [S.l.]: Springer Science & Business Media, 2006. Cited on page 49.

OSORIO, F.; PAULA, G.; GALEA, M. Assessment of local influence in elliptical linear models with longitudinal structure. **Computational Statistics and Data Analysis**, v. 51, p. 4354–4368, 2007. Cited on page 61.

PAULA, G. A.; MEDEIROS, M. J.; VILCA, F. Influence diagnostics for linear models with first autoregressive elliptical errors. **Statistical and Probability Letters**, v. 79, p. 339–346, 2009. Cited on page 61.

PENG, Y.; ZHANG, J. Estimation method of the semiparametric mixture cure gamma frailty model. **Statistics in Medicine**, v. 27, p. 5177–5194, 2008. Cited on page 80.

____. Identifiability of a mixture cure frailty model. **Statistics & Probability Letters**, Elsevier, v. 78, n. 16, p. 2604–2608, 2008. Cited on page 29.

PETERSEN, J. An additive frailty model for correlated life times. **Biometrics**, v. 54, p. 646–661, 1998. Cited on page 104.

PIEGORSCH, W. W. Maximum likelihood estimation for the negative binomial dispersion parameter. **Biometrics**, v. 46, p. 863–867, 1990. Cited 2 times on pages 45 and 79.

PRICE, D. L.; MANATUNGA, A. K. Modelling survival data with a cured fraction using frailty models. **Statistics in Medicine**, v. 20, p. 1515–1527, 2001. Cited on page 29.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2016. Available: <http://www.R-project.org/>. Cited 3 times on pages 49, 63, and 81.

RAHIMZADEH, M.; HAJIZADEH, E.; ESKANDARI, F. Non-mixture cure correlated frailty models in Bayesian approach. **Journal of Applied Statistics**, v. 38, p. 1651–1663, 2011. Cited on page 29.

RIECK, J.; NEDELMAN, J. A log-linear model for the Birnbaum-Saunders distribution. **Technometrics**, v. 3, p. 51–60, 1991. Cited on page 40.

RODRIGUES, J.; CANCHO, V. G.; CASTRO, M. de; LOUZADA, F. On the unification of long-term survival models. **Statistics and Probability Letters**, v. 79, p. 753–759, 2009. Cited 4 times on pages 28, 33, 45, and 79.

RODRIGUES, J.; CORDEIRO, G. M.; CANCHO, V. G.; BALAKRISHNAN, N. Relaxed poisson cure rate models. **Biometrical Journal**, Wiley Online Library, 2015. Cited on page 28.

RONDEAU, V.; SCHAFFNER, E.; CORBIÈRE, F.; GONZALEZ, J. R.; MATHOULIN-PÉLISSIER, S. Cure frailty models for survival data: Application to recurrences for breast cancer and to hospital readmissions for colorectal cancer. **Statistical methods in medical research**, Sage Publications, v. 22, n. 3, p. 243–260, 2013. Cited on page 29.

ROSS, G. J. S.; PREECE, D. A. The negative binomial distribution. **The Statistician**, v. 34, p. 323–335, 1985. Cited on page 80.

SAHA, K.; PAUL, S. Bias-corrected maximum likelihood estimator of the negative binomial dispersion parameter. **Biometrics**, v. 61, p. 179–185, 2005. Cited on page 45.

SANTOS-NETO, M.; CYSNEIROS, F. J. A.; LEIVA, V.; AHMED, S. E. On new parameterizations of the birnbaum-saunders distribution. **Pakistan Journal of Statistics**, v. 28, n. 1, 2012. Cited 7 times on pages 11, 13, 26, 27, 30, 38, and 44.

SANTOS-NETO, M.; CYSNEIROS, F. J. A.; LEIVA, V.; BARROS, M. A reparameterized birnbaum-saunders distribution and its moments, estimation and application. **REVSTAT–Statistical Journal**, v. 12, n. 3, p. 247–272, 2014. Cited 5 times on pages 11, 13, 30, 39, and 44.

SAULO, H.; LEIVA, V.; ZIEGELMANN, F. A.; MARCHANT, C. A nonparametric method for estimating asymmetric densities based on skewed Birnbaum-Saunders distributions applied to environmental data. **Stochastic Environmental Research and Risk Assessment**, v. 27, p. 1479–1491, 2013. Cited on page 27.

SAUNDERS, S. **Reliability, Life Testing and Prediction of Services Lives**. New York: Springer, 2007. Cited on page 33.

