Some Flexible Families of Mixture Cure Frailty Models and Associated Inference

SOME FLEXIBLE FAMILIES OF MIXTURE CURE FRAILTY MODELS AND ASSOCIATED INFERENCE

BY

MU HE, M.Sc.

A THESIS

SUBMITTED TO THE DEPARTMENT OF MATHEMATICS & STATISTICS

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Master of Science (2021) McMaster University
(Mathematics & Statistics) Hamilton, Ontario, Canada

TITLE:	Some Flexible Families of Mixture Cure Frailty Models and Associated Inference
AUTHOR:	Mu He M.Sc. Statistics, McMaster University B.Sc. Applied Mathematics, The Hong Kong Polytechnic University
SUPERVISOR:	Prof. N. Balakrishnan
NUMBER OF PAGES:	xii, 98

To my family

Abstract

In survival analysis or time-to-event analysis, one of the primary goals of analysis is to predict the occurrence of an event of interest for subjects within the study. Even though survival analysis methods were originally developed and used in medical research, those methods are also commonly used nowadays in other areas as well, such as in predicting the default of a loan and in estimating of the failure of a system.

To include covariates in the analysis, the most widely used models are the proportional hazard model developed by Cox (1972) and the accelerated failure time model developed by Buckley and James (1979). The proportional hazard (PH) model assumes subjects from different groups have their hazard functions proportionally, while the accelerated failure time (AFT) model assumes the effect of covariates is to accelerate or decelerate the occurrence of event of interest.

In some survival analyses, not all subjects in the study will experience the event. Such a group of individuals is referred to 'cured' group. To analyze a data set with a cured fraction, Boag (1948) and Berkson and Gage (1952) discussed a mixture cure model. Since then, the cure model and associated inferential methods have been widely studied in the literature. It has also been recognized that subjects in the study are often correlated within clusters or groups; for example, patients in a hospital would have similar conditions and environment. For this reason, Vaupel *et al.* (1979) proposed a frailty model to model the correlation among subjects within clusters and consequently the presence of heterogeneity in the data set. Hougaard (1989), McGilchrist and Aisbett (1991), and Klein (1992) all subsequently developed parametric frailty models. Balakrishnan and Peng (2006) proposed a Generalized Gamma frailty model, which includes many common frailty models, and discussed model fitting and model selection based on it.

To combine the key components and distinct features of the mixture cure model and the frailty model, a mixture cure frailty model is discussed here for modelling correlated survival data when not all the subjects under study would experience the occurrence of the event of interest. Longini and Halloran (1996) and Price and Manatunga (2001) developed several parametric survival models and employed the Likelihood Ratio Test (LRT) to perform a model discrimination among cure, frailty and mixture cure frailty models.

In this thesis, we first describe the components of a mixture cure frailty model, wherein the flexibility of the frailty distributions and lifetime survival functions are discussed. Both proportional hazard and accelerated failure time models are considered for the distribution of lifetimes of susceptible (or non-cured) individuals. Correlated random effect is modelled by using a Generalized Gamma frailty term, and an EM-like algorithm is developed for the estimation of model parameters. Some Monte Carlo simulation studies and real-life data sets are used to illustrate the models as well as the associated inferential methods.

KEY WORDS: Frailty model; Mixture cure model; Mixture cure frailty model; Proportional hazard model; Accelerated failure time model; Gamma distribution; Lognormal distribution; Weibull distribution; Generalized Gamma distribution; Censored data; Clustered data; Monte Carlo simulation; Model Discrimination; Maximum likelihood estimation; EM algorithm; Likelihood ratio test; Akaike information criterion; Bayesian information criterion; Bias; Mean square errors; Fisher information; Coverage probability.

Acknowledgements

I want to take this opportunity to thank my doctoral supervisor Professor Narayanaswamy Balakrishnan. He introduced me to the Ph.D. program and provided me the opportunity to pursue an academic career. Throughout my thesis-writing period, he guided me with great research ideas, consistent encouragement, care and unlimited patience. He is more than an academical mentor to me and it was a wonderful experience to work under his supervision.

I want to thank my doctoral supervisory committee members, Professor Ben Bolker and Professor Roman Viveros-Aguilera, for their useful suggestions, criticisms and support. I appreciate their help, advice and time gratefully. I also want to thank Professor Yingwei Peng and Professor Zhisheng Ye for their valuable suggestions and advice to my research work.

I am grateful to my parents for their unconditionally support and love. My special thanks to Yuexue Zhang, who walks into my life in my lowest time and is always accompany me with tolerance, understanding and love.

I would also wish to thank my friends Michael Li, Wenxing Guo, Kai Liu, Xiaojun Zhu, Yiming Liu, Sandip Barui and Suvra Pal for their discussion, help and support in the computational work. I want to thank my friends Katherine Davies, Yanqiao

Wang, Tian Feng, Chengwei Qin, Dejing Kong and all other students of my supervisor for their kind assistance and support.

Last, but not the least, I wish to express my appreciation to Department of Mathematics and Statistics for providing me the financial support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Basic Concepts in Survival Analysis

Time-to-event is of interest in many different fields; for example, whether a patient dies from a disease or gets cured due to a treatment and whether an operating system will fail. In general, statistical techniques for analyzing time-to-event data are referred to as survival analyses. A time-to-event variable in health applications correspond to the time until a participant in the clinical study has an event of interest, such as heart attack, cancer remission, death and so on.

The most direct way of analyzing the event time is to consider the lifetime as continuous response and use methods such as linear regression and logistic regression. However, it is common to have some lifetime data to be missing. For example, in many medical experiment studies, the clinical trials terminate before the event occurs for some individuals or some patients drop off from a study and researchers cannot follow up. An individual with unknown status of occurrence is said to be a censored individual, and the corresponding time is called censored time. It is not the event occurrence time, and the only information we know is that the event occurrence time is greater than the censored time. There are many kinds of censoring discussed in survival analysis; for example, in Type I censoring, the study is designed to end after n years, but censored subjects do not all have the same censoring time due to reason specified before, while in Type II censoring, a study ends when there is a pre-specified number of occurrence events. Regardless of the type of censoring, we assume here that the censoring is non-informative about the event. For such data, we develop a survival model to fit. For this purpose, suppose T denotes the response variable and $T > 0$. Then, the survival function is given by

$$S(t) = Pr(T > t) = 1 - F(t), \quad t > 0. \tag{1.1}$$

The survival function gives the probability that a subject will survive past time t. The hazard function $h(t)$ is the instantaneous failure rate given by

$$h(t) = \lim_{\Delta t \to 0} \frac{Pr(t < T \leq t + \Delta t)}{\Delta t} = \frac{f(t)}{S(t)}, \quad t > 0. \tag{1.2}$$

and the cumulative hazard, describing the accumulated risk up to time t is given by

$$H(t) = \int_0^t h(s)ds. \tag{1.3}$$

The relationships between $S(t)$, $H(t)$ and $h(t)$ are as follows:

$$\begin{aligned} h(t) &= -\frac{d\log\big(S(t)\big)}{dt}, \\ H(t) &= -\log(S(t)), \\ S(t) &= \exp(-H(t)). \end{aligned} \tag{1.4}$$

To model survival data with censored times properly, we denote W_i as the response for the ith subject, C_i as the censoring time for the ith subject and δ_i as the event censoring indicator given by

$$\delta_i = \begin{cases} 1, & W_i \leq C_i \\ 0, & W_i > C_i. \end{cases} \tag{1.5}$$

Then, the observed response is simply $T_i = \min(T_i, C_i)$.

To estimate the survival function, we could consider three different ways by assuming a parametric model, non-parametric model or semi-parametric model.

1.1.1 Parametric Survival Function

The most direct way to estimate the survival function is to assume a parametric form, such as Exponential, Weibull, Gamma and Log-normal distributions. In theory, the survival function is smooth based on the parametric setting. In practice, we may observe events on a discrete time scale, such as days, weeks, etc, in which case some discrete parametric choices may become necessary.

1.1.2 Non-parametric Survival Function

Instead of the parametric model which assumes the functional form to be known, we can consider the true survival distribution to be unknown. In this case, it will be preferable to model the data with some non-parametric method. The most popular one is the Kaplan-Meier estimator proposed by Kaplan and Meier (1958). Let $t_1 < t_2 < \cdots < t_k$ denote the unique event times, d_i denote the number of failures at time t_i, and r_i be the number of patients at risk just before time t_i. Then, the Kaplan-Meier estimator of the survival function $S(t)$ is given by

$$\hat{S}(t) = \prod_{t_i < t} \frac{r_i - d_i}{r_i}. \tag{1.6}$$

The corresponding cumulative hazard function is

$$\hat{H}(t) = -\log\big(\hat{S}(t)\big). \tag{1.7}$$

The variance of the Kaplan-Meier estimator is estimated by Greenwood's formula:

$$\hat{\sigma}^2(t) = \hat{S}(t)^2 \sum_{t_i < t} \frac{d_i}{r_i(r_i - d_i)}. \tag{1.8}$$

When there is no censoring, this formula reduces to the standard binomial variance estimator.

As no distributional assumptions are made, one important use of the estimator in (1.6) is to check graphically the fit of parametric models. Instead, Nelson (1969) and Nelson (1972) introduced a non-parametric estimator of the cumulative hazard function. The so-called Nelson-Aalen estimator is a non-parametric estimator that can be used to estimate the cumulative hazard function for censored survival data. It is given by

$$\hat{H}(t) = \prod_{t_i < t} \frac{d_i}{r_i}. \tag{1.9}$$

The variance of the Nelson-Aalen estimator can be estimated as

$$\hat{\sigma}^2(t) = \sum_{t_i < t} \frac{d_i(r_i - d_i)}{r_i^2(r_i - 1)}. \tag{1.10}$$

Breslow (1972) has described more the use of Nelson-Aalen estimator in survival analysis in great detail. A treatment of these two estimators from the point of view of counting processes can be found in Andersen *et al.* (1993).

1.1.3 Semi-parametric Survival Function

The most popular semi-parametric model is the Cox proportional hazard model (PH). The baseline survival (or hazard) function is not specified in a Cox model. The Cox PH model, developed by Cox (1972) and Cox and Oakes (1984), is usually written in terms of the hazard function as

$$h(t) = h_0(t) \exp(\beta' X), \tag{1.11}$$

where β is the regression parameter vectors for the covariates X. This model gives an expression for the hazard at time t for an individual with a given set of explanatory variables X. The hazard at time t is the product of an unspecified baseline hazard function $h_0(t)$ and a parametric explanatory variable part $\exp(\beta' X)$. Such a model is called a semi-parametric model because even if the regression parameters are known, the distribution of the outcome remains unknown. There are many other examples of semi-parametric models in the literature; these are discussed briefly in the next section.

1.2 Literature Review

1.2.1 Mixture Cure Model

In survival analysis, it is common to assume that all individuals will eventually experience the event of interest as long as the follow-up period is long enough. However, in some cancer clinical studies, a substantial proportion of subjects may get cured by the treatment they undergo and never experience the event of interest. The proportion of the cured patients is of primary interest while analyzing such data, but it cannot be estimated with a usual survival model, thus resulting the need for suitable models for this purpose. To model data with a proportion of cured patients, cure rate model was introduced by Boag (1948) and Berkson and Gage (1952). Let T denote the failure time and $S_{pop}(t)$ be the survival function of T. Then, the mixture cure model can be expressed as

$$S_{pop}(t) = p_0 + (1 - p_0)S_s(t), \tag{1.12}$$

where p_0 is cure rate and $S_s(t)$ is the survival function of the susceptible patients. This model can be fitted by using the maximum likelihood method. A detailed discussion of this approach can be found in Maller and Zhou (1996). Laska and Meisner (1993) proposed a non-parametric generalized maximum likelihood estimation method for the mixture cure model.

Equation (1.12) can also be expanded to include covariate effects. Let x and z denote some covariates that may affect the cure rate and the latency distribution. Such model can then be rewritten as

$$S_{pop}(t|x, z) = p_0(z) + \big(1 - p_0(z)\big)S_s(t|x), \tag{1.13}$$

where $p_0(z)$ is the probability of a patient being cured depending on z, and $S_s(t|x)$ is the survival function of the failure time distribution of susceptible patients depending on x.

Several authors have considered a parametric approach to Equation (1.13) by assuming a parametric distribution for the latency distribution and effects of x only on the scale of the latency distribution, including Farewell (1982), Yamaguchi (1992) and Peng *et al.* (1998). Semiparametric approaches to the model are also available to reduce the dependence of the model on a parametric assumption.

The most popular one is the semi-parametric Cox proportional hazards (PH) mixture cure model given by

$$S_{pop}(t|x, z) = p_0(z) + \big(1 - p_0(z)\big)\big(S_0(t)\big)^{\exp(\beta x)}, \tag{1.14}$$

where $S_0(t)$ is an unspecified baseline survival function. Several estimation methods have been discussed in the literature; for example, Kuk and Chen (1992) proposed a marginal likelihood method, using a Monte Carlo approximation for estimating the semi-parametric model. Peng and Dear (2000) and Sy and Taylor (2000) later discussed the proportional hazards cure model using a semi-parametric EM algorithm.

Other generalized mixture cure models have also been discussed in the literature. Lu and Ying (2004) proposed a general class of transformation cure models in which the linear transformation model is used for failure times of susceptible subjects. Besides the mixture model, Fine and Gray (1999) introduced the idea of competing risk to the cure fraction. Balakrishnan and Pal (2014), Balakrishnan and Pal (2016), Pal and Balakrishnan (2017) and Balakrishnan *et al.* (2017) summarized parametric competing risk models using the COM-Poisson distribution.

Another popular model associated with the cure model in survival analysis is the accelerated failure time (AFT) mixture model given by

$$\begin{aligned} \log(t) &= \exp(\beta' x) + \epsilon, \\ S_{pop}(\epsilon | x, z) &= p_0(z) + \big(1 - p_0(z)\big) S(\epsilon). \end{aligned} \tag{1.15}$$

In standard survival data analysis when there is no cure fraction, the accelerated failure time model, introduced by Kalbfleisch and Prentice (1973) and Cox and Oakes (1984), is another useful alternative to the proportional hazards model. The accelerated failure time model has a direct physical interpretation (Reid (1994)) and has been widely discussed in the literature. Different from the proportional hazard mixture cure model, the accelerated failure time mixture cure model utilizes the accelerated failure time model for the latency component. There are few studies of the semi-parametric accelerated failure time mixture cure model in the literature due to the complexity of the associated estimation. Li and Taylor (2002) considered the M-estimator for estimating the unknown parameters in the semi-parametric accelerated failure time mixture cure model. However, results from their method may depend on the form of the M-estimator and the resulting method of finding the estimates is computationally quite intensive. Zhang and Peng (2007b) developed a modified Gehan-type weighted log-rank estimation for parameter estimation later to improve the estimation method. However, the theoretical properties of these two estimates have not been studied yet. As these methods do not maximize the observed likelihood function, the estimators are not efficient and the bootstrap methods are used to obtain the estimated variance. To compensate for the lack of efficiency, Zeng and Lin (2007) incorporated a kernel estimation method and obtained an efficient estimator for the AFT model. Based on this work, Lu (2010) proposed a kernel-based non-parametric maximum likelihood estimation method for the accelerated failure time mixture cure model. Some more recent works have discussed the competing risk scenario, such as Choi *et al.* (2018), who has considered a bivariate competing risk for the mixture cure accelerated failure time model.

1.2.2 Frailty Model

The frailty model is another popular model used while analyzing clustered failure time data, wherein the frailty term is used to capture an association within each cluster. Due to the involvement of a random effect term in the hazard function, the frailty model is also known as the random effect model. The frailty model was first discussed by Vaupel *et al.* (1979), who modeled the correlation of clusters through a random effect PH model. Since then, Hougaard (1989), McGilchrist and Aisbett (1991) and Klein (1992) have modelled the frailty based on positive stable, Log-normal and Gamma distributions respectively. Balakrishnan and Peng (2006) discussed in detail the estimation method for the Generalized Gamma frailty model based on Monte

Carlo approach. In general, the frailty model can be written as

$$S_{pop}(t|x) = \exp\big(-yH_0(t)\big) = L_y\big(H_0(t)\big), \tag{1.16}$$

where $L_y(.)$ is the Laplace transform of y.
However, the frailty model based on the semi-parametric accelerated failure time model has attracted less attention than the one based on the proportional hazards model due to its computational difficulties. The model can be written as

$$\begin{aligned} \log(t) &= \exp(\beta x) + \epsilon, \\ S_{pop}(\epsilon|x,z) &= L_y(H_0\big(\epsilon)\big). \end{aligned} \tag{1.17}$$

The marginal approach is one of the most common methods for analyzing clustered data in the AFT model, as in Jin *et al.* (2006a) and Jin *et al.* (2006b). However, most non-parametric or semi-parametric models make use of EM algorithm due to the unspecified hazard function. Pan (2001) proposed the AFT frailty model by using a Gamma frailty on the error term and developed an EM-like algorithm for estimation of parameters in the AFT frailty model. This estimation procedure was improved by Zhang and Peng (2007a) through the M-estimator, and Xu and Zhang (2010) using rank estimation methods. However, none of these estimation methods are efficient. Liu *et al.* (2013) developed a non-parametric efficient maximum likelihood estimation method for the AFT frailty model, which is more efficient than the Gehan-type rank estimator in most cases.

1.2.3 Mixture Cure Rate Frailty Model

Little attention has been paid to the combination of cure model and frailty model due to its complex structure. The cure frailty model combines the advantages of the cure model and of the frailty model to deal with dependent curable survival data. Most of the work on cure frailty model uses a two-component mixture of the cured and uncured populations, commonly called a mixture cure frailty model, given by

$$S_{pop}(t|x,z) = p_0(z) + \big(1 - p_0(z)\big)S_s(t|x), \tag{1.18}$$

where the term $S_s(t|x)$ is the survival function for the susceptible group from the usual frailty model.

