***PARAMETRIC SURVIVAL MODELS WITH TIME-DEPENDENT COVARIATE FOR MIXED CASE INTERVAL-CENSORED DATA***

**KAVEH KIANI**

# PARAMETRIC SURVIVAL MODELS WITH TIME-DEPENDENT COVARIATE FOR MIXED CASE INTERVAL-CENSORED DATA

By

**KAVEH KIANI**

**Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy**

**May 2012**

## DEDICATIONS

*To*

*My parents*

*Who have supported me all the way since the beginning of my life*

*My parents-in-law*

*For their great encouragement and support*

*My wife, Marjan*

*Who has been a great source of love and motivation*

*My daughter, Niki*

*For her great patience*

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

**PARAMETRIC SURVIVAL MODELS WITH TIME-DEPENDENT COVARIATE FOR MIXED CASE INTERVAL-CENSORED DATA**

By

**KAVEH KIANI**

**May 2012**

**Chair: Jayanthi Arasan, PhD**

**Faculty: Faculty of Science**

The aim of this research is to analyze parametric survival models in the presence of left, right, interval and doubly interval censored data with time-dependent covariates. In this research we utilize and extend two important parametric survival models, the Gompertz and the exponential, to accommodate these censoring schemes and time-dependent covariates.

The analysis starts with the extension of the Gompertz model to incorporate time-dependent covariates in the presence of right-censored data. Then, the performance of the model is compared with the fixed covariate model. Following that, comparison is made when a fixed covariate model was fitted wrongly to a data set with time-dependent covariate. In addition, two methods of constructing confidence intervals, the Wald and jackknife are explored for the parameters of this model. Conclusions are drawn based on the coverage probability study.

In the next step, the Gompertz model is further extended to incorporate time-

dependent covariates with left, right and interval censored data as well as uncensored data. The model is then investigated thoroughly at dependent and independent covariate levels through a comprehensive simulation study. Following that, the model is compared with a fixed covariate model. Then, two methods of constructing confidence intervals the Wald and likelihood ratio are investigated for the parameters of the model and conclusions are drawn based on the coverage probability study.

Finally, a parametric survival model that accommodates doubly interval-censored data with time-dependent covariates is developed and studied. In order to formulate this censoring scheme let $V$ and $W$ be the times of two related consecutive events where both of them are interval-censored and $V \leq W$. Then, the survival time of interest could be defined as, $T = W - V$. Here it is assumed that the time to the first event, $V$, and the survival time, $T$, follow the exponential distribution (special case for Gompertz distribution).

In order to get to this final model, we had to explore three separate models in advance. Firstly, a simple model consisting doubly interval-censored data without any covariate was studied. Following that, a model with doubly interval-censored data and fixed covariates is considered. Lastly, a model with fixed covariates is studied where some of the covariates affect $T$ and the others affect $V$. All these models are studied by the simulation study and two methods of constructing confidence intervals, the Wald and jackknife are explored for the parameters of the models.

The results indicate that the Gompertz model with left, right and interval censored data with a time-dependent covariate works rather well despite its complexity. Similarly, although doubly interval-censored data with a time-dependent covariate requires more computational effort, the model will perform well if both $V$ and $T$ are exponentially distributed.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putran Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

**MODEL MANDIRIAN BERPARAMETER DENGAN KOVARIAT YANG BERSANDAR KEPADA MASA UNTUK KES DATA TERTAPIS SELANG BERCAMPUR**

Oleh

**KAVEH KIANI**

**May 2012**

**Pengerusi: Jayanthi Arasan, PhD**

**Fakulti: Fakulti Sains**

Matlamat kajian ini adalah untuk menganalisis model mandarian berparameter dengan data tertapis kiri, kanan, selang dan selang berganda dengan kovariat yang bersandar kepada masa. Dalam kajian ini, kita mengguna pakai dan melanjutkan dua model mandirian berparameter yang penting, *Gompertz* dan eksponen, untuk menampung skim penapisan tersebut dan kovariat yang bersandar kepada masa. Kajian bermula dengan lanjutan model *Gompertz* untuk mengambilkira kovariat-kovariat yang bersandar kepada masa dengan data tertapis kanan. Kemudian, prestasi model ini dibandingkan dengan dengan model berkovariat tetap. Seterusnya perbandingan dilakukan apabila model berkovariat tetap yang digunakan dan bukannya model sebenar dengan kovariat yang bersandar kepada masa. Di samping itu, dua kaedah membina selang keyakinan, iaitu *Wald* dan *jackknife* telah dikaji bagi parameter-parameter model ini. Kesimpulan telah dibuat berdasarkan kajian kebarangkalian liputan.