SERFLING, R. **Approximation Theorems of Mathematical Statistics**. New York: Wiley, 1980. Cited 2 times on pages 49 and 63.

SHAO, Q.; ZHOU, X. A new parametric model for survival data with long-term survivors. **Statistics in Medicine**, v. 23, p. 3525–3543, 2004. Cited on page 28.

SIQUEIRA, A. L.; TAYLOR, J. M. G. Treatment effects in a logistic model involving the box-cox transformation. **Journal of the American Statistical Association**, Taylor & Francis, v. 94, n. 445, p. 240–246, 1999. Cited on page 65.

STARE, J.; O'QUIGLEY, J. Fit and frailties in proportional hazards regression. **Biometrical Journal**, v. 46, p. 157–164, 2004. Cited on page 26.

SUZUKI, A. K.; CANCHO, V. G.; LOUZADA, F. The poisson–inverse-gaussian regression model with cure rate: a bayesian approach and its case influence diagnostics. **Statistical Papers**, Springer, v. 57, n. 1, p. 133–159, 2016. Cited on page 28.

TOMAZELLA, V. L. D. **Modelagem de Dados de Eventos Recorrentes via Processo de Poisson com Termo de Fragilidade**. 175 f. Phd Thesis (Doutorado em Ciências) — Instituto de Ciências Matemáticas e Computação, São Carlos, 2003. Cited on page 26.

TOURNOUD, M.; ECOCHARD, R. Application of the promotion time cure model with time-changing exposure to the study of HIV/AIDS and other infectious diseases. **Statistics in Medicine**, v. 26, p. 1008–1021, 2007. Cited on page 45.

VANEGAS, L.; PAULA, G. An extension of log-symmetric regression models: R codes and applications. **Journal of Statistical Simulation and Computation**, v. 86, p. 1709–1735, 2016. Cited on page 39.

_____. Log-symmetric distributions: statistical properties and parameter estimation. **Brazilian Journal of Probability and Statistics**, v. 30, p. 196–220, 2016. Cited on page 39.

VAUPEL, J. W.; MANTON, K. G.; STALLARD, E. The impact of heterogeneity in individual frailty on the dynamics of mortality. **Demography**, Springer, v. 16, n. 3, p. 439–454, 1979. Cited 5 times on pages 26, 27, 38, 43, and 44.

VILCA, F.; BALAKRISHNAN, N.; ZELLER, C. A robust extension of the bivariate Birnbaum-Saunders distribution and associated inference. **Journal of Multivariate Analysis**, v. 124, p. 418–435, 2014. Cited 2 times on pages 30 and 41.

_____. The bivariate sinh-elliptical distribution with applications to Birnbaum-Saunders distribution and associated regression and measurement error models. **Computational Statistics and Data Analysis**, v. 80, p. 1–16, 2014. Cited 2 times on pages 30 and 41.

WIENKE, A. **Frailty models in survival analysis**. [S.l.]: CRC Press, 2011. Cited 7 times on pages 26, 39, 44, 53, 68, 70, and 78.

WIENKE, A.; LOCATELLI, I.; YASHIN, A. I. The modelling of a cure fraction in bivariate time-to-event data. **Austrian Journal of Statistics**, v. 35, n. 1, p. 67–76, 2006. Cited on page 29.

YAKOVLEV, A. Y.; TSODIKOV, A. D. **Stochastic Models of Tumor Latency and their Biostatistical Applications**. New Jersey, US: World Scientific, 1996. Cited 2 times on pages 28 and 45.

YIN, G. Bayesian cure rate frailty models with application to a root canal therapy study. **Biometrics**, v. 61, p. 552–558, 2005. Cited on page 29.

YIQI, B.; CANCHO, V. G.; LOUZADA, F. On the bayesian estimation and influence diagnostics for the weibull-negative-binomial regression model with cure rate under latent failure causes. **Communications in Statistics-Theory and Methods**, Taylor & Francis, n. just-accepted, 2016. Cited on page 28.

YU, B. A frailty mixture cure model with application to hospital readmission data. **Biometrical Journal**, v. 50, p. 386–394, 2008. Cited 2 times on pages 26 and 29.

ZHOU, H.; HANSON, T.; JARA, A.; ZHANG, J. Modelling county level breast cancer survival data using a covariate-adjusted frailty proportional hazards model. **The annals of applied statistics**, NIH Public Access, v. 9, n. 1, p. 43, 2015. Cited on page 26.