The most basic cure frailty model assumes that the cure proportion and the frailty are independent and using a parametric distribution for the frailty part, such as the Gamma frailty model in Longini and Halloran (1996).

The cure frailty models are also applied to survival data among correlated individuals. Chatterjee and Shih (2001) proposed cure frailty models with shared frailty and assuming the cure part and frailty part to be dependent. Wienke *et al.* (2003) and Wienke *et al.* (2006) later extended the independent model using the correlated Gamma frailty and correlated Log-normal frailty models to enrich the family.

Other than the proposal of different models, general estimation procedures have also been discussed by Chatterjee and Shih (2001), Wienke *et al.* (2003) and Wienke *et al.* (2006). To overcome the computational complexity in the cure frailty model, Peng and Zhang (2008a) considered each susceptible individual as an individual cluster and introduced a combination of logistic regression estimation and Gehan-type weighted log-rank estimation for the cure frailty model.

In addition to the marginal likelihood methods and EM-like algorithm in the literature, Bayesian approach has also been implemented by some researchers, including Yin (2005) and Diao and Yin (2012).

Peng and Zhang (2008b) discussed conditions under which the cure frailty model is identifiable, and also showed that the model is identifiable when constructed using the mixture cure model and containing covariates in both the cure fraction and frailty distribution.

Instead of the usual mixture cure model without frailty term, AFT model can be considered as well. The mixture cure frailty AFT model has not been discussed due to its complex structure and computational difficulties. In this thesis, we discuss the mixture cure frailty AFT model and the associated estimation. The model setting for the AFT part follows Pan (2001)'s work, who developed the EM algorithm based on the error term. The mixture cure frailty AFT model can be written as

$$\begin{aligned} &\log(t) = \exp(\beta' x) + \epsilon, \\ &h(\epsilon | Y_{ij} = y_{ij}) = y_{ij} h_0(\epsilon), \\ &S_{pop}(\epsilon | x, z) = p_0(z) + \big(1 - p_0(z)\big) S(\epsilon). \end{aligned} \tag{1.19}$$

It should be mentioned that the frailty term y_i or y_{ij} depends on whether it is a shared frailty or an individual random effect. Given Y, the conditional hazard function $h(.)$ follows the PH frailty model.

Note that we have standard mixture cure AFT model with $y_{ij} \equiv 1$. The estimation can be done directly through maximum likelihood if the baseline hazard function is specified. However, the baseline hazard function is usually unknown and a EM-like algorithm can then be applied.

1.3 Frailty Distribution

For the frailty term in the frailty model, we usually assume some common parametric distribution such as Gamma distribution, Log-normal distribution, Generalized Gamma distribution, and so on. We present details of Gamma distribution and Generalized Gamma distribution which are made use of in the latter chapters.

1.3.1 Gamma Distribution

Suppose $f(y)$ corresponds to a Gamma distribution with parameter p. Then,

$$f(y|p) = \frac{p^p y^{p-1} e^{-py}}{\Gamma(p)}, \tag{1.20}$$

where $\mathbb{E}(y) = 1$, which is required for the purpose of identifiability.

1.3.2 Generalized Gamma Distribution

Suppose $g(y)$ corresponds to a Generalized Gamma distribution. Then,

$$g(y|q,\sigma,\lambda) = \begin{cases} |q|(q^{-2})^{q^{-2}}(\lambda y)^{q^{-2}(q/\sigma)} \exp[-q^{-2}(\lambda y)^{q/\sigma}]/[\Gamma(q^{-2})\sigma y], & q \neq 0, \\ (\sqrt{2\pi}\sigma y)^{-1} \exp\{-[\log(\lambda y)]^2/(2\sigma^2)\}, & q = 0, \end{cases} \tag{1.21}$$

where -$\infty < q < \infty$ and $\sigma > 0$ are shape parameters and $\lambda > 0$ is the scale parameter. The mean of the Generalized Gamma distribution in (1.21) is

$$\frac{\Gamma(q^{-2} + \frac{\sigma}{q})}{\Gamma(q^{-2})(q^{-2})^{\sigma/q}\lambda} \tag{1.22}$$

and exists only when $q > -\frac{1}{\sigma}$; when the mean equals one, we have

$$\lambda = \frac{\Gamma(q^{-2} + \frac{\sigma}{q})}{\Gamma(q^{-2})(q^{-2})^{\sigma/q}}. \tag{1.23}$$

Then, the corresponding the variance is

$$\frac{\Gamma(q^{-2} + 2\frac{\sigma}{q})\Gamma(q^{-2})}{\Gamma^2(q^{-2} + \frac{\sigma}{q})} - 1. \tag{1.24}$$

The Generalized Gamma distribution includes many well known distributions as special cases. For example, it reduces to the Weibull distribution when $q = 1$, the Log-normal distribution when $q = 0$, the Gamma distribution when $q/\sigma = 1$, and the positive stable distribution with index $\frac{1}{2}$ when $q^{-2} = \frac{1}{2}$ and $\sigma/q = -1$. It possesses considerable flexibility to capture the characteristics in a distribution that might have been missed by using these particular frailty distributions.

The Generalized Gamma frailty model can be represented as a transformation to the Gamma frailty model. From the work of Balakrishnan and Peng (2006), if Z denotes the frailty in the Gamma frailty model with p.d.f.

$$f(z|p) = \frac{p^p z^{p-1} e^{-pz}}{\Gamma(p)}, \tag{1.25}$$

then

$$Y = \frac{1}{\lambda} Z^{\sigma\sqrt{p}}, \tag{1.26}$$

where Y is the Generalized Gamma frailty model with $q = \frac{1}{\sqrt{p}}$ and λ in (1.23).
Thus, the frailty term in Equation (1.25) can be written as a random effect term as

$$h(t_{ij}|y_i) = h_0(t_{ij}) \exp(\beta' x_{ij} + w_i), \tag{1.27}$$

where

$$w_i = \log y_i \tag{1.28}$$

can be written as

$$w_i = -\log(\lambda) + \frac{\sigma}{q} \log z_i, \tag{1.29}$$

with the z_i being a realisation of Z. Therefore, $\frac{\sigma}{q}$ can be considered as a coefficient of $\log z_i$ and varies independently in the Generalized Gamma frailty model.
The limiting property and the estimation methods of the Generalized Gamma distribution as a frailty distribution has also been discussed in the literature; See also Prentice (1977), Lawless (1980), Johnson *et al.* (1994) and Balakrishnan and Peng (2006).
For simplification, we can also use the alternative form of the Generalized Gamma distribution, given by Stacy (1962) as follows:

$$f(y_i|a, d, p) = \frac{\frac{p}{a^d} y_i^{d-1} \exp(-(\frac{y_i}{a})^p)}{\Gamma(\frac{d}{p})}, \tag{1.30}$$

and the moment generating function of Generalized Gamma is given by

$$M_{y_i}(u) = \sum_{r=0}^{\infty} \frac{(ua)^r}{r!} \left\{ \frac{\Gamma(\frac{d+r}{p})}{\Gamma(\frac{d}{p})} \right\}, \tag{1.31}$$

where the link with the original form is that with $E(y_i) = 1$:

$$\begin{aligned} d &= (\frac{q^{-2}q}{\sigma}) = \frac{1}{q\sigma} < 1, \\ p &= \frac{q}{\sigma}, \\ a &= \frac{(q^2)^{\sigma/q}}{\lambda} = \frac{\Gamma(q^{-2})}{\Gamma(q^{-2} + \sigma/q)}. \end{aligned} \tag{1.32}$$

1.4 Bone marrow transplant data

Kersey *et al.* (1987) first introduced the bone marrow transplant data from a clinical trial study. This study was designed to compare the treatment effect of autologous (AL, treatment 0) and allogeneic (AG, treatment 1) marrow transplantation for the disease acute lymphoblastic leukemia. There were 91 participants in the study and followed up to 1845 days. 46 of the participants received allogeneic marrow transplant while the others were in the autologous treatment group. The time to a recurrence of leukemia or the censoring time was subsequently recorded in days. The Kaplan-Meier curves for these data are presented in Figure 1.1.

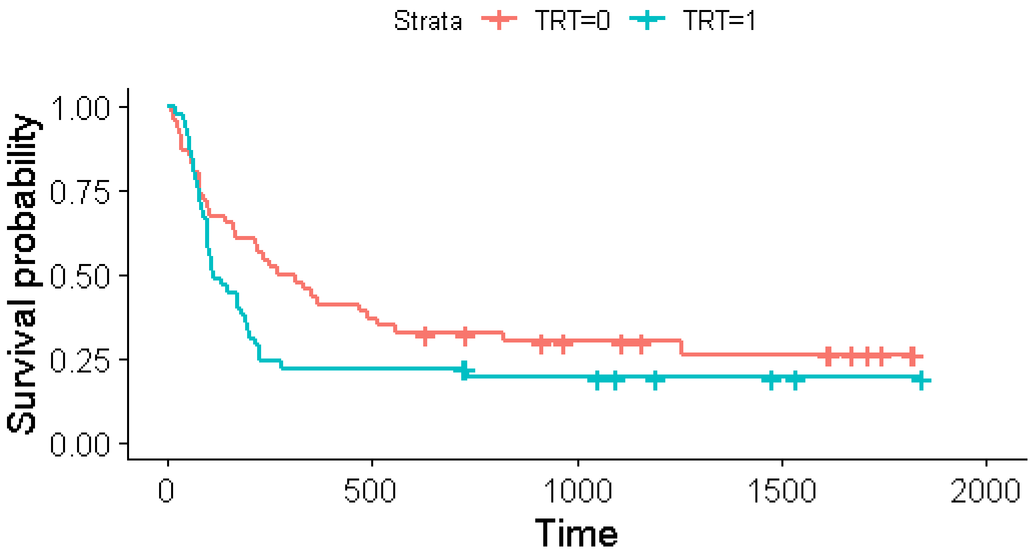

Figure 1.1: Bone marrow transplant data: Kaplan-Meier estimates of the survival curves for the autologous and allogeneic groups

There were correspondingly 28.26% and 20% patients who were censored in the allogeneic and autologous treatment groups. Meanwhile, the survival curves show limiting survival probabilities of 26.34% and 19.44% for allogeneic and autologous groups respectively. The leveling-off in Figure 1.1 might be caused by long-term censored times, which could potentially be either cured patients or patients dropping-off from the study for some other reasons.

This data set has been extensively studied in the literature. Maller and Zhou (1996) analyzed these data with exponential mixture cure model. Based on their results, it is known that the exponential mixture cure model fits the AG group satisfactorily, but performed poorly in fitting the AL group. They also provided evidence of existence of the cure fractions in the AG group using the likelihood-ratio test (LRT). Peng *et al.*

(2001) fitted the data with Weibull, Gamma and Log-normal mixture cure models, and also tested the goodness-of-fit of each model through LRT by using the more general distribution, namely the Generalized Gamma distribution. The Generalized Gamma distribution contains Weibull, Gamma and Log-normal distributions all as special cases, as discussed in the last section. They observed that the Log-normal fitted the AL group better. Price and Manatunga (2001) extended the mixture cure model to mixture cure frailty model to incorporate heterogeneity between patients, possibly caused by other biological phenomena. Focusing their study on the AL group, they compared the homogeneous model, mixture cure model, Gamma frailty model, inverse Gaussian frailty model, mixture cure Gamma frailty model, mixture cure inverse Gaussian frailty model and mixture cure compound Poisson frailty model, with baseline distribution being Weibull. They concluded that the mixture cure Gamma frailty model provides a better fit for these data. They also proved the existence of heterogeneity using LRT. Further, Zhang and Peng (2007b) fitted the data set with a semi-parametric mixture cure AFT model with a new estimation method. Peng and Zhang (2008a) fitted a semi-parametric mixture cure Gamma frailty model using EM algorithm and multiple imputation estimation methods; their results are consistent with those of Price and Manatunga (2001). However, due to the inefficiency in the estimation method, such a semi-parametric model is usually not suitable in model discrimination analysis.

1.5 Likelihood Inference

Let $O = \{t_{ij}, \delta_{ij}, x_{ij}, i = 1, 2, \ldots, n\}$ denote the observed data, where t_{ij} is the observed time, δ_{ij} is the censoring indicator, and x_{ij} are the covariates. Under the assumption of the cure rate, the cure fraction can be written as

$$p_0(b) = \frac{1}{1 + \exp(b'X)}, \tag{1.33}$$

where $b = (b_0, b_1, ..., b_p)$ is a vector of unknown parameters. We also define β as a vector of unknown parameters in the Cox PH setting in the frailty model, and thus $\Theta = \{b, \beta, H_0(t)\}$. In the mixture cure frailty model, apart from the frailty latent variable y_{ij}, we define the censoring indicator I_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, \ell_i$, as

$$I_{ij} = \begin{cases} 0 & \text{if the subject is cured,} \\ 1 & \text{if the subject is not cured.} \end{cases} \tag{1.34}$$

We develop a EM-type algorithm since usually the EM Algorithm's base is a model with discrete latent value.

At each iteration in the M step, we introduce $\{Q(I_{ij}, y_{ij}) : \sum Q(I_{ij}, y_{ij}) = 1, Q(I_{ij}, y_{ij}) >$

0} to satisfy Jensen's inequality. Then, the corresponding log-likelihood is derived as

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{\ell_i}\log(p(O|\Theta)) &= \sum_{i=1}^{n}\sum_{j=1}^{\ell_i}\log\left(\sum_{y_{ij}}\left[\sum_{I=\{0,1\}} p(O,I_{ij},y_{ij}|\Theta)\right]\right)\\ &= \sum_{i=1}^{n}\sum_{j=1}^{\ell_i}\log\left(\sum_{y_{ij}}\left[\sum_{I=\{0,1\}} Q(I_{ij},y_{ij})\frac{p(O,I_{ij},y_{ij}|\Theta)}{Q(I_{ij},y_{ij})}\right]\right)\\ &= \sum_{i=1}^{n}\sum_{j=1}^{\ell_i}\log\left(\mathbb{E}_{I_{ij},y_{ij}\sim Q}\left(\frac{p(O,I_{ij},y_{ij}|\Theta)}{Q(I_{ij},y_{ij})}\right)\right)\\ &\geq \sum_{i=1}^{n}\sum_{j=1}^{\ell_i}\mathbb{E}_{I_{ij},y_{ij}\sim Q}\left(\log\left(\frac{p(O,I_{ij},y_{ij}|\Theta)}{Q(I_{ij},y_{ij})}\right)\right)\\ &= \sum_{i=1}^{n}\sum_{j=1}^{\ell_i}\int_0^\infty\left[\sum_{I=\{0,1\}} Q(I_{ij},y_{ij})\log\left(\frac{p(O,I_{ij},y_{ij}|\Theta)}{Q(I_{ij},y_{ij})}\right)\right]dy_{ij}\end{aligned}\tag{1.35}$$

To further satisfy the lower-bound tightly, we have

$$\begin{aligned}Q(I_{ij},y_{ij}) &= \frac{p(O,I_{ij},y_{ij}|\Theta)}{\sum_{I_{ij},y_{ij}} p(O,I_{ij},y_{ij}|\Theta)}\\ &= p(I_{ij},y_{ij}|O,\Theta).\end{aligned}\tag{1.36}$$

In the above, the y_{ij} are assumed to follow a specific distribution. Then, after using the expected values in the E-step, we repeat the M-step until convergence to the desired level of accuracy.

1.6 Simulation Study

To assess the performance of the cure frailty model and the associated estimation method, a Monte Carlo simulation study is performed. For comparative purpose, the Gamma and Log-normal frailty models are fitted to simulated data and the baseline hazard function is taken as the Weibull distribution.

The performance of the proposed mixture cure frailty model and mixture cure frailty model with AFT hazard, along with the precision of the estimates, are evaluated through simulation. Different settings in the simulation study are used for the purpose of investigating the effects of sample size, cure rate, and censoring proportion. The results from the simulation study, including parameter estimates, standard errors, bias, mean square errors, root mean square errors and 95% coverage probabilities, for the parameters are all determined. The asymptotic normality property of the MLEs can be used to construct confidence intervals for the model parameters of interest.

Different set-ups have been considered in each simulation study. The number of replications depend on the computational intensity and the complexity of the corresponding models. For each set-up, suppose the estimate of β is $\hat{\beta}_i$ based on the i-th simulated sample, $i = 1, ...n$. Then, the parameter estimates, standard errors, mean square errors are determined as follows:

$$m(\hat{\beta}) = \frac{\sum \hat{\beta}_i}{n} \quad sd(\hat{\beta}) = \sqrt{\frac{\sum \left(\hat{\beta}_i - m(\hat{\beta})\right)^2}{n-1}} \quad MSE(\hat{\beta}) = \frac{\sum(\hat{\beta}_i - \beta)^2}{n} \tag{1.37}$$

We also considered the coverage probability using the following three steps:

1. Simulate n samples of size n_s from the population;
2. Compute the 95% confidence interval based each sample;
3. Compute the proportion of samples for which the (known) population parameter is contained within the confidence interval. Then, that proportion is an estimate of the empirical coverage probability for the CI.

However, mostly in this thesis, the calculation of standard errors of the parameter estimates is done based on bootstrap, then the coverage probability is not easy to determine based on the simulated data sets due to the computational intensity involved. Therefore, we only evaluate the bootstrap variances under a few settings in the simulation study due to the limit in computational time.

For some of the simulations involving MCMC approximation, the convergence is not achieved sometimes, and in such cases we stop at the 30th iteration as suggested by Cai *et al.* (2012).

Lastly, for the sake of computational efficiency, the Shared Hierarchical Academic Research Computing Network (SHARC net) was made use of in for all the simulation work in this thesis.

1.7 Model Discrimination

Model discrimination is used to assess the relative performance between Generalized Gamma mixture frailty AFT model in Chapter 4, as we apply a kernel-smoothed profile likelihood method. The method is similar to the partial likelihood method, but the estimates can be obtained by maximization through Newton-Raphson algorithm. However, as mentioned before, the estimates considered in Chapter 2 and Chapter 3 are not efficient, and so the model discrimination based on them is not suitable.