Dalam langkah seterusnya, model *Gompertz* telah dilanjutkan lagi untuk mengambilkira kovariat-kovariat yang bersandar kepada masa dengan data tertapis kiri,

kanan dan selang disamping data tidak tertapis. Model ini kemudiannya telah dikaji secara menyeluruh pada tahap kovariat yang bersandar dan tidak bersandar melalui kajian simulasi yang komperhensif. Berikutan itu, model ini dibandingkan dengan model berkovariat tetap. Kemudian, dua kaedah membina selang keyakinan iaitu *Wald* dan kemungkinan nisbah dikaji bagi parameter-parameter model ini dan kesimpulan telah dibuat berdasarkan kajian kebarangkalian liputan.

Akhirnya, model mandirian berparameter yang mengambilkira data tertapis selang berganda dengan kovariat-kovariat yang bersandar kepada masa telah dikira dan dikaji. Bagi tujuan untuk memformulasikan skim data tertapis ini, biar $V$ dan $W$ menjadi masa-masa bagi kejadian berturutan yang berkaitan dimana kedua-duanya adalah data tertapis selang dan $V \leq W$. Kemudian, masa mandirian yang dikehendaki boleh di definasikan sebagai, $T = W - V$. Disini, adalah dianggap bahawa masa ke kejadian yang pertama, $V$ dan masa mandirian, $T$, bertaburan eksponen (kes khas bagi *Gompertz*).

Untuk mencapai model terakhir, kita terpaksa mengkaji tiga model yang berasingan dahulu. Pertama sekali, model ringkas yang terdiri daripada data tertapis selang berganda tanpa sebarang kovariat telah dikaji. Seterusnya, model dengan data tertapis selang berganda dan satu kovariat tetap telah diambilkira. Akhir sekali, model dengan kovariat-kovariat tetap telah dikaji dimana sebahagian daripada kovariat tersebut memberi kesan kepada $V$. Kesemua model-model ini telah dikaji dengan menggunakan kajian simulasi dan dua kaedah pembinaan selang keyakinan, *Wald* dan *jackknife* telah dikaji bagi parameter-parameter model ini. Keputusan menunjukkan bahawa model *Gompertz* dengan data tertapis kiri, kanan dan selang dangan kovariat yang bersandar kepada masa berfungsi dengan agak baik walaupun ianya kompleks. Begitu juga, walaupun data tertapis selang yang bersandar kepada masa memerlukan usaha komputasi yang lebih, model tersebut berfungsi dengan baik apabila kedua dua $V$ dan $T$ bertaburan eksponen

.

## ACKNOWLEDGEMENTS

First and foremost, I am grateful to Allah s.w.t for providing an opportunity for me to continue my studies in Malaysia.

I am extremely indebted to my supervisor, Dr. Jayanthi Arasan for her excellent supervision, invaluable guidance, patience and financial support via grant for the past three years. Her charisma and enthusiasm during these years has made these challenging life period a very enjoyable and useful experience. She is the one always leading me to keep faith during some difficult times. She really deserves special recognition because without her help, this work would not be accomplished. I have learned immensely from her not only in survival analysis but also in various other aspects of life. She has helped me so much in so many ways that I do have not enough words to express them.

Special thanks and appreciation goes to Prof. Dr. Habshah Midi as a member of supervisory committee for her helpful comments, suggestions and cooperation. I would like to thank Dr. Bashar Abdul Aziz Majeed Al-Talib being as a member of supervisory committee for his cooperation.