APPENDIX

A

MATHEMATICAL RESULTS FOR BS FRAILTY MODEL

A.1 Score vector

The elements of the score vector are obtained from the first derivatives of the log-likelihood function given in (3.7) with respect to the two parameters. Throughout this appendix, for the sake of simplification, we define $\tau_{i,1} = \delta + 4\gamma t_i + 1$ and $\tau_{i,2} = \delta + \sqrt{(\delta+1)\tau_{i,1}} + 4\gamma t_i + 3$. Then, elements of the score vector are expressed as

$$\begin{aligned}\frac{\partial l(\boldsymbol{\xi})}{\partial \delta} &= -\sum_{i=1}^{n}\frac{\varsigma_i}{\tau_{i,1}} - \sum_{i=1}^{n}\frac{\varsigma_i\left(\frac{2\delta+4\gamma t_i+2}{2\sqrt{(\delta+1)\tau_{i,1}}}+1\right)}{\delta+\sqrt{(\delta+1)\tau_{i,1}}+1} - \sum_{i=1}^{n}\frac{1}{\tau_{i,1}} + \frac{\delta\sum_{i=1}^{n}\sqrt{\tau_{i,1}}}{4(\delta+1)^{3/2}} \\ &\quad - \frac{\delta\sum_{i=1}^{n}\frac{1}{2\sqrt{\tau_{i,1}}}}{2\sqrt{\delta+1}} + \sum_{i=1}^{n}\frac{\frac{1}{2\sqrt{\delta+1}}+\frac{1}{2\sqrt{\tau_{i,1}}}}{\sqrt{\delta+1}+\sqrt{\tau_{i,1}}} - \frac{\sum_{i=1}^{n}\sqrt{\tau_{i,1}}}{2\sqrt{\delta+1}} \\ &\quad + \sum_{i=1}^{n}\frac{\varsigma_i\left(\delta+\delta\left(\frac{2\delta+4\gamma t_i+2}{2\sqrt{(\delta+1)\tau_{i,1}}}+1\right)+\sqrt{(\delta+1)\tau_{i,1}}+4\gamma t_i+3\right)}{\delta\left(\delta+\sqrt{(\delta+1)\tau_{i,1}}+4\gamma t_i+3\right)+2},\end{aligned}$$

$$\begin{aligned}\frac{\partial l(\boldsymbol{\xi})}{\partial \gamma} &= -\sum_{i=1}^{n}\frac{4\varsigma_i t_i}{\tau_{i,1}} - \sum_{i=1}^{n}\frac{2(\delta+1)\varsigma_i t_i}{\sqrt{(\delta+1)\tau_{i,1}}\left(\delta+\sqrt{(\delta+1)\tau_{i,1}}+1\right)} \\ &\quad + \sum_{i=1}^{n}\frac{\varsigma_i\left(\delta\tau_{i,2}+\delta\gamma\left(\frac{2(\delta+1)t_i}{\sqrt{(\delta+1)\tau_{i,1}}}+4t_i\right)+2\right)}{\gamma(\delta\tau_{i,2}+2)} \\ &\quad - \frac{\delta\sum_{i=1}^{n}\frac{2t_i}{\sqrt{\tau_{i,1}}}}{2\sqrt{\delta+1}} - \sum_{i=1}^{n}\frac{4t_i}{\tau_{i,1}} + \sum_{i=1}^{n}\frac{2t_i}{\sqrt{\tau_{i,1}}\left(\sqrt{\delta+1}+\sqrt{\tau_{i,1}}\right)}.\end{aligned}$$

A.1.1 Observed information matrix

Let $T_1,\ldots,T_n$ be a random sample from the BS frailty model and $t_1,\ldots,t_n$ their observations. From the log-likelihood function given in (3.7), we have that the observed information

matrix of the BS frailty model is $\boldsymbol{I}(\boldsymbol{\xi}) = \begin{pmatrix} I_{\theta_1\xi_1} & I_{\xi_1\theta_2} \\ I_{\theta_2\xi_1} & I_{\xi_2\theta_2} \end{pmatrix}$, where $I_{\xi_i\xi_j} = -\partial^2 l(\boldsymbol{\xi})/\partial\theta_i\theta_j$, for $i,j = 1,2$, with $\xi_1 = \delta$ and $\xi_2 = \gamma$. Thus,