Samples were generated from several true models and fitted with some candidate models. The fitted results can be compared by information-based criteria such as Akaike information criterion (AIC) and Bayesian information criterion (BIC), which are given by

$$AIC = -2\hat{\ell} + 2p, \quad BIC = -2\hat{\ell} + p\log(n), \tag{1.38}$$

where $\hat{\ell}$ is the log-likelihood value computed at the MLEs, p is the number of parameters and n is the sample size. Models with the smallest AIC or BIC are chosen for each sample and the percentages of each candidate model selected can then be calculated.

The Likelihood Ratio Test (LRT) is useful in the case of nested models, and so it could be used in the mixture cure Generalized Gamma frailty model as it contains the mixture Gamma cure frailty model as a special case.

The LRT statistic is given by

$$\Lambda = -2(\hat{\ell}_0 - \hat{\ell}), \tag{1.39}$$

where $\hat{\ell}_0$ and $\hat{\ell}$ are the maximized log-likelihood values for the reduced model and the full model, respectively.

Under some suitable regularity conditions, the asymptotic distribution of the LRT statistic follows a χ^2 distribution under the null hypothesis with degrees of freedom $df_{full} - df_{reduced}$, where df_{full} and $df_{reduced}$ are the numbers of parameters in the reduced model and the full model, respectively.

Furthermore, the boundary condition of LRT has been discussed by Maller and Zhou (1996). They proposed that the large sample distribution of $-2\log(\hat{\ell}_0 - \hat{\ell})$ is a 50 - 50 mixture of a chi-square random variable with 1 degree of freedom and a point mass at 0. This boundary condition can be applied when testing the existence of cure proportion as H_0.

1.8 Scope of the thesis

In Chapter 2, we study the mixture cure frailty model with Generalized Gamma distribution for the frailty term. We employ a Breslow-type estimator for the baseline hazard function $H_0(t)$, motivated by the work of Breslow (1972), Peng (2003) and Peng and Zhang (2008a). We then fit a mixture cure frailty model with Gamma, Log-normal and Weibull distributions for comparison, as these are special cases in the Generalized Gamma distribution. A simulation and real-life data study are then performed to examine the estimation methods using EM-type algorithm. In Chapter 3, a mixture cure AFT frailty model is considered. The survival function follows a proportional hazard frailty distribution as discussed in the preceding sections, while the frailty term is assumed to have a Gamma distribution. The EM algorithm is implemented due to the involvement of the latency variables I and Y. In the M-step, a Gehan-type weighed function is introduced and made use of in maximizing the Q function. As the baseline distribution is not specified, we employ a Breslow-type estimator and update the baseline distribution in each iteration. Both simulation and real-life data studies show that the developed estimates are accurate. In Chapter 4, to further generalize the mixture cure AFT frailty model, we assume the frailty term to follow Generalized Gamma distribution. To find an efficient estimator, the normal kernel smoothed methods suggested by Zeng and Lin (2007) are then applied. This model provides greater flexibility in general for modelling correlated survival data when not all the subjects under study would experience the occurrence of the event of interest. Some concluding remarks are finally presented in Chapter 5, wherein some further research problems of interest are also mentioned.

Chapter 2

Likelihood Inference for Semiparametric Mixture Cure Generalized-Gamma Frailty Model

2.1 Basic Model

Let $O = \{t_{ij}, \delta_{ij}, x_{ij}, i = 1, 2, \ldots, n\}$ denote the observed data, where t_{ij} is the observed time, δ_{ij} is the censoring indicator, and x_{ij} are the covariates. With the assumption of the cure rate, the conditional population survival function can be obtained, when the latency distribution is given, as follows:

$$S_p(t) = p_0 + (1 - p_0)S_s(t), \tag{2.1}$$

where $S_s(t)$ is the susceptible group's survival function with a random effect term y_i or y_{ij}. It is easy to incorporate the frailty term into the latency distribution to deal with the unobserved information as

$$S_p(t) = p_0 + (1 - p_0)L_y(H_0(t)), \tag{2.2}$$

where $L_y(t)$ is the Laplace transform of the frailty variable y_i. However, in this case, when we consider the Generalized Gamma distribution as the frailty term, the Laplace transform would involve an infinite sum and might pose some problem with convergence. For this reason, we use a Monte Carlo approximation instead.

The cure proportion is modelled by the logistic link function of the form

$$p_0 = \frac{1}{1 + \exp(b'X_{ij})}, \tag{2.3}$$

where $b = (b_0, b_1, ..., b_p)$ is a vector of the unknown parameters. As in the proportional hazard model, we assume $\phi(x) = \exp(\beta x)$, where β is a vector of unknown parameters. The corresponding hazard function and survival function for the susceptible group are as follows:

$$h_i(t_{ij}) = y_i h_0(t_{ij}) \exp(\beta' x_{ij}), \quad S_i(t_{ij}) = \exp(-y_i \exp(\beta' x_{ij}) H_0(t_{ij})), \tag{2.4}$$

where the frailty term y_i is assumed to follow the generalized gamma distribution with mean 1. It is not feasible to maximize the observed likelihood function directly if $H_0(t)$ is unknown, and so we apply the EM Algorithm for the estimation problem.

2.2 Estimation Procedure

As estimation procedure is discussed here for the semi-parametric mixture cure frailty PH model. As we have two latent variables I and Y, the EM algorithm is considered to estimate the unknown parameters β, b, θ and H_0 in the considered model. Given the values of the frailty term y_i and indicator I_i, the conditional likelihood function can be expressed as

$$\begin{aligned} L_c(\Theta|O, y, I) = \prod_{i=1}^{n}\prod_{j=1}^{\ell_i} & p_0^{1-I_{ij}}(1-p_0)^{I_{ij}}\big\{\exp(-y_i\exp(\beta' x_{ij})H_0(t_{ij})) \\ & \times [y_i\exp(\beta' x_{ij})h_0(t_{ij})]^{\delta_{ij}}\big\}^{I_{ij}} g(y_i), \end{aligned} \tag{2.5}$$

where I_i is an indicator function with $I_i = 0$ if the patient is non-susceptible or cured and 1 otherwise.

Let us denote the model parameter by Θ=$\{b, \beta, H_0(.)\}$. Then, the corresponding complete log-likelihood function can be obtained as

$$\ell_c(\Theta) = l_{c1}(b) + l_{c2}(\beta, H_0(t)) + l_{c3}(q, \sigma), \tag{2.6}$$

where

$$\begin{aligned} \ell_{c1} &= \sum_{i=1}^{n}\sum_{j=1}^{\ell_i}(1-I_{ij})(-\log(1+\exp(b' X_{ij}))) + I_{ij}(\log(\exp(b' X_{ij})) - \log(1+\exp(b' X_{ij}))) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{\ell_i} I_{ij}\log(\exp(b' X_{ij})) - \log(1+\exp(b' X_{ij})), \end{aligned} \tag{2.7}$$

$$\ell_{c2} = \sum_{i=1}^{n}\sum_{j=1}^{\ell_i} I_{ij}\big(-y_i\exp(\beta' x_{ij})H_0(t_{ij}) + \delta_{ij}(\log(y_i) + \beta' x_{ij} + \log(h_0(t_{ij})))\big), \tag{2.8}$$

$$\ell_{c3} = \sum_{i=1}^{n}\sum_{j=1}^{\ell_i}\Bigg(\log(q) + \frac{1}{q\sigma}\log\big(\Gamma\big(q^{-2}+\frac{\sigma}{q}\big)\big) + (\frac{1}{q\sigma}-1)\log(y_i) - \Bigg(\frac{\Gamma(q^{-2}+\frac{\sigma}{q})}{\Gamma(q^{-2})}y_i\Bigg)^{q/\sigma}\Bigg). \tag{2.9}$$

E-step:

The E-step computes the conditional expectation of the complete log-likelihood with respect to the latent variable I and y_i, given the current estimate. The corresponding Q function, based on the given information $(\Theta^{(m)}, O)$ at the mth iteration, is

$$Q_1 = \sum_{i=1}^{n}\sum_{j=1}^{\ell_i}\mathbb{E}(I_{ij})\log(\exp(b'X_{ij})) - \log(1+\exp(b'X_{ij})), \tag{2.10}$$

$$Q_2 = \sum_{i=1}^{n}\sum_{j=1}^{\ell_i} -\mathbb{E}(y_i I_{ij})\exp(\beta' x_{ij})H_0(t_{ij}) + \sum_{i=1}^{n}\sum_{j=1}^{\ell_i}\delta_{ij}(\beta' x_{ij} + \log(h_0(t_{ij}))), \tag{2.11}$$

$$\begin{aligned} Q_3 = \sum_{i=1}^{n}\sum_{j=1}^{l_i}\Bigg(&\log(q) + \frac{1}{q\sigma}\log\big(\Gamma\big(q^{-2}+\frac{\sigma}{q}\big)\big) + \big(\delta_{ij} + \frac{1}{q\sigma} - 1\big)\mathbb{E}(\log(y_i)) \\ &- (\frac{\Gamma(q^{-2}+\frac{\sigma}{q})}{\Gamma(q^{-2})})^{q/\sigma}\mathbb{E}\bigg(y_i^{q/\sigma}\bigg)\Bigg), \end{aligned} \tag{2.12}$$

where

$$\begin{aligned} \pi_{ij} &= \mathbb{E}(I_{ij}|\Theta^{(m)}, O) = \delta_{ij} + (1-\delta_{ij})\frac{(1-p_0)\times L_{y_i}\big(\exp(\beta' x_{ij})H_0(t)\big)}{p_0 + (1-p_0)\times L_{y_i}\big(\exp(\beta' x_{ij})H_0(t)\big)}, \\ a_{ij} &= \mathbb{E}(y_i I_{ij}|\Theta^{(m)}, O) = \mathbb{E}(y_i|I_{ij}=1, \Theta^{(m)}, O)\times\pi_{ij}, \\ b_{ij} &= \mathbb{E}(\log(y_i)|\Theta^{(m)}, O) \\ &= \mathbb{E}(\log(y_i)|I_{ij}=1, \Theta^{(m)}, O)\times\pi_{ij} + \mathbb{E}(\log(y_i)|I_{ij}=0, \Theta^{(m)}, O)\times(1-\pi_{ij}), \\ c_{ij} &= \mathbb{E}(y_i^{q/\sigma}|\Theta^{(m)}, O) = \mathbb{E}(y_i^{q/\sigma}|I_{ij}=1, \Theta^{(m)}, O)\times\pi_{ij} + \mathbb{E}(y_i^{q/\sigma}|I_{ij}=0, \Theta^{(m)}, O)\times(1-\pi_{ij}), \end{aligned} \tag{2.13}$$

with $\mathbb{E}(I_{ij}\delta_{ij}) = \delta_{ij}$ and $L_{y_i}(.)$ being the Laplace transform of the Generalized Gamma distribution. We performed a Markov chain Monte Carlo (MCMC) approximation in calculating the value, given the parameter in each step.

Obtaining the above conditional expectations of y_i, $\log(y_i)$ and $(y_i)^{q/\sigma}$ are not straight forward. The conditional distribution of y_i, given the information at mth iteration, is proportional to

$$g(y_i|I_{ij}, \Theta^{(m)}, O) \propto y_i^{1/q\sigma+\sum_{j=1}^{\ell_i}\delta_{ij}-1} \exp(-\sum_{j=1}^{\ell_i} I_{ij} y_i \exp\big(\beta' x_{ij}) H_0(t_{ij}) - q^{-2}(\lambda y_i)^{q/\sigma}\big), \tag{2.14}$$

which is not a common distribution. So, we can only solve the problem by considering each subject as one cluster, which replaces y_i by y_{ij}. Note that when $\sigma = q$, $g(y_i)$ is a Gamma distribution $\Gamma(\frac{1}{q^2}, \frac{1}{q^2})$, which is the special that has been studied by Peng and Zhang (2008a).

The expectations in the E-step need to be calculated numerically by using the integration approximation methods such as Markov Chain Monte Carlo (MCMC) methods. Similar work has been performed by Balakrishnan and Peng (2006) and Chen *et al.* (2013) using 'MCMC' package in software R. The idea is to sample posterior distributions from y_{ij} to calculate the above expectations. The procedure is discussed briefly below.

Metropolis Hasting Procedure:

Step 1 For each chain, initialize $y^{(0)}$, given Θ, I and O.

Step 2 For iteration $i = 1, 2, ...$, generate a random proposal y^* near $y^{(i-1)}$ by a jumping distribution $G_t(y^*)$.

Step 3 Calculate the ratio:

$$r = \frac{f(y^{*}|\mathbf{\Theta}, \mathbf{I}, \mathbf{O})/G_t(y^{*})}{f(y^{(i-1)}|\mathbf{\Theta}, \mathbf{I}, \mathbf{O})/G_t(y^{(i-1)})} \tag{2.15}$$

Step 4 Accept the proposal y_i as y^* if the ratio is larger than 1 or a uniform(0,1) random variable.

Step 5 Iterate Steps 2, 3 and 4 until the expected values in the E-step converge to the desired level of accuracy.

The acceptance ratio is kept between 0.2 and 0.25 as suggested by Gelman *et al.* (1996).

M-Step:

The M-Step consists of maximizing Equation (2.10), (2.11) and (2.12), to update the parameters b , β and $H_0(t)$.

As Q_2 contains an unspecified cumulative hazard baseline function $H_0(t)$, we develop a semi-parametric baseline, motivated by Peng and Dear (2000) and Sy and Taylor (2000). This method is based on the estimation method developed by Breslow (1972), and for this reason is called a Breslow-type estimator.

Let $\tau_1 < \cdots < \tau_k$ be the distinct uncensored failure times. D_j is the set of d_j, which represents the uncensored failures at τ_j. Let R_j be the individual at risk set at time τ_j, that is, the set of individuals alive and uncensored prior to τ_j. Let E_j be the set of censored observations in $[\tau_j, \tau_{j+1})$. Denote $h_0(\tau_j) = \alpha_j$ if $\tau_{j-1} < t < \tau_j$, $j = 1, \ldots, k$, and $\tau_0 = 0$.

We obtain Q_2 in Equation (2.11) as

$$Q_2 = \log\Big[\prod_{j=1}^{k}\Big(\prod_{i\in D_j}\alpha_j \exp(\beta' x)\exp\big(-\sum_{i\in D_j\cup E_j} a_{ij}\exp(\beta' x)\sum_{m=1}^{j}\alpha_m(\tau_m-\tau_{m-1})\big)\Big)\Big]. \tag{2.16}$$

Given the current estimate of $\beta^{(m)}$, the baseline cumulative hazard function $\hat{H_0}^{(m)}(t)$ can be obtained by maximizing Q_2 with respect to α_j. Thus, a Breslow-type or Nelson-Aalon type estimator can be obtained as

$$\hat{H}_0^{(m)}(t_{ij}) = \sum_{t_i<t}\frac{d_{t_{ij}}}{\sum_{j\in R(t_{ij})} a_{ij}\exp(\beta' x_{ij})}, \tag{2.17}$$

where a_{ij} is the expectation term in Equation (2.13), $d_{t_{ij}}$ denotes the number of uncensored individuals at time t_{ij}, and $R(t_{ij})$ is the risk set at time t_{ij}.

Correspondingly, the survival function can be obtained at the mth iteration as

$$\hat{S}_0^{(m)}(t_{ij}) = \exp\Big(-\sum_{t_i<t}\frac{d_{t_{ij}}}{\sum_{j\in R(t_{ij})} a_{ij}\exp(\beta' x_{ij})}\Big). \tag{2.18}$$

As suggested by Peng and Zhang (2008a), maximizing Equation (2.10) with respect to b can be calculated through standard logistic regression.

Meanwhile, to maximize Equation (2.11) with respect to β and $H_0()$ can be performed using the standard PH model with covariates $\log(a_{ij})$ with the fixed coefficient 1. Both maximization process can be carried out with package 'smcure' by Cai *et al.* (2012) in software R. Lastly, as the process of maximizing Q_3 function is to update the parameters σ and q in each iteration, the Newton-Raphson algorithm can be applied to maximize Equation (2.12) with the cumulative hazard function $H_0(t)$ following the Breslow-type estimator.

Estimation Procedure:

Step 1 Given the initial values $b_{(0)}$, $\beta_{(0)}$, $\sigma_{(0)}$ and $q_{(0)}$ for the cure frailty model, $\hat{H}_0(t_{ij})$ can be estimated correspondingly;

Step 2 E-Step: Sample y_i from the posterior distribution and compute the corresponding expectations in the Q function;

Step 3 M-Step: Estimate $b_{(m)}$, $\beta_{(m)}$, $\sigma_{(m)}$, $q_{(m)}$ by maximizing the Q function and update the estimation of $\hat{H}_0(t_{ij})$;

Step 4 Iterate Steps 2 and 3 until b, β, σ and q converge to the desired level of accuracy.

Remarks: Based on personal discussion with the authors and open source code in 'smcure' package in R, the convergence condition is defined as setting a threshold value K for parameters of interest, where

$$\sum_{\Theta}(\Theta_{(m+1)} - \Theta_{(m)})^2 \leq K, \tag{2.19}$$

where Θ is the set of parameters, then we stopped the iteration.

In addition, Sy and Taylor (2000) and Peng (2003) discussed the tail of the survival functions and pointed out that the zero tail constraint for the baseline survival function can improve the estimation. Hence, we set $\hat{S}_0(t) = 0$ for t greater than the maximum failure time.

Furthermore, the EM algorithm usually does not produce the standard errors of the estimated parameters. In particular in the mixture cure frailty model, it is difficult to find the information matrix corresponding to the Q functions. The complexity of the second derivatives leads to computational issues. Therefore, we propose bootstrap method for estimating the standard errors of the estimates of b and β in the real data analysis, which is the commonly used variance estimation method in mixture cure frailty model.