$$
\begin{aligned}
I_{\xi_1\xi_1} = & -\sum_{i=1}^{n} -\frac{\varsigma_i}{{\tau_{i,1}}^2} - \sum_{i=1}^{n} \varsigma_i \left(\frac{\frac{1}{\sqrt{(\delta+1)\tau_{i,1}}} - \frac{(2\delta+4\gamma t_i+2)^2}{4\left((\delta+1)\tau_{i,1}\right)^{3/2}}}{\delta+\sqrt{(\delta+1)\tau_{i,1}}+1} - \frac{\left(\frac{2\delta+4\gamma t_i+2}{2\sqrt{(\delta+1)\tau_{i,1}}}+1\right)^2}{\left(\delta+\sqrt{(\delta+1)\tau_{i,1}}+1\right)^2} \right) \\
& + \sum_{i=1}^{n} \varsigma_i \left(\frac{\frac{2\delta+4\gamma t_i+2}{\sqrt{(\delta+1)\tau_{i,1}}} + \delta\left(\frac{1}{\sqrt{(\delta+1)\tau_{i,1}}} - \frac{(2\delta+4\gamma t_i+2)^2}{4\left((\delta+1)\tau_{i,1}\right)^{3/2}}\right)+2}{\delta\left(\delta+\sqrt{(\delta+1)\tau_{i,1}}+4\gamma t_i+3\right)+2} \right. \\
& \left. - \frac{\left(\delta+\delta\left(\frac{2\delta+4\gamma t_i+2}{2\sqrt{(\delta+1)\tau_{i,1}}}+1\right)+\sqrt{(\delta+1)\tau_{i,1}}+4\gamma t_i+3\right)^2}{\left(\delta\left(\delta+\sqrt{(\delta+1)\tau_{i,1}}+4\gamma t_i+3\right)+2\right)^2} \right) \\
& - \frac{1}{2}\delta\left(-\frac{\sum_{i=1}^{n}\frac{1}{2\sqrt{\tau_{i,1}}}}{(\delta+1)^{3/2}} + \frac{3\sum_{i=1}^{n}\sqrt{\tau_{i,1}}}{4(\delta+1)^{5/2}} + \frac{\sum_{i=1}^{n} -\frac{1}{4{\tau_{i,1}}^{3/2}}}{\sqrt{\delta+1}} \right) - \frac{\sum_{i=1}^{n}\frac{1}{2\sqrt{\tau_{i,1}}}}{\sqrt{\delta+1}} \\
& + \frac{\sum_{i=1}^{n}\sqrt{\tau_{i,1}}}{2(\delta+1)^{3/2}} - \sum_{i=1}^{n} -\frac{1}{{\tau_{i,1}}^2} + \sum_{i=1}^{n} \left(\frac{-\frac{1}{4(\delta+1)^{3/2}} - \frac{1}{4{\tau_{i,1}}^{3/2}}}{\sqrt{\delta+1}+\sqrt{\tau_{i,1}}} - \frac{\left(\frac{1}{2\sqrt{\delta+1}}+\frac{1}{2\sqrt{\tau_{i,1}}}\right)^2}{\left(\sqrt{\delta+1}+\sqrt{\tau_{i,1}}\right)^2} \right),
\end{aligned}
$$