2.3 Simulation Study

The purpose of the simulation study is to evaluate the estimation performance of the mixture cure generalized gamma frailty model. Here, we take our simulation in order to demonstrate the process of simulation study. We generate 1000 data sets with sample size $N = 200$ from a mixture cure gamma frailty model. One covariate is considered, which is the binary variable x taking on 0 (control) or 1 (treatment). We assume that x influences the cure rate with $b_0 = 2$ and $b_1 = -1$. Correspondingly, the cure fraction is 12% in the control group and 27% in the treatment group. The correlated covariates $\beta = \log(2)$ is influenced by x. The baseline survival function distribution is generated from the standard Log-normal distribution. The frailty is generated by Gamma distribution or Log-normal distribution with variance 0.5. The censoring time is generated from the uniform distribution to obtain a fraction of approximately 25% for the combined groups.

For a detailed simulation study, we change the setting to test the performance with various sample size, frailty variance and censoring proportions, respectively. The details of the settings used in the study are as Table 2.1.

	Sample size	b_0	b_1	Variance	Censoring proportion	β
1	100	2	-1	0.5	0.25	log(2)
2	100	2	-1	0.5	0.5	log(2)
3	100	2	-1	1	0.25	log(2)
4	100	2	-1	1	0.5	log(2)
5	100	2	-1	1.5	0.25	log(2)
6	100	2	-1	1.5	0.5	log(2)
7	200	2	-1	0.5	0.25	log(2)
8	200	2	-1	0.5	0.5	log(2)
9	200	2	-1	1	0.25	log(2)
10	200	2	-1	1	0.5	log(2)
11	200	2	-1	1.5	0.25	log(2)
12	200	2	-1	1.5	0.5	log(2)

Table 2.1: Simulation Settings.

The simulation procedure can be stated as follows:

1. Generate $N = n_i \times \ell_i$ frailty values from the frailty distribution y_{ij};

2. Assign each simulated individual to treatment group and control group with probability 0.5, and generate covariates x from normal distribution correspondingly;

3. Based on the survival function, we have the frailty term y_{ij} substituted into the equation

$$S(t|y) = \exp\big(-y_{ij}H_0(t)\exp(\beta'X)\big). \tag{2.20}$$

Therefore, the cumulative distribution function can be given as

$$F(t|y) = 1 - \exp\big(-y_{ij} H_0(t) \exp(\beta' X)\big), \tag{2.21}$$

which is assumed to follow a uniform distribution (0,1). So, we can generate $U \sim \text{Uniform(0,1)}$ and set $U = F(t|y)$;

4. Find the cumulative hazard function

$$H_0(.) = -\frac{\log(1 - u_{ij})}{y_{ij} \exp(\beta' X)}; \tag{2.22}$$

5. Find the inverse of the true baseline function, for standard log-normal as

$$t_{ij} = \exp\big(\Phi^{-1}(1 - \exp(-H_0(.)))\big) \tag{2.23}$$

6. Generate the censoring time C from Uniform distribution. Compare t_{ij} and c_{ij} and determine δ_{ij} for each subject, so that we can adjust the censoring rate to satisfy our requirements.

Variance	True Frailty	Censoring proportion	Parameters	Bias	MSE(σ^2_{boot})	Coverage Probability
0.5	Gamma	25	$\hat{\sigma}^2_f$	-0.021	0.038(0.041)	0.947
			$\hat{b}_0$	0.122	0.257(0.291)	0.958
			$\hat{b}_1$	-0.087	0.189(0.141)	0.943
			$\hat{\beta}_1$	0.013	0.016 (0.022)	0.934
	Log-normal	25	$\hat{\sigma}^2_f$	0.028	0.014 (0.021)	0.959
			$\hat{b}_0$	0.158	0.202 (0.196)	0.963
			$\hat{b}_1$	-0.089	0.161 (0.154)	0.954
			$\hat{\beta}_1$	0.022	0.011 (0.014)	0.935
	Gamma	50	$\hat{\sigma}^2_f$	-0.036	0.052	
			$\hat{b}_0$	0.201	0.358	
			$\hat{b}_1$	0.125	0.338	
			$\hat{\beta}_1$	-0.045	0.129	
	Log-normal	50	$\hat{\sigma}^2_f$	-0.077	0.101	
			$\hat{b}_0$	0.146	0.346	
			$\hat{b}_1$	-0.152	0.104	
			$\hat{\beta}_1$	-0.048	0.067	

Table 2.2: Simulation results with sample size 200 and true variance $\sigma^2_f = 0.5$

Variance	True Frailty	Censoring proportion	Parameters	Bias	MSE
	Gamma	25	$\hat{\sigma}^2$	0.060	0.048
			$\hat{b}_0$	0.125	0.096
			$\hat{b}_1$	-0.175	0.133
			$\hat{\beta}_1$	0.012	0.032
	Log-normal	25	$\hat{\sigma}^2$	-0.064	0.018
			$\hat{b}_0$	0.098	0.178
			$\hat{b}_1$	-0.121	0.292
			$\hat{\beta}_1$	0.016	0.041
	Gamma	50	$\hat{\sigma}^2$	-0.087	0.066
			$\hat{b}_0$	0.207	0.271
			$\hat{b}_1$	-0.184	0.149
			$\hat{\beta}_1$	0.022	0.014
	Log-normal	50	$\hat{\sigma}^2$	0.089	0.079
			$\hat{b}_0$	0.185	0.167
			$\hat{b}_1$	-0.162	0.256
			$\hat{\beta}_1$	0.025	0.019

Table 2.3: Simulation results with sample size 200 and true variance $\sigma_f^2 = 1$

Variance	True Frailty	Censoring proportion	Parameters	Bias	MSE
1.5	Gamma	25	$\hat{\sigma}^2$	0.081	0.111
			$\hat{b}_0$	-0.023	0.092
			$\hat{b}_1$	0.120	0.273
			$\hat{\beta}_1$	0.052	0.092
	Log-normal	25	$\hat{\sigma}^2$	-0.064	0.198
			$\hat{b}_0$	-0.224	0.315
			$\hat{b}_1$	-0.119	0.173
			$\hat{\beta}_1$	0.063	0.165
	Gamma	50	$\hat{\sigma}^2$	0.097	0.099
			$\hat{b}_0$	0.184	0.245
			$\hat{b}_1$	0.152	0.121
			$\hat{\beta}_1$	0.041	0.059
	Log-normal	50	$\hat{\sigma}^2$	0.039	0.052
			$\hat{b}_0$	0.121	0.182
			$\hat{b}_1$	0.145	0.173
			$\hat{\beta}_1$	0.077	0.097

Table 2.4: Simulation results with sample size 200 and true variance $\sigma_f^2 = 1.5$

Variance	True Frailty	Censoring proportion	Parameters	Bias	MSE
0.5	Gamma	25	$\hat{\sigma}_f^2$	0.045	0.073
			$\hat{b}_0$	-0.208	0.173
			$\hat{b}_1$	-0.107	0.098
			$\hat{\beta}_1$	0.031	0.042
	Log-normal	25	$\hat{\sigma}_f^2$	0.039	0.025
			$\hat{b}_0$	0.109	0.144
			$\hat{b}_1$	0.104	0.125
			$\hat{\beta}_1$	0.031	0.040
	Gamma	50	$\hat{\sigma}_f^2$	-0.033	0.029
			$\hat{b}_0$	-0.104	0.123
			$\hat{b}_1$	0.106	0.128
			$\hat{\beta}_1$	-0.024	0.094
	Log-normal	50	$\hat{\sigma}_f^2$	-0.057	0.077
			$\hat{b}_0$	-0.203	0.106
			$\hat{b}_1$	0.122	0.222
			$\hat{\beta}_1$	-0.033	0.074

Table 2.5: Simulation results with sample size 100 and true variance $\sigma_f^2 = 0.5$

Variance	True Frailty	Censoring proportion	Parameters	Bias	MSE
1	Gamma	25	$\hat{\sigma}^2$	-0.042	0.056
			$\hat{b}_0$	0.165	0.183
			$\hat{b}_1$	0.126	0.128
			$\hat{\beta}_1$	0.044	0.028
	Log-normal	25	$\hat{\sigma}^2$	-0.056	0.017
			$\hat{b}_0$	0.086	0.079
			$\hat{b}_1$	0.163	0.111
			$\hat{\beta}_1$	-0.076	0.062
	Gamma	50	$\hat{\sigma}^2$	0.029	0.040
			$\hat{b}_0$	0.109	0.137
			$\hat{b}_1$	0.137	0.184
			$\hat{\beta}_1$	0.067	0.081
	Log-normal	50	$\hat{\sigma}^2$	-0.045	0.039
			$\hat{b}_0$	0.124	0.146
			$\hat{b}_1$	0.192	0.204
			$\hat{\beta}_1$	0.103	0.130

Table 2.6: Simulation results with sample size 100 and true variance $\sigma_f^2 = 1$

Variance	True Frailty	Censoring proportion	Parameters	Bias	MSE
1.5	Gamma	25	$\hat{\sigma}^2$	0.108	0.135
			$\hat{b}_0$	0.125	0.096
			$\hat{b}_1$	-0.175	0.133
			$\hat{\beta}_1$	0.012	0.032
	Log-normal	25	$\hat{\sigma}^2$	0.154	0.118
			$\hat{b}_0$	-0.079	0.128
			$\hat{b}_1$	0.081	0.109
			$\hat{\beta}_1$	0.076	0.051
	Gamma	50	$\hat{\sigma}^2$	0.097	0.126
			$\hat{b}_0$	-0.106	0.120
			$\hat{b}_1$	0.176	0.163
			$\hat{\beta}_1$	-0.041	0.052
	Log-normal	50	$\hat{\sigma}^2$	0.112	0.104
			$\hat{b}_0$	-0.162	0.198
			$\hat{b}_1$	0.125	0.181
			$\hat{\beta}_1$	0.038	0.049

Table 2.7: Simulation results with sample size 100 and true variance $\sigma_f^2 = 1.5$

In Tables 2.2-2.7, we have presented the results of simulation study from which we observe that the estimation is satisfactory; the bootstrap is applied for estimating the variances of the parameter estimates obtained from the EM algorithm. To check the performance of the bootstrap variance estimator, we computed the estimated variances from 500 bootstrap samples for the Case 7 and the mean of estimated variances is shown in Table 2.2. We observe from this table that the MSE and the bootstrap variance estimates are quite close, which reveals that the bootstrap variance estimator can be trusted. Meanwhile, as the censoring proportion increases, the bias and MSE of the cure coefficient b increase, which means that higher censoring rate will reduce the accuracy of the cure rate estimation. The frailty variance estimation is stable regardless the true baseline. The proportional hazard coefficient β estimation remains stable in all simulation cases. In most cases, the coverage probabilities are quite close to the nominal level. Moreover, larger sample size and smaller censoring rate lead to more accurate estimates. The change of frailty variance does not have a clear effect on the accuracy or precision of the parameter estimates. In conclusion, the simulation results show that the EM algorithm provides satisfactory results.

2.4 Application to Bone Marrow Transplant Data

Price and Manatunga (2001) considered bone marrow transplant study for the leukaemia patients. Leukaemia patients received either an allogeneic transplant or an autologous transplant. The maximum followed up time is 1845 days and the time to recurrence or censoring is recorded. The details of the data set have been described earlier in Section 1.4.

To test our model and compare our model with some common mixture cure frailty models, we fit the bone marrow transplant data set and Table 2.8 shows the obtained results. As the log-likelihood is calculated separately from the Q functions and may not maximize the log-likelihood, they are just listed for illustration. When we compare the frailty variance estimated, the Generalized Gamma frailty shows the greatest value, which indicates that the single parameter distributions, such as Gamma, Lognormal and Weibull distributions usually have less flexibility and underestimate the frailty variance. This result is consistent with Peng and Zhang (2008a) and Chen *et al.* (2013).

	Gamma		Log-normal		Weibull		GG	
	Estimate	Var	Estimate	Var	Estimate	Var	Estimate	Var
$\hat{\beta}_1$	0.6266	0.1637	0.6523	0.1932	0.6341	0.1856	0.6293	0.1724
b_0	1.0817	0.1222	1.0276	0.1788	1.0572	0.2082	1.0750	0.2422
b_1	0.4355	0.2805	0.4236	0.4021	0.4327	0.3398	0.4739	0.3663
σ							0.9187	0.2895
q							0.7498	0.3781
$\hat{\theta}$	1.5832	0.3741						
$\hat{\sigma}_f^2$	0.6316		0.6738		0.7607		0.8547	
ℓ	-196.615		-197.28		-196.89		-194.514	

Table 2.8: Estimation results using the mixture cure PH Generalized Gamma frailty model for bone marrow transplant data

Note: The estimated variances(Var) are from the bootstrap method using 500 repeated samples. The estimated frailty variance ($\hat{\sigma}_f^2$) denotes the variance calculated based on the estimation.

Note: The log-likelihood ℓ presented is not the maximized log-likelihood, and so the our estimation methods carried out can not be proved to be efficient, they are just presented for reference.

According to the estimation results, choosing the Gamma, Log-normal, Weibull and Generalized Gamma distributions as frailty do not influence the cure rate significantly. Correspondingly, the cure fractions for allogeneic group are 25.3%, 26.4%, 25.8% and 25.4%, compared to 27% obtained directly from the Kaplan-Meier curve in Figure 1.1. Furthermore, the cure fractions for autologous group are 18.0%, 19.0%, 18.4% and 17.5%, respectively. The results are consistent with the work by Peng and Zhang (2008a). The results indicate that the cure frailty model tends to give a lower estimate of the cure proportion compared to the mixture cure model without frailty assumption, and if the frailty model is more flexible (with more parameters), the cure proportion tends to be smaller. This model helps reveal the heterogeneity in the BMT data and also as stated by Balakrishnan and Peng (2006), Generalized Gamma distribution has considerable flexibility to capture the characteristics in a distribution that might have been missed by the use of any of its special cases. This motivates us to use it as the frailty distribution to model the frailty term. For the purpose of comparing the nested models, we can employ the Likelihood Ratio Test (LRT). If we perform the LRT accordingly, we can find the test statistic as $-2\log(\hat{\ell}_0 - \hat{\ell}) = (4.202, 5.532, 4.752)$ accordingly. All the test statistic are greater than 3.8414 at significance level 5%, which indicates the Generalized Gamma distribution is a satisfactory choice as frailty distribution.

Chapter 3

Estimation of the Mixture Cure Accelerated Failure Time Model with Gamma Frailty

3.1 Basic Model

Let T_i^* be the failure time of the ith subject. Then, the observed time from the subject is denoted by T_i=min(T_i^*, C_i), where C_i is a censoring time. The censoring indicator $\delta_i = 1$ if $T_i{}^* = T_i$ and 0 otherwise and the censoring is assumed to be non-informative. Also, X_i is a vector of covariates for the i-th individual.

Based on the given information, the accelerated failure time (AFT) model for failure time is given by

$$\log(T_i) = \beta' x_i + \epsilon_i, \tag{3.1}$$

where β is the coefficient of interest and ϵ are independent random errors. Assuming each object is a cluster with single element, that is for every cluster we have only one individual, we can propose the Mixture cure AFT model with a frailty term. Following Zhang and Peng (2007a), the hazard function of ϵ_i can be written as

$$h(\epsilon_i|Y_i = y_i) = y_i h_0(\epsilon_i), \tag{3.2}$$

where $h_0(\epsilon)$ is an arbitrary baseline hazard function and y_i is an independent random term for each subject. Here, if we do not include covariate effects, this would be a special case of the general frailty model considered by Kalbfleisch and Prentice (1973). The most common and convenient choice of frailty is the Gamma frailty with mean 1 and variance $1/\theta$:

$$f(y_i) = \theta^\theta {y_i}^{\theta-1} \exp(-y_i\theta)/\Gamma(\theta), \quad y_i > 0, \quad \theta > 0 \tag{3.3}$$

The advantage of the gamma distribution as a frailty model is that the distribution in the E-step of the EM algorithm, will still remain as gamma distribution. This expression for the expectations of interest, which makes the E-Step easier.

Furthermore, to accommodate the mixture cure model, we define a cure fraction for some subjects using the incidence component:

$$p_0 = \frac{1}{1 + \exp(b'x)}, \tag{3.4}$$

where b is a vector of unknown parameters of interest modelling the cure fraction. The conditional survival function of T, given that the patient is not cured, is $S(\epsilon_i)$.

For the sake of convenience, let us denote the observed information $O = \{t_i, \delta_i, x_i\}$ for the ith individual, $i = 1, ..., n$. Then, the complete likelihood function can be written as

$$\prod_{i=1}^{n}(p_0)^{1-I_i} \times \{(1-p_0)h(\epsilon_i)^{\delta_i}S(\epsilon_i)\}^{I_i} \times f(y_i), \tag{3.5}$$

where I_i is an indicator function with $I_i = 0$ if the patient is non-susceptible or cured and 1 otherwise.

3.2 Estimation Procedure

An estimation procedure is discussed here for the semi-parametric mixture cure frailty AFT model. As we have two latent variables I and Y, the EM algorithm is considered to estimate the unknown parameters β, b, θ and H_0 in the proposed model. Given the values of the frailty term y_i and indicator I_i, the conditional likelihood function can be expressed as

$$\ell_c(\beta, \sigma, q) \propto$$
$$\prod_{i=1}^{n}(p_0)^{1-I_i}(1-p_0)^{I_i} \times \prod_{i=1}^{n}\Big(h_0(\epsilon_i)\Big)^{I_i\delta_i} {y_i}^{I_i\delta_i} \exp\big(-I_i y_i H_0(\epsilon_i)\big) \times \prod_{i=1}^{n}\theta^{\theta}{y_i}^{\theta-1}\exp(-y_i\theta)/\Gamma(\theta). \tag{3.6}$$

The logarithm of the three complete likelihood functions can be given as:

$$\begin{aligned}
\ell_{c1}(b_0, b_1) &= \sum_{i=1}^{n}(1-I_i)\log(p_0) + I_i\log(1-p_0),\\
\ell_{c2}(\beta, H_0(t)) &= \sum_{i=1}^{n}\delta_i\log\big(h_0(\epsilon_i)\big) - I_i y_i H_0(\epsilon_i),\\
\ell_{c3}(\theta) &= \sum_{i=1}^{n}[\theta\log(\theta) - \log(\Gamma(\theta)) - y_i\theta + (\delta_i+\theta-1)\log(y_i)],
\end{aligned} \tag{3.7}$$

where $\delta_i I_i = \delta_i$ always holds.

E-step:

The E-step calculates the conditional expectation of the complete log-likelihood corresponding to the latent variables I_i and y_i, given the current estimated parameters, denoted by $\Theta^{(m)} = \{b_0^{(m)}, b_1^{(m)}, H_0^{(m)}(t)\}$ and the observed data O. In this case, we find

$$\pi_i = \mathbb{E}(I_i|\Theta^{(m)}) = \delta_i + (1-\delta_i)\frac{(1-p_0)L_{y_i}(H_0^{(m)}(\epsilon_i))}{p_0 + (1-p_0) \times L_{y_i}(H_0^{(m)}\big(\epsilon_i\big))}, \tag{3.8}$$

$$a_i = \mathbb{E}(y_i|\Theta^{(m)}) = \mathbb{E}(y_i|I=1,\Theta^{(m)}) \times \pi_i + \mathbb{E}(y_i|I=0,\Theta^{(m)}) \times (1-\pi_i), \tag{3.9}$$

$$b_i = \mathbb{E}(y_i I_i|\Theta^{(m)}) = \mathbb{E}(y_i I_i|I=1,\Theta^{(m)}) \times \pi_i, \tag{3.10}$$

$$c_i = \mathbb{E}(\log(y_i)|\Theta^{(m)}) = \mathbb{E}(\log(y_i)|I=1,\Theta^{(m)}) \times \pi_i + \mathbb{E}(\log(y_i)|I=0,\Theta^{(m)}) \times (1-\pi_i), \tag{3.11}$$

where $L_{y_i}(t) = (1+\theta^{-1}t)^{-\theta}$ is the Laplace transform of the gamma random variable. However, some expectation terms involve y_i in the E-step, and are therefore not very straightforward to calculate. To find the expectation with respect to the conditional distribution of the frailty term y_i, we observe that the conditional distribution of y_i is proportional to

$$\exp(y_i(I_i H_0(\epsilon_i) + \theta)){y_i}^{\delta_i+\theta-1}; \tag{3.12}$$

thus, we have

$$
\begin{aligned}
y_i|(I_i = 0) &\sim \text{Gamma}(\delta_i + \theta, \theta^{-1}), && (3.13)\\
y_i|(I_i = 1) &\sim \text{Gamma}\big(\delta_i + \theta, \frac{1}{\theta + H_0(\epsilon_i)}\big). && (3.14)
\end{aligned}
$$

This indicates that for the E-step, given the current estimate of $\Theta^{(m)}$, the conditional expectations are

$$
\begin{aligned}
a_i &= \frac{\delta_i + \theta}{\theta + H_0(\epsilon_i)}\pi_i + \frac{\delta + \theta}{\theta}(1 - \pi_i), && (3.15)\\
b_i &= \frac{\delta_i + \theta}{\theta + H_0(\epsilon_i)}\pi_i, && (3.16)\\
c_i &= \big(\phi(\delta_i + \theta) - \log(\theta + H_0(\epsilon_i)\big)\pi_i + (\phi(\delta_i + \theta) - \log(\theta))(1 - \pi_i), && (3.17)
\end{aligned}
$$

where $\phi(.)$ is the digamma function. Substituting them into the complete log-likelihood function, the conditional expectations of the three complete log-likelihood functions in the E-Step are given by Q-functions as

$$
\begin{aligned}
Q_1(b_0, b_1) &= \sum_{i=1}^{n}(1 - \pi_i)\log(p_0) + \pi_i \log(1 - p_0),\\
Q_2(\beta, H_0(t)) &= \sum_{i=1}^{n} \delta_i \log\big(h_0(\epsilon_i)\big) - b_i H_0(\epsilon_i), \qquad (3.18)\\
Q_3(\theta) &= \sum_{i=1}^{n}[\theta\log(\theta) - \log(\Gamma(\theta)) - a_i\theta + c_i(\delta_i + \theta - 1)].
\end{aligned}
$$

M-Step:

The M-Step is to maximize Q_1, Q_2 and Q_3 with respect to the unknown parameters b, β and $H_0()$.

For finding $H_0(.)$, Zhang and Peng (2007b) and Zhang and Peng (2009) introduced a method for the estimation of β. As we take a ${\epsilon_i}^*$ as an error term with unknown distribution following the usual AFT model

$${\epsilon_i}^* = \log(T_i) - \beta x,$$

Zhang and Peng (2007b) stated that the form of Q_2 can be turned into a standard semi-parametric AFT mixture (except for b_i). This enables us to estimate β based on the methods for the semi-parametric AFT model.

Following the methods of Wei (1992) and Pan (2001), we use a rank estimation method. If we take derivative of the logarithm of the partial likelihood function Q_2 for the model with respect to β and extend to include a general (predictable) weight function under suitable assumptions, the function of β is obtained as

$$\Psi(\beta, k(.)) = \sum_{i=1}^{n} \delta_i k(\epsilon_i^*)\left(x_i - \frac{\sum_{j=1}^{n} x_j b_i^{(m)} I(\epsilon_j^* \geq \epsilon_i^*)}{\sum_{j=1}^{n} b_i^{(m)} I(\epsilon_j^* \geq \epsilon_i^*)} \right), \tag{3.19}$$

where $I(.)$ is the indicator function and $k(.)$ is a general (predictable) weight function. Fygenson and Ritov (1994) proved that when the Gehan weight function $k(u) = \sum_{j=1}^{n} I(\epsilon_j^* \geq u) b_i^{(m)}/n$ is used, the estimating equation is monotone (without the constant term ${b_i}^{(m)}$).

Under the Gehan-type weight function, $\Psi(\beta, k(.))$ in (3.19) becomes a monotone function of β, and gets simplified as

$$\Psi(\beta, k(.)) = \sum_{i=1}^{n} \sum_{j=1}^{n} n^{-1} \delta_i b_i^{(m)} (x_i - x_j) I(\epsilon_i^* < \epsilon_j^*). \tag{3.20}$$

Therefore, if there is a solution to $\Psi(\beta, k(.)) = 0$, it will be unique and consistent. Moreover, another advantage of using the Gehan-type weight function, suggested by Zhang and Peng (2007b), is that it can be considered as the gradient of a convex function

$$L_G(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{n} n^{-1} \delta_i b_i^{(m)} |\epsilon_i^* - \epsilon_j^*| I(\epsilon_i^* < \epsilon_j^*). \tag{3.21}$$

As $L_G(\beta)$ is convex, finding the root of $\Psi(\beta, k(.)) = 0$ is the same as minimizing $L_G(\beta)$, which can be carried out by using the linear programming method.

Hence, a Breslow-type baseline estimator can be obtained and updated as

$$\hat{S}_0(t_i) = \exp(-\sum_{t_i < t} \frac{d_{t_i}}{\sum_{j \in R(t_i)} b_i^{(m)}}), \tag{3.22}$$

where $b_i^{(m)}$ is the expectation term in Equation (3.16), d_{t_i} denotes the number of uncensored times at time t_i, and $R(t_i)$ is the risk set at time t_i.

Estimation Procedure:

Step 1 Given the initial values of $b_{(0)}$, $\beta_{(0)}$ and $\theta_{(0)}$ from the cure frailty model, $\hat{H}_0(t_i)$ can be estimated correspondingly;

Step 2 E-Step: Calculate the corresponding expectations and substitute the expectations in the Q function;

Step 3 M-Step: Estimate $b_{(m)}$, $\beta_{(m)}$ and $\theta_{(m)}$ by maximizing the Q function and update the estimation of $\hat{H}_0(t_{ij})$;

Step 4 Iterate Steps 2 and 3 until b, β and θ converge to the desired level of accuracy.

Remarks: Based on personal discussion with the authors and open source code in 'smcure' package in R, the convergence condition is defined as setting a threshold value K for the sum of squared error of parameters of interest, where

$$\sum_{\Theta}(\Theta_{(m+1)} - \Theta_{(m)})^2 \leq K, \tag{3.23}$$

where Θ is the set of parameters.

3.3 Simulation Study

As mentioned by Zhang and Peng (2009), the choice of the latency distribution function could introduce non-identifiability of the survival model S_{pop} and consequently affect the estimate of the cure fraction. For the purpose of testing the identifiablility and the estimation performance of the semi-parametric mixture cure AFT frailty model, a simulation study is carried out in this section.

Here, we take our simulation settings in order to demonstrate the process of simulation study. For example, in one simulation study, we generate 1000 data sets from the model specified. In each data set, we assume that a single covariate x, has effects on both the incidence and the latency component of the cure model and that there are 50% of patients with $x = 0$ (control group) and 50% of patients with $x = 1$ (treatment group). The effect of x on the incidence is through $b_0 = 2$ and $b_1 = -1$. Then, the corresponding cure rates will be 11.9% and 26.9% respectively, in the control and treatment groups. The effect of x on the latency is through $\beta_1 = \log(2)$.

The baseline survival function distribution is assumed to be standard Log-normal distribution for simplicity. The frailty is generated by Gamma or Log-normal distributions with variance 0.5. The censoring time is generated from the uniform distribution to obtain a censoring fraction of 25%. The method of simulating the data is similar to what has been discussed in Section 2.3.

To carry out a detailed simulation study, we consider different simulation settings to test the performance with various sample sizes, frailty variance and censoring proportions. The details of the settings are presented in Table 3.1.

	Sample size	b_0	b_1	Variance	Censoring proportion	β
1	100	2	-1	0.5	0.25	log(2)
2	100	2	-1	0.5	0.5	log(2)
3	100	2	-1	1	0.25	log(2)
4	100	2	-1	1	0.5	log(2)
5	100	2	-1	2	0.25	log(2)
6	100	2	-1	2	0.5	log(2)
7	200	2	-1	0.5	0.25	log(2)
8	200	2	-1	0.5	0.5	log(2)
9	200	2	-1	1	0.25	log(2)
10	200	2	-1	1	0.5	log(2)
11	200	2	-1	2	0.25	log(2)
12	200	2	-1	2	0.5	log(2)

Table 3.1: Simulation Settings.

Table 3.2: Simulation results with sample size 200 with true frailty distribution as Gamma

θ	Censoring proportion	Parameters	Bias	MSE
0.5	25	σ_f^2	-0.032	0.036
		$\hat{b}_0$	0.098	0.262
		$\hat{b}_1$	0.074	0.323
		$\hat{\beta}_1$	-0.052	0.042
0.5	50	σ_f^2	-0.019	0.018
		$\hat{b}_0$	0.411	0.382
		$\hat{b}_1$	-0.304	0.550
		$\hat{\beta}_1$	-0.038	0.132
1	25	σ_f^2	-0.011	0.006
		$\hat{b}_0$	0.207	0.271
		$\hat{b}_1$	-0.184	0.149
		$\hat{\beta}_1$	0.083	0.135
1	50	σ_f^2	0.023	0.013
		$\hat{b}_0$	0.337	0.549
		$\hat{b}_1$	-0.218	0.472
		$\hat{\beta}_1$	0.128	0.119
2	25	σ_f^2	-0.098	0.041
		$\hat{b}_0$	0.087	0.131
		$\hat{b}_1$	-0.062	0.176
		$\hat{\beta}_1$	0.044	0.032
2	50	σ_f^2	0.027	0.009
		$\hat{b}_0$	0.189	0.421
		$\hat{b}_1$	-0.114	0.556
		$\hat{\beta}_1$	0.094	0.203

Table 3.3: Simulation results with sample size 100 with true frailty distribution as Gamma

.

θ	Censoring proportion	Parameters	Bias	MSE (σ_f^2)	Coverage Probability
0.5	25	σ_f^2	-0.022	0.029(0.023)	0.953
		$\hat{b}_0$	0.189	0.331(0.363)	0.951
		$\hat{b}_1$	-0.158	0.314(0.321)	0.948
		$\hat{\beta}_1$	-0.013	0.074 (0.081)	0.937
0.5	50	σ_f^2	0.014	0.013(0.019)	0.941
		$\hat{b}_0$	0.310	0.291(0.308)	0.957
		$\hat{b}_1$	-0.207	0.498(0.449)	0.938
		$\hat{\beta}_1$	0.028	0.068(0.079)	0.934
1	25	σ_f^2	0.018	0.023	
		$\hat{b}_0$	0.210	0.258	
		$\hat{b}_1$	-0.108	0.289	
		$\hat{\beta}_1$	0.015	0.052	
1	50	σ_f^2	0.005	0.014	
		$\hat{b}_0$	0.256	0.499	
		$\hat{b}_1$	-0.161	0.517	
		$\hat{\beta}_1$	0.036	0.089	
2	25	σ_f^2	-0.004	0.018	
		$\hat{b}_0$	0.093	0.244	
		$\hat{b}_1$	-0.188	0.355	
		$\hat{\beta}_1$	0.023	0.044	
2	50	σ_f^2	-0.024	0.031	
		$\hat{b}_0$	0.362	0.402	
		$\hat{b}_1$	-0.144	0.488	
		$\hat{\beta}_1$	0.044	0.101	

Table 3.4: Simulation results with sample size 200 with true frailty distribution as Log-normal

θ	Censoring proportion	Parameters	Bias	MSE
0.5	25	σ_f^2	0.013	0.012
		$\hat{b}_0$	0.162	0.305
		$\hat{b}_1$	-0.063	0.321
		$\hat{\beta}_1$	-0.035	0.098
0.5	50	σ_f^2	0.021	0.016
		$\hat{b}_0$	0.283	0.271
		$\hat{b}_1$	-0.148	0.379
		$\hat{\beta}_1$	0.059	0.144
1	25	σ_f^2	-0.008	0.019
		$\hat{b}_0$	0.166	0.239
		$\hat{b}_1$	-0.084	0.208
		$\hat{\beta}_1$	-0.029	0.073
1	50	σ_f^2	-0.019	0.029
		$\hat{b}_0$	0.268	0.316
		$\hat{b}_1$	-0.121	0.493
		$\hat{\beta}_1$	0.031	0.116
2	25	σ_f^2	0.016	0.023
		$\hat{b}_0$	0.198	0.285
		$\hat{b}_1$	-0.139	0.317
		$\hat{\beta}_1$	-0.033	0.178
2	50	σ_f^2	-0.025	0.036
		$\hat{b}_0$	0.256	0.391
		$\hat{b}_1$	-0.169	0.512
		$\hat{\beta}_1$	0.086	0.160

Table 3.5: Simulation results with sample size 100 with true frailty distribution as Log-normal

θ	Censoring proportion	Parameters	Bias	MSE
0.5	25	σ_f^2	0.029	0.019
		$\hat{b}_0$	0.191	0.362
		$\hat{b}_1$	-0.079	0.435
		$\hat{\beta}_1$	0.067	0.132
0.5	50	σ_f^2	0.023	0.026
		$\hat{b}_0$	0.340	0.468
		$\hat{b}_1$	-0.190	0.523
		$\hat{\beta}_1$	0.082	0.093
1	25	σ_f^2	-0.025	0.032
		$\hat{b}_0$	0.172	0.262
		$\hat{b}_1$	-0.143	0.370
		$\hat{\beta}_1$	-0.061	0.059
1	50	σ_f^2	0.021	0.025
		$\hat{b}_0$	0.258	0.389
		$\hat{b}_1$	-0.233	0.467
		$\hat{\beta}_1$	0.077	0.088
2	25	σ_f^2	0.033	0.028
		$\hat{b}_0$	0.261	0.321
		$\hat{b}_1$	-0.157	0.240
		$\hat{\beta}_1$	-0.099	0.140
2	50	σ_f^2	0.048	0.037
		$\hat{b}_0$	0.229	0.461
		$\hat{b}_1$	-0.199	0.394
		$\hat{\beta}_1$	0.105	0.217

Tables 3.2 - 3.5 present the results of the simulation study. The Bias and MSE are calculated for the estimates of all the parameters. Based on the simulation results, we observe that, the MSEs of b_0 and b_1 increase greatly when the censoring rate increases from 25% to 50%. Meanwhile, the performance of the variance and the proportional hazard coefficient β estimation is satisfactory in all cases, regardless of the changes in the true variances and the true frailty distribution.

Here, as also done in Section 2.3, we implemented a 500 bootstrap simulations for estimating the variance of the estimates for Cases 1 and 2. As the MSE, estimated variance and the bootstrap variance are all close in general, we can conclude that the bootstrap variance is quite reliable as mentioned by Peng and Zhang (2008a). In most cases, the CPs are quite close to the nominal level. In summary, larger sample size and less censoring results in accurate estimates. The change of frailty variance does not have a clear effect on the accuracy or precision of the parameter estimates.

3.4 Application to Bone Marrow Transplant Data

As an application of the studied model, we consider the bone marrow transplant study for leukaemia patients. This data set, first studied by Kersey *et al.* (1987), was described earlier in Chapter 1.

In this data set, patients with leukaemia received either an allogeneic transplant or an autologous transplant. There were 46 patients in the allogeneic treatment group and the other patients in the autologous treatment group. They were followed up to maximum 1845 days, and time to recurrence or censoring is recorded. Among them, 33 patients experienced a recurrence of leukaemia in the allogeneic treatment and 35 patients experienced a recurrence in the autologous treatment group.