$$
\begin{aligned}
I_{\xi_1\xi_2} = & -\sum_{i=1}^{n} \left(\frac{\varsigma_i\left(\frac{2t_i}{\sqrt{(\delta+1)\tau_{i,1}}} - \frac{(\delta+1)t_i(2\delta+4\gamma t_i+2)}{\left((\delta+1)\tau_{i,1}\right)^{3/2}}\right)}{\delta+\sqrt{(\delta+1)\tau_{i,1}}+1} - \frac{2(\delta+1)\varsigma_i t_i\left(\frac{2\delta+4\gamma t_i+2}{2\sqrt{(\delta+1)\tau_{i,1}}}+1\right)}{\sqrt{(\delta+1)\tau_{i,1}}\left(\delta+\sqrt{(\delta+1)\tau_{i,1}}+1\right)^2} \right) \\
& + \sum_{i=1}^{n} \left(\frac{\varsigma_i\left(\frac{2(\delta+1)t_i}{\sqrt{(\delta+1)\tau_{i,1}}} + \delta\left(\frac{2t_i}{\sqrt{(\delta+1)\tau_{i,1}}} - \frac{(\delta+1)t_i(2\delta+4\gamma t_i+2)}{\left((\delta+1)\tau_{i,1}\right)^{3/2}}\right)+4t_i\right)}{\delta\left(\delta+\sqrt{(\delta+1)\tau_{i,1}}+4\gamma t_i+3\right)+2} \right. \\
& \left. - \frac{\delta\varsigma_i\left(\frac{2(\delta+1)t_i}{\sqrt{(\delta+1)\tau_{i,1}}}+4t_i\right)\left(\delta+\delta\left(\frac{2\delta+4\gamma t_i+2}{2\sqrt{(\delta+1)\tau_{i,1}}}+1\right)+\sqrt{(\delta+1)\tau_{i,1}}+4\gamma t_i+3\right)}{(\delta\tau_{i,2}+2)^2} \right) \\
& - \frac{\sum_{i=1}^{n}\frac{2t_i}{\sqrt{\tau_{i,1}}}}{2\sqrt{\delta+1}} + \frac{\delta\sum_{i=1}^{n}\frac{2t_i}{\sqrt{\tau_{i,1}}}}{4(\delta+1)^{3/2}} - \sum_{i=1}^{n} -\frac{4t_i}{{\tau_{i,1}}^2} + \sum_{i=1}^{n}\frac{4\varsigma_i t_i}{{\tau_{i,1}}^2} \\
& + \sum_{i=1}^{n} \left(-\frac{t_i}{{\tau_{i,1}}^{3/2}\left(\sqrt{\delta+1}+\sqrt{\tau_{i,1}}\right)} - \frac{2t_i\left(\frac{1}{2\sqrt{\delta+1}}+\frac{1}{2\sqrt{\tau_{i,1}}}\right)}{\sqrt{\tau_{i,1}}\left(\sqrt{\delta+1}+\sqrt{\tau_{i,1}}\right)^2} \right) - \frac{\delta\sum_{i=1}^{n} -\frac{t_i}{{\tau_{i,1}}^{3/2}}}{2\sqrt{\delta+1}},
\end{aligned}
$$

$$
\begin{aligned}
I_{\xi_2\xi_2} \quad = \quad & \sum_{i=1}^{n}\left(-\frac{\varsigma_i\left(\delta\left(\delta+\sqrt{(\delta+1)\tau_{i,1}}+4\gamma t_i+3\right)+\delta\gamma\left(\frac{2(\delta+1)t_i}{\sqrt{(\delta+1)\tau_{i,1}}}+4t_i\right)+2\right)}{\gamma^2\left(\delta\tau_{i,2}+2\right)}\right. \\
& +\frac{\varsigma_i\left(2\delta\left(\frac{2(\delta+1)t_i}{\sqrt{(\delta+1)\tau_{i,1}}}+4t_i\right)-\frac{4\delta(\delta+1)^2\gamma t_i^2}{\left((\delta+1)\tau_{i,1}\right)^{3/2}}\right)}{\gamma(\delta\tau_{i,2}+2)} \\
& \left.-\frac{\delta\varsigma_i\left(\frac{2(\delta+1)t_i}{\sqrt{(\delta+1)\tau_{i,1}}}+4t_i\right)\left(\delta\tau_{i,2}+\delta\gamma\left(\frac{2(\delta+1)t_i}{\sqrt{(\delta+1)\tau_{i,1}}}+4t_i\right)+2\right)}{\gamma\left(\delta\tau_{i,2}+2\right)^2}\right) \\
& -\sum_{i=1}^{n}-\frac{16\varsigma_i t_i^2}{{\tau_{i,1}}^2}-\sum_{i=1}^{n}\left(-\frac{4(\delta+1)^2\varsigma_i t_i^2}{\left((\delta+1)\tau_{i,1}\right)^{3/2}\left(\delta+\sqrt{(\delta+1)\tau_{i,1}}+1\right)}\right. \\
& \left.-\frac{4(\delta+1)\varsigma_i t_i^2}{\tau_{i,1}\left(\delta+\sqrt{(\delta+1)\tau_{i,1}}+1\right)^2}\right)-\sum_{i=1}^{n}-\frac{16t_i^2}{{\tau_{i,1}}^2} \\
& +\sum_{i=1}^{n}\left(-\frac{4t_i^2}{{\tau_{i,1}}^{3/2}\left(\sqrt{\delta+1}+\sqrt{\tau_{i,1}}\right)}-\frac{4t_i^2}{\tau_{i,1}\left(\sqrt{\delta+1}+\sqrt{\tau_{i,1}}\right)^2}\right)-\frac{\delta\sum_{i=1}^{n}-\frac{4t_i^2}{{\tau_{i,1}}^{3/2}}}{2\sqrt{\delta+1}}.
\end{aligned}
$$