Based on the plots of Kaplan-Meier survival functions in Figure 1.1, we observe that the two treatment groups are not quite proportional to each other, and so the assumption of AFT model may be more appropriate in this situation. It also shows that both curves level off at a value substantially greater than 0 after one or two years of follow-up, which means that some of the patients will not experience a recurrence after the treatments and should be considered as cured subjects.

Zhang and Peng (2007b) analyzed these data using a semi-parametric mixture cure AFT model and pointed out that the PH model studied in the literature is not appropriate since the proportional hazard assumption is not satisfied for this data set. Also, Peng and Zhang (2008a) suggested using a semi-parametric mixture cure frailty model for the purpose of accounting for the lack of proportionality.

Based on the cumulative hazard functions in Figure 3.1, we can confirm that the AFT model is more appropriate for this data set.

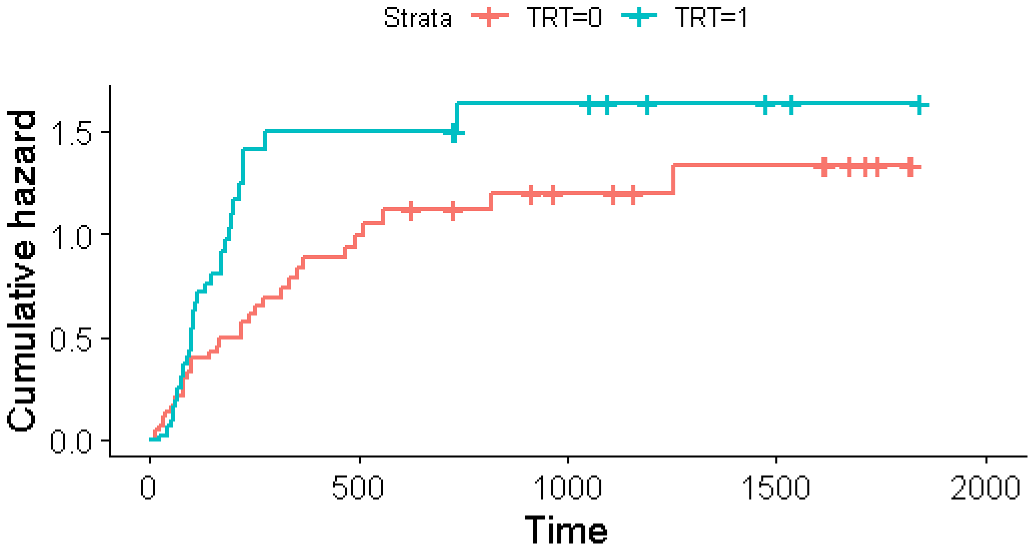

Figure 3.1: Bone marrow transplant data: Cumulative Hazard Functions

Therefore, we apply the semi-parametric mixture cure AFT frailty model for the data. We fit the data with the semi-parametric AFT mixture cure model using the method described in the previous sections. The standard errors of the parameter estimates are obtained based on 500 bootstrap samples.

	Bias	Var
$\hat{\beta}_1$	-0.359	0.080
$\hat{b}_0$	1.021	0.118
$\hat{b}_1$	0.456	0.268
$\hat{\theta}$	1.287	0.507
$\hat{\sigma}_f^2$	0.777	

Table 3.6: Estimated parameters from the mixture cure AFT Gamma frailty model for bone marrow transplant data

.

The estimate of β_1 is -0.359 and the corresponding standard error is $\sqrt{0.080} = 0.282$, yielding the p-value as 0.1015. This is consistent with the conclusion of Zhang and Peng (2007b) that we can state that we do not see a significant difference in the occurrence of the bone marrow engraftment between the patients treated with autologous bone marrow transplant and allogeneic bone marrow transplant if they are not cured. The corresponding frailty variance is 0.777, which is greater than the estimated value in the mixture cure Gamma frailty distribution and less than the estimated value in the mixture cure Generalized Gamma frailty distribution. This may indicate an underestimate of frailty variance due to the lack of flexibility in the model.

The cure fraction for the allogeneic group is $1/(1 + \exp(1.021)) = 0.2648$, and it is close to 26.28% reported by Zhang and Peng (2007a)'s result, both of which are close to 27%, observed directly from the Kaplan-Meier survival curve in Figure 1.1. The cure fraction for the autologous group is $1/(1 + \exp(1.021 + 0.456)) = 0.1859$. Furthermore, based on the estimation results, we conclude that the mixture cure AFT frailty model also helps in identifying the heterogeneity in the BMT data.

Chapter 4

Mixture Cure Model with Accelerated Failure time and Flexible Random Effects

4.1 Basic Model

Let T_{ij}^{*} be the failure time of the jth subject in the ith cluster, and X_{ij} be a vector of covariates, for $i = 1, ..., n$ and $j = 1, ..., \ell_i$. The observed time from the subject is denoted by T_{ij}=min(T_{ij}^{*}, C_{ij}), where C_{ij} is a censoring time. The censoring indicator $\delta_{ij} = 1$ if ${T_{ij}}^{*} = T_{ij}$ and 0 otherwise, and the censoring is assumed to be non-informative.

Based on the information $O = \{T_{ij}, \delta_{ij}, X_{ij}\}$, the AFT model for dependent failure times can be specified by

$$\begin{aligned}
&T_{ij} = \exp(\beta' X_{ij}) V_{ij}, \\
&h(V_{ij}|y_{ij}) = y_{ij} h_0(V_{ij}), \\
&h(t_{ij}|X_{ij}, y_{ij}) = y_{ij} \exp(-\beta' X_{ij}) h_0(t_{ij} \exp(-\beta' X_{ij})), \\
&S(t_{ij}|X_{ij}, y_{ij}) = S_0(t_{ij} \exp(-\beta' X_{ij}))^{y_{ij}}.
\end{aligned} \tag{4.1}$$

This is an alternative representation for the one presented in Chapter 3. For the sake of simplicity, we will use T_{ij} instead of $\log(T_{ij})$. Both representations are modelling the same mixture cure frailty structure.

Next, let us define a cure fraction using the incidence component

$$p_0 = \frac{1}{1 + \exp(b'x)}. \tag{4.2}$$

Using the cure fraction p_0, the complete likelihood function can be written as

$$\prod_{i=1}^{n} \prod_{j=1}^{\ell_i} (p_0)^{1-I_{ij}} \times \{(1 - p_0) h(t_{ij}|X_{ij}, y_{ij})^{\delta_{ij}} S(t_{ij}|X_{ij}, y_{ij})\}^{I_{ij}} f(y_{ij}), \tag{4.3}$$

where I_i is an indicator function with $I_i = 0$ if the patient is non-susceptible or cured and 1 otherwise.

For the frailty term, we assume the Gamma form representation of the Generalized Gamma distribution in Balakrishnan and Peng (2006), expressed as

$$f(z; p) = \frac{p^p z^{p-1} e^{-pz}}{\Gamma(p)}, \quad z > 0, \quad p > 0, \tag{4.4}$$

with

$$Y = \frac{1}{\lambda} Z^{\sigma\sqrt{p}}, \tag{4.5}$$

where Y is the Generalized Gamma variable with $q = \frac{1}{\sqrt{p}}$. The details of the transformation can be found in Section 1.3.

4.2 Estimation Procedure

Using the simplified notation, the complete likelihood function can be written as

$$\begin{aligned}
&\ell_c(\beta, \sigma, q) \propto \\
&\prod_{i=1}^{n}\prod_{j=1}^{\ell_i}(p_0)^{1-I_{ij}}(1-p_0)^{I_{ij}} \\
&\times \exp(-I_{ij}\delta_{ij}\beta' X_{ij}) {h_0}^{I_{ij}\delta_{ij}}(t_{ij}\exp(-\beta' X_{ij})) \left(\frac{{z_{ij}}^{\sigma/q}}{\lambda}\right)^{I_{ij}\delta_{ij}} \exp\left(-I_{ij}\frac{{z_{ij}}^{\sigma/q}}{\lambda} H_0(t_{ij}\exp(-\beta' X_{ij}))\right) \\
&\times \frac{(q^{-2})^{q^{-2}} {z_{ij}}^{q^{-2}-1}\exp(-q^{-2}z_{ij})}{\Gamma(q^{-2})}.
\end{aligned} \tag{4.6}$$

The logarithm of the complete likelihood functions can then be expressed as

$$\ell = \ell_{c1}(b_0, b_1) + \ell_{c2}(\sigma, q) + \ell_{c3}(\beta, \sigma, q), \tag{4.7}$$

where

$$\begin{aligned}
\ell_{c1}(b_0, b_1) &= \sum_{i=1}^{n} \sum_{j=1}^{\ell_i} (1 - I_{ij}) \log(p_0) + I_{ij} \log(1 - p_0), \\
\ell_{c2}(\sigma, q) &= \sum_{i=1}^{n} \Bigg(\sum_{j=1}^{\ell_i} I_{ij}\delta_{ij} \bigg(\frac{\sigma}{q} \log(z_{ij}) - \log(\lambda) \bigg) \\
&\quad + q^{-2} \log(q^{-2}) + (q^{-2} - 1) \log(z_{ij}) - q^{-2} z_{ij} - \log(\Gamma(q^{-2})) \Bigg), \\
\ell_{c3}(\beta, \sigma, q) &= \sum_{i=1}^{n} \sum_{j=1}^{\ell_i} -\delta_{ij} I_{ij} \beta + \delta_{ij} I_{ij} \log(h_0(t_{ij} \exp(-\beta X_{ij}))) - I_{ij} \frac{{z_{ij}}^{\sigma/q}}{\lambda} H_0(t_{ij} \exp(-\beta' X_{ij})),
\end{aligned} \tag{4.8}$$

where the frailty variable Z_{ij} is a random effect and I_{ij} is a latent indicator variable. We cannot maximize the log-likelihood function directly, and so we adopt the EM algorithm. The steps in the iteration are specified as follows:

E-Step

In the E-step, we calculate the conditional Q functions given by

$$\begin{aligned}
Q_{c1}(b_0, b_1) &= \sum_{i=1}^{n} \sum_{j=1}^{\ell_i} (1 - \mathbb{E}(I_{ij})) \log(p_0) + \mathbb{E}(I_{ij}) \log(1 - p_0), \\
Q_{c2}(\sigma, q) &= \sum_{i=1}^{n} \Bigg(\sum_{j=1}^{\ell_i} \delta_{ij} \bigg(\frac{\sigma}{q} \mathbb{E}(\log(z_{ij})) - \log(\lambda) \bigg) \\
&\quad + q^{-2} \log(q^{-2}) + (q^{-2} - 1)\mathbb{E}(\log(z_{ij})) - q^{-2}\mathbb{E}(z_{ij}) - \log(\Gamma(q^{-2})) \Bigg), \\
Q_{c3}(\beta, \sigma, q) &= \sum_{i=1}^{n} \sum_{j=1}^{\ell_i} -\delta_{ij} \beta + \delta_{ij} \log(h_0(t_{ij} \exp(-\beta X_{ij}))) - \mathbb{E}\bigg(I_{ij} \frac{{z_{ij}}^{\sigma/q}}{\lambda} \bigg) H_0(t_{ij} \exp(-\beta' X_{ij})),
\end{aligned} \tag{4.9}$$

where the conditional expectations of the unknown variables Z_{ij} and I_{ij} are given by

$$\begin{aligned}
\pi_{ij} &= \mathbb{E}(I_{ij}) = \delta_{ij} + (1-\delta_{ij})\frac{(1-p_0)\times L_{y_i}\big(H_0(t\exp(-\beta' x_{ij}))\big)}{p_0 + (1-p_0)\times L_{y_i}\big(H_0(t\exp(-\beta' x_{ij}))\big)},\\
a_{ij} &= \mathbb{E}(I_{ij}\frac{{z_{ij}}^{\sigma/q}}{\lambda}) = \mathbb{E}(\frac{{z_{ij}}^{\sigma/q}}{\lambda}|I_{ij}=1)\times\pi_{ij},\\
b_{ij} &= \mathbb{E}(z_{ij})|\Theta^{(m)},O) = \mathbb{E}(z_{ij}|I_{ij}=1,\Theta^{(m)},O)\times\pi_{ij} + \mathbb{E}(z_{ij}|I_{ij}=0,\Theta^{(m)},O)\times(1-\pi_{ij}),\\
c_{ij} &= \mathbb{E}(\log(z_{ij}))|\Theta^{(m)},O) = \mathbb{E}(\log(z_{ij})|I_{ij}=1)\times\pi_{ij} + \mathbb{E}(\log(z_{ij})|I_{ij}=0)\times(1-\pi_{ij}).
\end{aligned} \tag{4.10}$$

As the conditional distribution of Z does not have a closed form, but we can find the conditional density of $Z_{ij}(\Theta, O)$ as

$$\propto {z_{ij}}^{\sigma/q\delta_{ij}+q^{-2}-1}\exp\big(-\frac{{z_{ij}}^{\sigma/q}}{\lambda}I_iH_0(t_{ij}\exp(-\beta' X_{ij})) - q^{-2}z_{ij}\big). \tag{4.11}$$

We use the MCMC method to find the expected values in Equation (4.10). This step is similar to the approach in Chapter 2.

M-Step

In the M-Step, our goal is to maximize the conditional log-likelihood and update the unknown parameters. Because we did not specify the baseline hazard function, it is hard to evaluate the value of Q_3 in Equation (4.9).

As in Equation (3.18), we can employ a Breslow-type estimator and find an efficient estimate correspondingly. But, motivated by Zeng and Lin (2007) and Chen *et al.* (2013), we want to introduce an efficient estimator and study the MLE of the model parameters. Then, we can compare different frailty models corresponding to the special cases of Generalized Gamma frailty distribution. For this purpose, we adopt a piecewise constant hazard function, and then follow Zeng and Lin (2007)'s idea to find a kernel smoothed baseline. First, we partition the real line containing all $\exp(-\beta' X_{ij})$ into J_n equally spaced intervals, $0 \equiv t_0 < t_1 < \cdots < t_{J_n} \equiv M$, where M represents the upper bound for $\exp(-\beta' X_{ij})$ over all possible β's in a bounded set. Then, we assume a piecewise constant baseline hazard function in each interval as

$$\begin{aligned} h(t) &= \sum_{k=1}^{J_n} d_k I(t \in [t_{k-1}, t_k)), \\ H(t) &= \sum_{k=1}^{J_n} d_k (t - t_k) I(t \in [t_{k-1}, t_k)) + \frac{M}{J_n} \sum_{k=1}^{J_n} d_k I(t \geq t_k), \end{aligned} \tag{4.12}$$

where $I(.)$ is the indicator function. After partitioning the baseline, we can introduce it into the Q_3 function as

$$\begin{aligned} Q_{c3} = & \sum_{i=1}^{n} \sum_{j=1}^{\ell_i} -\delta_{ij}\beta + \sum_{k=1}^{J_n} \log(d_k) \sum_{i=1}^{n} \sum_{j=1}^{\ell_i} \delta_{ij} I\big(\exp(-\beta X_{ij}) t_{ij} \in [t_{k-1}, t_k)\big) \\ & - \mathbb{E}\bigg(I_{ij} \frac{{z_{ij}}^{\sigma/q}}{\lambda}\bigg) \sum_{k=1}^{J_n} d_k \bigg(\sum_{i=1}^{n} \sum_{j=1}^{\ell_i} (\exp(\beta' X_{ij}) t - t_k) I\big(\exp(-\beta X_{ij}) t_{ij} \in [t_{k-1}, t_k)\big) \\ & + \frac{M}{J_n} \sum_{k=1}^{J_n} I(\exp(-\beta X_{ij}) t \geq t_k) \bigg). \end{aligned} \tag{4.13}$$

Upon differentiating Q_3 with respect to d_k and letting it equal to 0, we can estimate the baseline as follows:

$$d_k = \frac{\sum_{i=1}^{n} \sum_{j=1}^{\ell_i} \delta_{ij} I\big(\exp(-\beta X_{ij}) t_{ij} \in [t_{k-1}, t_k)\big)}{\mathbb{E}(I_{ij} \frac{{z_{ij}}^{\sigma/q}}{\lambda}) \bigg(\sum_{i=1}^{n} \sum_{j=1}^{\ell_i} (\exp(\beta' X_{ij}) t - t_k) I\big(\exp(-\beta X_{ij}) t_{ij} \in [t_{k-1}, t_k)\big) + \frac{M}{J_n} \sum_{k=1}^{J_n} I(\exp(-\beta X_{ij}) t \geq t_k) \bigg)}. \tag{4.14}$$

The pertinent details of the maximazation and smoothing process are presented in the Appendix for Chapter 4. Finally, we get the expected value of log-likelihood function

Q_3 as follows:

$$\begin{aligned} Q_3 = &\sum_{i=1}^{n}\sum_{j=1}^{\ell_i}\left[\delta_{ij}\log\left(\frac{1}{n\ell_i a_n}\sum_{k=1}^{n}\sum_{l=1}^{n_i}\delta_{kl}K(\frac{r_{kl}-r_{ij}}{a_n})\right)\right] \\ &-\sum_{i=1}^{n}\sum_{j=1}^{\ell_i}\delta_{ij}\log\left[\frac{1}{n\ell_i}\sum_{k=1}^{n}\sum_{l=1}^{n_i}\int_{-\infty}^{\frac{r_{kl}-r_{ij}}{a_n}}\mathbb{E}(I_{ij}\frac{{z_{ij}}^{\sigma/q}}{\lambda})K(s)ds\right], \end{aligned} \tag{4.15}$$

where $r_{ij} = \log(t_{ij}) - \beta' X_{ij}$ and $K(S)$ is the kernel function and a_n is the bandwidth. Thus, the complete log-likelihood function can be approximated using the kernel smoothed function.

The details of kernel function and bandwidth selection can be found in Zeng and Lin (2007). We will also discuss briefly in the next section. The cumulative hazard function can be updated in each iteration as

$$\hat{H}_0(t) = \int_0^t h_0(s)ds = \int_{-\infty}^{\log(t)} \frac{\frac{1}{n\ell_i a_n}\sum_{k=1}^{n}\sum_{l=1}^{\ell_i}\delta_{kl}K(\frac{r_{kl}-s}{a_n})}{\frac{1}{n\ell_i}\sum_{k=1}^{n}\sum_{l=1}^{\ell_i}\int_{-\infty}^{\frac{r_{kl}-s}{a_n}}\mathbb{E}(I_{ij}\frac{{z_{ij}}^{\sigma/q}}{\lambda})K(u)du} ds. \tag{4.16}$$

Finally, the estimated survival function can be obtained as $\exp(-\hat{H}_0(t))$.

Estimation Procedure:

Step 1 Given the initial values of $b_{(0)}$, $\beta_{(0)}$, $\sigma_{(0)}$ and $q_{(0)}$ for the cure frailty model, $\hat{H}_0(t_{ij})$ can be estimated correspondingly;

Step 2 E-Step: Calculate the corresponding expectations in the Q function;

Step 3 M-Step: Estimate $b_{(m)}$, $\beta_{(m)}$, $\sigma_{(m)}$, $q_{(m)}$ by maximizing the Q function and update the estimation of $\hat{H}_0(t_{ij})$ using Newton-Raphson algorithm;

Step 4 Iterate Steps 2 and 3 until b, β, σ and q converge to the desired level of accuracy.

4.3 Simulation Study

The purpose of our simulation study is to evaluate the performance of the proposed estimation method. We generate 1000 simulated data sets with sizes 100 and 200. Then, we generate baseline distribution as a standard log-normal distribution. The frailty term is generated with mean 1 and variance 0.5 or 1 from Gamma, Weibull and Log-normal distributions. Finally, a singe covariate x is generated by randomly assigning the subjects into two groups with probability 0.5.

The effect of x on the incidence is $b_0 = 2$ and $b_1 = -1$ for the cure fraction. Therefore, the corresponding cure rate will be 11.9% and 26.9% respectively, in the control and treatment group. The effect of x on the latency is $\beta_1 = \log(1.5)$. The censoring time is generated from the uniform distribution to obtain a fraction of 25% or 50%.

The method of simulating the data is similar to that discussed in Section 2.3.

From the discussion on the optimal choice of the kernel function and the bandwidth by Zeng and Lin (2007), we use a standard normal distribution as the kernel density and the bandwidth $\sigma_k n^{-1/5}$, where σ_k is the sample standard deviation of $\log(T)$.

As stated above, to evaluate the performance of the model and estimation methods, we use various sample size, frailty variance and censoring proportions, and these are listed in Tables 4.1 and 4.2.

Based on Tables 4.3 - 4.6, we observe that the estimation method works well. As the censoring proportion increases, the cure coefficient b_0 and b_1 tend to have greater bias and MSE, and this is similar to situation in Section 3.4. The frailty variance $\hat{\theta}_f^2$ and the covariate's coefficient β are not influenced much by the true frailty model, which indicates that the Generalized Gamma distribution is a good frailty assumption to model random effects. In most cases, the CPs are quite close to the nominal level. In summary, larger sample size and less censoring rate lead to more accurate estimates, and that the change in frailty variance does not have a clear effect on the accuracy or precision of the parameter estimates.

	Sample size	b_0	b_1	Variance	Censoring proportion	β	True frailty
1	100	2	-1	0.5	0.25	log(1.5)	Gamma
2	100	2	-1	0.5	0.5	log(1.5)	Gamma
3	100	2	-1	1	0.25	log(1.5)	Gamma
4	100	2	-1	1	0.5	log(1.5)	Gamma
5	100	2	-1	0.5	0.25	log(1.5)	Weibull
6	100	2	-1	0.5	0.5	log(1.5)	Weibull
7	100	2	-1	1	0.25	log(1.5)	Weibull
8	100	2	-1	1	0.5	log(1.5)	Weibull
9	100	2	-1	0.5	0.25	log(1.5)	Log-normal
10	100	2	-1	0.5	0.5	log(1.5)	Log-normal
11	100	2	-1	1	0.25	log(1.5)	Log-normal
12	100	2	-1	1	0.5	log(1.5)	Log-normal

Table 4.1: Simulation Setting

	Sample size	b_0	b_1	Variance	Censoring proportion	β	True frailty
1	200	2	-1	0.5	0.25	log(1.5)	Gamma
2	200	2	-1	0.5	0.5	log(1.5)	Gamma
3	200	2	-1	1	0.25	log(1.5)	Gamma
4	200	2	-1	1	0.5	log(1.5)	Gamma
5	200	2	-1	0.5	0.25	log(1.5)	Weibull
6	200	2	-1	0.5	0.5	log(1.5)	Weibull
7	200	2	-1	1	0.25	log(1.5)	Weibull
8	200	2	-1	1	0.5	log(1.5)	Weibull
9	200	2	-1	0.5	0.25	log(1.5)	Log-normal
10	200	2	-1	0.5	0.5	log(1.5)	Log-normal
11	200	2	-1	1	0.25	log(1.5)	Log-normal
12	200	2	-1	1	0.5	log(1.5)	Log-normal

Table 4.2: Simulation Setting

Table 4.3: Simulation results with sample size 200 and true variance $\sigma_f^2 = 0.5$

True Frailty	Censoring proportion	Parameters	Bias	MSE	Coverage Probability
Gamma	25	$\hat{\sigma}_f^2$	0.017	0.091	0.951
		$\hat{b}_0$	0.115	0.215	0.942
		$\hat{b}_1$	0.096	0.224	0.957
		$\hat{\beta}_1$	-0.029	0.033	0.952
Log-normal		$\hat{\sigma}_f^2$	0.037	0.061	0.952
		$\hat{b}_0$	0.092	0.204	0.961
		$\hat{b}_1$	0.202	0.406	0.946
		$\hat{\beta}_1$	-0.026	0.025	0.958
Weibull		$\hat{\sigma}_f^2$	0.048	0.114	0.960
		$\hat{b}_0$	0.071	0.177	0.946
		$\hat{b}_1$	0.074	0.307	0.951
		$\hat{\beta}_1$	-0.033	0.046	0.949
Gamma	50	$\hat{\sigma}_f^2$	0.026	0.066	
		$\hat{b}_0$	0.310	0.399	
		$\hat{b}_1$	0.103	0.299	
		$\hat{\beta}_1$	-0.028	0.043	
Log-normal		$\hat{\sigma}_f^2$	-0.032	0.081	
		$\hat{b}_0$	0.333	0.391	
		$\hat{b}_1$	0.192	0.295	
		$\hat{\beta}_1$	-0.041	0.055	
Weibull		$\hat{\sigma}_f^2$	0.048	0.092	
		$\hat{b}_0$	0.401	0.502	
		$\hat{b}_1$	0.183	0.309	
		$\hat{\beta}_1$	-0.045	0.062	

Table 4.4: Simulation results with sample size 100 and true variance $\sigma_f^2 = 0.5$

True Frailty	Censoring proportion	Parameters	Bias	MSE
Gamma	25	$\hat{\sigma}_f^2$	0.045	0.096
		$\hat{b}_0$	0.084	0.136
		$\hat{b}_1$	0.056	0.291
		$\hat{\beta}_1$	-0.032	0.046
Log-normal		$\hat{\sigma}_f^2$	0.068	0.086
		$\hat{b}_0$	0.051	0.119
		$\hat{b}_1$	0.141	0.266
		$\hat{\beta}_1$	-0.025	0.055
Weibull		$\hat{\sigma}_f^2$	-0.039	0.112
		$\hat{b}_0$	0.130	0.165
		$\hat{b}_1$	0.063	0.179
		$\hat{\beta}_1$	0.011	0.052
Gamma	50	$\hat{\sigma}_f^2$	0.028	0.081
		$\hat{b}_0$	0.362	0.403
		$\hat{b}_1$	0.201	0.454
		$\hat{\beta}_1$	-0.018	0.039
Log-normal		$\hat{\sigma}_f^2$	0.014	0.065
		$\hat{b}_0$	0.184	0.235
		$\hat{b}_1$	0.125	0.338
		$\hat{\beta}_1$	0.013	0.030
Weibull		$\hat{\sigma}_f^2$	-0.016	0.108
		$\hat{b}_0$	0.329	0.396
		$\hat{b}_1$	0.099	0.421
		$\hat{\beta}_1$	-0.047	0.057

Table 4.5: Simulation results with sample size 200 and true variance $\sigma_f^2 = 1$

True Frailty	Censoring proportion	Parameters	Bias	MSE
Gamma	25	$\hat{\sigma}_f^2$	0.021	0.119
		$\hat{b}_0$	0.129	0.204
		$\hat{b}_1$	0.197	0.332
		$\hat{\beta}_1$	-0.022	0.041
Log-normal		$\hat{\sigma}_f^2$	0.027	0.127
		$\hat{b}_0$	0.102	0.153
		$\hat{b}_1$	-0.178	0.329
		$\hat{\beta}_1$	-0.031	0.034
Weibull		$\hat{\sigma}_f^2$	0.038	0.138
		$\hat{b}_0$	0.112	0.163
		$\hat{b}_1$	0.106	0.357
		$\hat{\beta}_1$	-0.028	0.037
Gamma	50	$\hat{\sigma}_f^2$	-0.018	0.182
		$\hat{b}_0$	0.266	0.304
		$\hat{b}_1$	0.172	0.282
		$\hat{\beta}_1$	0.020	0.035
Log-normal		$\hat{\sigma}_f^2$	0.045	0.124
		$\hat{b}_0$	0.339	0.347
		$\hat{b}_1$	0.169	0.406
		$\hat{\beta}_1$	-0.036	0.056
Weibull		$\hat{\sigma}_f^2$	0.098	0.093
		$\hat{b}_0$	0.313	0.371
		$\hat{b}_1$	0.114	0.367
		$\hat{\beta}_1$	-0.025	0.042

Table 4.6: Simulation results with sample size 100 and true variance $\sigma_f^2 = 1$

True Frailty	Censoring proportion	Parameters	Bias	MSE
Gamma	25	$\hat{\sigma}_f^2$	0.055	0.104
		$\hat{b}_0$	0.143	0.252
		$\hat{b}_1$	-0.031	0.177
		$\hat{\beta}_1$	-0.039	0.036
Log-normal		$\hat{\sigma}_f^2$	0.098	0.108
		$\hat{b}_0$	0.072	0.116
		$\hat{b}_1$	0.050	0.291
		$\hat{\beta}_1$	-0.031	0.056
Weibull		$\hat{\sigma}_f^2$	-0.033	0.071
		$\hat{b}_0$	0.095	0.129
		$\hat{b}_1$	-0.026	0.322
		$\hat{\beta}_1$	0.015	0.023
Gamma	50	$\hat{\sigma}_f^2$	0.050	0.127
		$\hat{b}_0$	0.246	0.301
		$\hat{b}_1$	0.207	0.336
		$\hat{\beta}_1$	-0.035	0.041
Log-normal		$\hat{\sigma}_f^2$	-0.016	0.142
		$\hat{b}_0$	0.121	0.223
		$\hat{b}_1$	0.228	0.375
		$\hat{\beta}_1$	-0.033	0.049
Weibull		$\hat{\sigma}_f^2$	-0.039	0.119
		$\hat{b}_0$	0.312	0.352
		$\hat{b}_1$	0.101	0.343
		$\hat{\beta}_1$	-0.036	0.066

4.4 Model Selection and Discrimination

Model discrimination is carried out in this section to examine further the performance of the developed model and the estimation method. Based on the simulation setting described in the last section, the four mixture cure AFT frailty models with Gamma, Log-normal, Weibull and Generalized Gamma distribution are fitted to each data set.

True model	Fitted Model			
	Gamma	Log-normal	Weibull	Generalized Gamma
Gamma	0.29	0.25	0.18	0.21
Log-normal	0.22	0.29	0.20	0.29
Weibull	0.26	0.27	0.23	0.24

Table 4.7: Selection rates based on Akaike information criterion (Sample size 100, frailty variance 1 and censoring rate 0.25)

True model	Fitted Model			
	Gamma	Log-normal	Weibull	Generalized Gamma
Gamma	0.35	0.21	0.13	0.31
Log-normal	0.26	0.30	0.15	0.29
Weibull	0.25	0.20	0.24	0.31

Table 4.8: Selection rates based on Akaike information criterion (Sample size 200, frailty variance 1 and censoring rate 0.25)

Tables 4.7 and 4.8 present the selection rates of models based on the Akaike information criterion. The selection rates of the correct models increase as the sample size increases. The Generalized Gamma frailty distribution performs better in most cases when the sample size is large enough.

4.5 Application to Bone Marrow Transplant Data

As an application of the proposed model, we fit the bone marrow transplant study for leukaemia patients with the mixture cure AFT frailty model.
Price and Manatunga (2001) considered this bone marrow transplant study for the leukaemia patients. Leukaemia patients received either an allogeneic transplant or an autologous transplant. The maximum followed up time is 1845 days and the time to recurrence or censoring is recorded. The details of the data set have been described earlier in Section 1.4.

	Gamma		Log-normal		Weibull		GG	
	Estimate	Var	Estimate	Var	Estimate	Var	Estimate	Var
$\hat{\beta}_1$	-0.356	0.288	-0.361	0.271	-0.392	0.446	-0.362	0.303
b_0	1.008	0.341	1.018	0.334	0.989	0.330	1.009	0.320
b_1	0.448	0.497	0.447	0.488	0.416	0.539	0.423	0.481
σ							1.187	0.467
q							0.982	0.294
$\hat{\theta}$	1.623	0.440						
$\hat{\sigma}_f^2$	0.616		0.668		0.626		0.814	
ℓ	-199.02		-195.27		-197.39		-191.87	
AIC	406.04		398.54		402.78		393.74	

Table 4.9: Estimated parameters from the mixture cure AFT Generalized Gamma frailty model for bone marrow transplant data

Note: The estimated variances (Var) are from the bootstrap method based on 500 bootstrap samples. The estimated frailty variance ($\hat{\sigma}_f^2$) denotes the variance calculated from the estimates.

Table 4.7 displays the estimation results of the fits of the bone marrow transplant data set. As done by Price and Manatunga (2001), we consider a nested model and are interested in testing the presence of immune fraction. This requires the comparison of the mixture cure AFT frailty model and the AFT frailty model by Chen *et al.* (2013). For this purpose, we can employ the Likelihood Ratio Test (LRT). Furthermore, since the parameter is at the boundary of the parameter space, we need to consider the boundary condition of LRT, as discussed in Maller and Zhou (1996). They suggested the large sample distribution of $-2\log(\hat{\ell}_0 - \hat{\ell})$ as a 50 - 50 mixture of a chi-square random variable with 1 degree of freedom and a point mass at 0.

Then, to formally test the presence of immunes in the data set, the null hypothesis H_0: $p_0 = 0$ is tested. The 95th percentile of the distribution of $-2\log(\Lambda)$ is given by $\frac{1}{2} + \frac{1}{2}P(\chi^2 \leq \chi^{2^*}) = 0.95$. Based on χ^2-table, the critical value is χ^{2^*} =2.71. Because $-2\log(\hat{\ell}_0 - \hat{\ell}) = -2\log(-180.03 - 191.87) = 23.68 > 2.71$, the null hypothesis is not supported at the 5% significance level. So, we can conclude that there exists strong evidence of a cured proportion.

Further, the cure proportions for the allogeneic transplant group are $1/(1+\exp(1.0085)) = 0.2673$, $1/(1 + \exp(1.0183)) = 0.2654$, $1/(1 + \exp(1.0.9894)) = 0.2710$ and $1/(1 + \exp(1.0094)) = 0.2671$, respectively. On the other hand, the cure proportions for the autologous transplant group are $1/(1 + \exp(1.0085 + 0.4476)) = 0.1891$, $1/(1 + \exp(1.0183 + 0.4465)) = 0.1877$, $1/(1 + \exp(0.9894 + 0.4161)) = 0.1969$ and $1/(1 + \exp(1.0094 + 0.423)) = 0.1928$. These results are consistent with those of Peng and Zhang (2008a) and Zhang and Peng (2007a).

Chapter 5

Concluding Remarks

5.1 Conclusions

In clinical studies, survival data do not often follow the assumption of proportional hazards. One of the common approaches in such situations is to employ the frailty model with proportion hazard (PH) assumption or to employ the accelerated failure time (AFT) model to fit the hazard. One potential issue is that commonly used frailty distributions, such as Gamma, Log-normal, Weibull and Inverse-Gaussian distributions, are not flexible enough to capture the heterogeneity among individual subjects or groups. Moreover, the mixture cure accelerated time model with frailty term has not been discussed in the literature. For these reason, we have introduced the mixture cure Generalized Gamma frailty models under proportional hazard and accelerated failure time assumptions.

In Chapter 2, with the aim of proposing a general mixture cure frailty model that can describe the random effects among individuals in a flexible way, we have introduced a mixture cure Generalized Gamma frailty model. Due to the complexity in the structure of the model and in not specifying the underling baseline distribution, we propose an EM-type algorithm to develop semi-parametric inference for the model parameters. Moreover, the expectations in the E-step do not have a closed form and so Markov chain Monte Carlo (MCMC) approach is needed. As we optimize not by maximizing the log-likelihood function directly, the Fisher information can not be obtained and so bootstrap variance estimation is performed. This is also the variance estimation approach suggested in the literature. The model and estimation procedure provide satisfactory results in both simulation study and real-life data analysis. Based on the results obtained, we observe that the common frailty distributions usually underestimate the frailty variance, and so a more flexible frailty distribution model will be very useful.

In Chapter 3, for the purpose of establishing a mixture cure frailty model with an accelerated failure time model, we introduce a mixture cure AFT Gamma frailty model. Due to the convenience of Gamma distribution as a frailty model, we can find a closed-form expression for the conditional frailty variable. This simplifies the estimation procedure significantly. The semi-parametric approach implements a Gehan-type weight function resulting in a monotone function of parameters, and so the required computation can be carried out through linear programming methods. The proposed model is fitted to the bone marrow transplant data which demonstrates that the mixture cure accelerated failure time frailty model is a suitable alternative to the usual mixture cure frailty model.

In Chapter 4, motivated by the work of Zeng and Lin (2007) and to improve the estimation methods developed in Chapters 2 and 3, we employ the normal kernel smoothing methods to develop an efficient estimation method. As in the case of the usual proportional hazard model, EM algorithm is needed for estimation and the MCMC approach is utilized. By assuming infinite partitions of the time interval, we can find estimates of the Q function by a kernel density with certain bandwidth. Thus, in the M-step, the Newton-Raphson method can be applied, and we can thus find an efficient estimator. Then, we perform model discrimination based on Akaike information criterion and Likelihood Ratio Test (LRT) for the purpose of comparing various mixture cure AFT frailty models and for testing the necessity of the cure assumption, respectively. The results show that the Generalized Gamma distribution performs better than the common frailty distributions used and the cure fraction is always significant under the accelerated failure time assumption.

5.2 Possible Future Directions

In the present work, we have assumed the cure probability and the random effect of individual subjects to be independent of each other. This is only for the purpose of simplifying the computation of effort; so, some further correlation analysis can be performed for the mixture cure frailty model to relax this assumption made.

Also, limited by the structure of the mixture cure frailty itself, it is not possible to consider frailty shared by subjects from the same cluster. There may be some other ways to introduce a similar structure to propose a mixture cure clustered frailty model.

In this work, we have developed efficient estimation of the use of a kernel smoothed function. There are other ways to produce estimators with good properties, such as Box-cox transformation, multiple imputation, numerical quadrature and Bayesian analysis. Some estimation methods along these lines may be used to develop for the mixture cure AFT frailty model.

Chapter 6

Appendix

6.1 Results for Special Cased of Generalized Gamma Distribution

As specified in (1.21), when the parameter $q \to 0$, the generalized gamma distribution will tends to the Log-normal distribution. When we can set $k = q^{-2}$ and $z = \frac{\log(\lambda w)}{\sigma}$, the limiting density function can be obtained as:

$$\lim_{q \to 0} g(w|q,\sigma,\lambda) = |q|(q^{-2})^{q^{-2}}(\lambda w)^{q^{-2}(q/\sigma)} \exp\big(-q^{-2}(\lambda w)^{q/\sigma}\big/\Gamma(q^{-2}\sigma w). \tag{6.1}$$

If we rearrange the above expression, the density is given by

$$\begin{aligned} \lim_{k \to \infty} g(w|k) &= \frac{k^{k-1/2}}{\Gamma(k)} \exp(\left(\sqrt{k}z - k\exp(z/\sqrt{k})\right))\frac{1}{\sigma w} \\ &= \exp\left[\left(k - \frac{1}{2}\right)\log(k) - \log\Gamma(k)\right]\exp\left(\sqrt{k}z - ke^{z/\sqrt{k}}\right)\frac{1}{\sigma w}. \end{aligned} \tag{6.2}$$

Based on the property of Log-Gamma function described by Prentice (1977) and Lawless (1980), the function can be expressed as

$$\begin{aligned}\log(\Gamma(k)) &= (k-\frac{1}{2})\log(k) - k + \frac{1}{2}\log(2\pi) + \frac{1}{12} + O(k^{-3}),\\ (k-\frac{1}{2})\log(k) - \log(\Gamma(k)) &= k - \frac{1}{2}\log(2\pi) - \frac{1}{12} + O(k^{-3}).\end{aligned} \tag{6.3}$$

Thus, as $k \to \infty$, the limiting density function is given by

$$\begin{aligned}\lim_{k\to\infty} g(w|k) &= \exp\left(k - \frac{1}{2}\log(2\pi) - \frac{1}{12k}\right)\exp\left(\sqrt{k}z - ke^{z/\sqrt{k}}\right)\frac{1}{\sigma w}\\ &= \exp\left(k - \frac{1}{2}\log(2\pi) - \frac{1}{12k} + \sqrt{k}z - ke^{z/\sqrt{k}}\right)\\ &= \frac{1}{\sqrt{2\pi}}\exp\left(k(1-e^{z/\sqrt{k}}) + \sqrt{k}z - \frac{1}{12k}\right)\\ &= \frac{1}{\sqrt{2\pi}\sigma w}\exp\left(-\frac{z^2}{2}\right)\\ &= \frac{1}{\sqrt{2\pi}\sigma w}\exp\left(-\frac{\log(\lambda w)}{2\sigma^2}\right).\end{aligned} \tag{6.4}$$

This is consistent with the result in Equation (1.21); See also Balakrishnan and Peng (2006) and Chen *et al.* (2013).

6.2 Details of smoothing Q_3 in Chapter 4

Based on Zeng and Lin (2007) and Chen *et al.* (2013), the kernel smoothed approximation of Q_3 can be calculated as follows.

Suppose that $n \times \ell = N$. When $N \to \infty$, $J_n \to \infty$, and $\frac{J_n}{N} \Rightarrow 0$, according to the Donsker theorem, we can obtain the following limits:

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{\ell}\delta_{ij}\beta' X_{ij} \to \mathbb{E}(\delta\beta' X),\\
&\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{\ell}\delta_{ij} I(e^{-\beta' X} t_{ij} \in [t_{k-1}, t_k)) \to P(\delta = 1, e^{-\beta' X} t \in [t_{k-1}, t_k)),\\
&\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{\ell}\delta_{ij} I(e^{-\beta' X} t_{ij} \in [t_{k-1}, t_k)) \to \frac{dP(\delta = 1, e^{-\beta' X} t \le s)}{ds}.
\end{aligned} \tag{6.1}$$

Definition: If for every $\epsilon > 0$, there exist constants C_ϵ and n_ϵ such that $P(|X_n| \le C_\epsilon a_n) > 1 - \epsilon$ for every $n \ge n_\epsilon$, then we say X_n is $O_p(a_n)$.

Here, we can apply the multiple central limit theorem in Van Der Vaart and Wellner (1996), and employ the maximization within a limit as follows:

$$\begin{aligned}
&\max\Bigg|\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{\ell}\mathbb{E}\bigg(I_{ij}\frac{{z_{ij}}^{\sigma/q}}{\lambda}\bigg)(t_{ij}e^{-\beta' X_{ij}} - t_k) I(t_{k-1} \le t_{ij}e^{-\beta' X_{ij}} \le t_k)\\
&- \mathbb{E}\Bigg(\mathbb{E}\bigg(I_{ij}\frac{{z_{ij}}^{\sigma/q}}{\lambda}\bigg)(t_{ij}e^{-\beta' X_{ij}} - t_k) I(e^{-\beta' X} t \in [t_{k-1}, t_k))\Bigg)\Bigg| = O_p\bigg(\frac{1}{\sqrt{N}}\bigg).
\end{aligned} \tag{6.2}$$

In other words, Q_3 is bounded, up to an exceptional event of arbitrarily small (but fixed) positive probability.

Furthermore, the maximization can be simplified as

$$\max\left|\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{\ell_u}\mathbb{E}\bigg(I_{ij}\frac{{z_{ij}}^{\sigma/q}}{\lambda}\bigg)I(t_{ij}e^{-\beta' X_{ij}}\geq t_k)-\mathbb{E}\bigg[\mathbb{E}\bigg(I_{ij}\frac{{z_{ij}}^{\sigma/q}}{\lambda}\bigg)I(te^{-\beta' X_{ij}}\geq t_k)\bigg]\right| = O_p(\frac{1}{\sqrt{N}}). \tag{6.3}$$

According to the assumption that $\frac{J_n}{N}\to 0$, the equation will converge uniformly in β and t_k.

Finally, upon choosing a kernel function $K(.)$ with bandwidth a_n, under suitable regularity conditions, Q_3 can be obtained as

$$\begin{aligned} Q_3 = \sum_{i=1}^{n}\sum_{j=1}^{\ell_i}\bigg[&\delta_{ij}\log\bigg(\frac{1}{n\ell_i a_n}\sum_{k=1}^{n}\sum_{l=1}^{n_i}\delta_{kl}K(\frac{r_{kl}-r_{ij}}{a_n})\bigg)\\ &-\delta_{ij}\log\big(\frac{1}{n\ell_i}\sum_{k=1}^{n}\sum_{l=1}^{n_i}\int_{-\infty}^{\frac{r_{kl}-r_{ij}}{a_n}}\mathbb{E}(I_{ij}\frac{{z_{ij}}^{\sigma/q}}{\lambda})K(s)ds\big)\bigg], \end{aligned} \tag{6.4}$$

where

$$r_{ij} = \log(t_{ij}) - \beta' X_{ij}, \tag{6.5}$$

$K(S)$ is the kernel function and a_n is the bandwidth. Thus, the complete log-likelihood function can be approximated by using the kernel smoothed function. Details of the proof and the asymptotic property can be found in Zeng and Lin (2007).

Bibliography

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical Models based on Counting Processes*. Springer, New York.

Balakrishnan, N. and Pal, S. (2014). COM-Poisson cure rate models and associated likelihood-based inference with Exponential and Weibull lifetimes. In *Applied Reliability Engineering and Risk Analysis*, page 308. John Wiley & Sons, New York.

Balakrishnan, N. and Pal, S. (2016). Expectation maximization-based likelihood inference for flexible cure rate models with Weibull lifetimes. *Statistical Methods in Medical Research*, **25**(4), 1535–1563.

Balakrishnan, N. and Peng, Y. (2006). Generalized gamma frailty model. *Statistics in Medicine*, **25**(16), 2797–2816.

Balakrishnan, N., Barui, S., and Milienos, F. (2017). Proportional hazards under Conway-Maxwell-Poisson cure rate model and associated inference. *Statistical Methods in Medical Research*, **25**(13), 2055.

Berkson, J. and Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, **47**(259), 501–515.

Boag, J. W. (1948). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society, Series B*, **11**(1), 15–53.

Breslow, N. E. (1972). Contribution to discussion of paper by Dr. Cox. *Journal of Royal Statistical Society, Series B*, **34**, 216–217.

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika*, **66**(3), 429–436.

Cai, C., Zou, Y., Peng, Y., and Zhang, J. (2012). smcure: An R-package for estimating semiparametric mixture cure models. *Computer Methods and Programs in Biomedicine*, **108**(3), 1255–1260.

Chatterjee, N. and Shih, J. (2001). A bivariate cure-mixture approach for modeling familial association in diseases. *Biometrics*, **57**(3), 779–786.

Chen, P., Zhang, J., and Zhang, R. (2013). Estimation of the accelerated failure time frailty model under generalized gamma frailty. *Computational Statistics & Data Analysis*, **62**, 171–180.

Choi, S., Zhu, L., and Huang, X. (2018). Semiparametric accelerated failure time cure rate mixture models with competing risks. *Statistics in Medicine*, **37**(1), 48–59.

Cox (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*, **34**(2), 187–220.

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data.* Chapman & Hall, London.

Diao, G. and Yin, G. (2012). A general transformation class of semiparametric cure rate frailty models. *Annals of the Institute of Statistical Mathematics*, **64**(5), 959–989.

Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, **38**(4), 1041–1046.

Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, **94**(446), 496–509.

Fygenson, M. and Ritov, Y. (1994). Monotone estimating equations for censored data. *The Annals of Statistics*, **22**(2), 732–746.

Gelman, A., Roberts, G. O., and Gilks, W. R. (1996). Efficient metropolis jumping rules. *Bayesian Statistics*, (5), 599–607.

Hougaard, P. (1989). Fitting a multivariate failure time distribution. *IEEE Transactions on Reliability*, **38**(4), 444–448.

Jin, Z., Lin, D., and Ying, Z. (2006a). On least-squares regression with censored data. *Biometrika*, **93**(1), 147–161.

Jin, Z., Lin, D., and Ying, Z. (2006b). Rank regression analysis of multivariate failure time data based on marginal linear models. *Scandinavian Journal of Statistics*, **33**(1), 1–23.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions*. John Wiley & Sons, New York.

Kalbfleisch, J. D. and Prentice, R. L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika*, **60**(2), 267–278.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**(282), 457–481.

Kersey, J. H., Weisdorf, D., Nesbit, M. E., LeBien, T. W., Woods, W. G., McGlave, P. B., Kim, T., Vallera, D. A., Goldman, A. I., Bostrom, B., *et al.* (1987). Comparison of autologous and allogeneic bone marrow transplantation for treatment of high-risk refractory acute lymphoblastic leukemia. *New England Journal of Medicine*, **317**(8), 461–467.

Klein, J. P. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics*, **48**(3), 795–806.

Kuk, A. Y. and Chen, C.-H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, **79**(3), 531–541.

Laska, E. M. and Meisner, M. J. (1993). Nonparametric estimation and testing in a cure model. *Biometrics*, **48**(4), 1223–1234.

Lawless, J. F. (1980). Inference in the Generalized Gamma and Log Gamma distributions. *Technometrics*, **22**(3), 409–419.

Li, C.-S. and Taylor, J. M. (2002). A semi-parametric accelerated failure time cure model. *Statistics in Medicine*, **21**(21), 3235–3247.

Liu, B., Lu, W., and Zhang, J. (2013). Kernel smoothed profile likelihood estimation in the accelerated failure time frailty model for clustered survival data. *Biometrika*, **100**(3), 741–755.

Longini, I. M. and Halloran, M. E. (1996). A frailty mixture model for estimating vaccine efficacy. *Journal of the Royal Statistical Society, Series C*, **45**(2), 165–173.

Lu, W. (2010). Efficient estimation for an accelerated failure time model with a cure fraction. *Statistica Sinica*, **20**(2), 661.

Lu, W. and Ying, Z. (2004). On semiparametric transformation cure models. *Biometrika*, **91**(2), 331–343.

Maller, R. A. and Zhou, X. (1996). *Survival Analysis with Long-Term Survivors*. John Wiley & Sons, New York.

McGilchrist, C. and Aisbett, C. (1991). Regression with frailty in survival analysis. *Biometrics*, **47**(2), 461–466.

Nelson, W. (1969). Hazard plotting for incomplete failure data. *Journal of Quality Technology*, **1**(1), 27–52.

Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, **14**(4), 945–966.

Pal, S. and Balakrishnan, N. (2017). Likelihood inference for COM-Poisson cure rate model with interval-censored data and weibull lifetimes. *Statistical Methods in Medical Research*, pages 2093–2113.

Pan, W. (2001). Using frailties in the accelerated failure time model. *Lifetime Data Analysis*, **7**(1), 55–64.

Peng, Y. (2003). Estimating baseline distribution in proportional hazards cure models. *Computational Statistics & Data Analysis*, **42**(1-2), 187–201.

Peng, Y. and Dear, K. B. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics*, **56**(1), 237–243.

Peng, Y. and Zhang, J. (2008a). Estimation method of the semiparametric mixture cure gamma frailty model. *Statistics in Medicine*, **27**(25), 5177–5194.

Peng, Y. and Zhang, J. (2008b). Identifiability of a mixture cure frailty model. *Statistics & Probability Letters*, **78**(16), 2604–2608.

Peng, Y., Dear, K. B., and Denham, J. (1998). A generalized F mixture model for cure rate estimation. *Statistics in Medicine*, **17**(8), 813–830.

Peng, Y., Dear, K. B., and Carriere, K. (2001). Testing for the presence of cured patients: a simulation study. *Statistics in Medicine*, **20**(12), 1783–1796.

Prentice, F. R. L. (1977). A study of distributional shape in life testing. *Technometrics*, **19**(1), 69–75.

Price, D. L. and Manatunga, A. K. (2001). Modelling survival data with a cured fraction using frailty models. *Statistics in Medicine*, **20**(9-10), 1515–1527.

Reid, N. (1994). A conversation with Sir David Cox. *Statistical Science*, **9**(3), 439–455.

Stacy, E. W. (1962). A generalization of the Gamma distribution. *The Annals of Mathematical Statistics*, **33**(3), 1187–1192.

Sy, J. P. and Taylor, J. M. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics*, **56**(1), 227–236.

Van Der Vaart, A. W. and Wellner, J. A. (1996). Weak convergence. In *Weak convergence and empirical processes*, pages 16–28. Springer, New York.

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

Wei, L.-J. (1992). The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine*, **11**(14-15), 1871–1879.

Wienke, A., Lichtenstein, P., and Yashin, A. I. (2003). A bivariate frailty model with a cure fraction for modeling familial correlations in diseases. *Biometrics*, **59**(4), 1178–1183.

Wienke, A., Locatelli, I., and Yashin, A. I. (2006). The modelling of a cure fraction in bivariate time-to-event data. *Austrian Journal of Statistics*, **35**(1), 67–76.

Xu, L. and Zhang, J. (2010). An EM-like algorithm for the semiparametric accelerated failure time gamma frailty model. *Computational Statistics & Data Analysis*, **54**(6), 1467–1474.

Yamaguchi, K. (1992). Accelerated failure-time regression models with a regression model of surviving fraction: an application to the analysis of permanent employment in Japan. *Journal of the American Statistical Association*, **87**(418), 284–292.

Yin, G. (2005). Bayesian cure rate frailty models with application to a root canal therapy study. *Biometrics*, **61**(2), 552–558.

Zeng, D. and Lin, D. (2007). Efficient estimation for the accelerated failure time model. *Journal of the American Statistical Association*, **102**(480), 1387–1396.

Zhang, J. and Peng, Y. (2007a). An alternative estimation method for the accelerated failure time frailty model. *Computational Statistics & Data Analysis*, **51**(9), 4413–4423.

Zhang, J. and Peng, Y. (2007b). A new estimation method for the semiparametric accelerated failure time mixture cure model. *Statistics in Medicine*, **26**(16), 3157–3171.

Zhang, J. and Peng, Y. (2009). Accelerated hazards mixture cure model. *Lifetime Data Analysis*, **15**(4), 455–467.