I would like to thank all members of Institute for Mathematical Research (INSPEM) and Dept. of Mathematics for all their friendly support. This particularly goes to Prof. Dato Dr. Kamel Arifin M. Atan, Director of INSPEM, Prof. Dr. Noor Akma Ibrahim, Deputy Director of INSPEM and Assoc. Prof. Dr. Mahenderan Shitan, Head of Laboratory of Computational Statistics & Operations Research, INSPEM, for their great effort to provide a unique academic environment for the graduate students. I am highly grateful to the Universiti Putra Malaysia (UPM) for all the fruitful years of my study that left an enduring positive impression on my life and professional development.

I would like to express my sincere appreciation to all who have helped and supported me and contributed in many ways big and small to the success of this research.

My special thanks goes to my grandmothers (late Behjat Javadi and Bozorg Kiani), my beloved parents (Mr. Jebraeel Kiani and Mrs. Parvin Rumi), my special parents-in-law (Mr. Mohammadhosein Shahverdi and Mrs. Ameneh Ghasemi), my dear brothers (Kambiz and Kamal) and their wives (Talayeh and Nasim), my sister-in-law's family (Sara and Reza), my brother-in-law's family (Masoud and Maryam), for their love, encouragement, support and all their favours.

Finally my deep gratitude goes to my beloved wife Marjan, without her patience, sacrifice and moral support this thesis would not have been possible. During this journey, she also struggled for her PhD program and simultaneously, carried out overwhelming motherly obligations. I am also grateful to my sweet daughter Niki, she has filled our life with joie de vivre.

I certify that a Thesis Examination Committee has met on **May 2012** to conduct the final examination of **Kaveh Kiani** on his thesis entitled **"Parametric Survival Models with Time-Dependent Covariate for Mixed Case Interval-Censored Data"** in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the **Doctor of Philosophy**.

Members of the Thesis Examination Committee were as follows:

**Isthrinayagy a/p S. Krishnarajah, PhD**
Senior Lecturer
Faculty of Science, Department of Mathematics
Universiti Putra Malaysia
(Chairman)

**Noor Akma Ibrahim, PhD**
Professor
Faculty of Science, Department of Mathematics
Universiti Putra Malaysia
(Internal Examiner)

**Mohd Bakri Adam, PhD**
Senior Lecturer
Faculty of Science, Department of Mathematics
Universiti Putra Malaysia
(Internal Examiner)

**Muhammad Hanif Mian, PhD**
Professor
Department of Mathematics, School of Science and Engineering
Lahore University of Management Sciences
Pakistan
(External Examiner)

---

**SEOW HENG FONG, PhD**
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of **Doctor of Philosophy**. The members of the Supervisory Committee were as follows:

**Jayanthi Arasan, PhD**
Senior Lecturer
Faculty of Science, Department of Mathematics
Universiti Putra Malaysia
(Chairman)

**Habshah Midi, PhD**
Professor
Faculty of Science, Department of Mathematics
Universiti Putra Malaysia
(Member)

**Bashar Abdul Aziz Majeed Al-Talib, PhD**
Senior Lecturer
College of Computer Science and Mathematics, Department of Statistics and Informatics
University of Mosul, Iraq
(Member)

---

**BUJANG KIM HUAT, PhD**
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

## DECLARATION

I declare that the thesis is my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously, and is not concurrently, submitted for any other degree at Universiti Putra Malaysia or at any other institution.

**KAVEH KIANI**

Date:

## TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF APPENDICES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| AC | Anti-Conservative |
| Asy | Asymmetrical |
| C | Conservative |
| CI | Confidence Interval |
| CP | Censoring Proportion |
| DIC | Doubly Interval-Censored |
| EDIC | Exponential Model with Doubly Interval-Censored Data |
| EDICF1 | Exponential Model with Doubly Interval-Censored Data and Fixed Covariates, Case 1 |
| EDICF2 | Exponential Model with Doubly Interval-Censored Data and Fixed Covariates, Case 2 |
| EDICTD | Exponential Model with Doubly Interval-Censored Data and Time-Dependent Covariates |
| GICF | Gompertz Model with Interval-Censored Data and Fixed Covariates |
| GICTD | Gompertz Model with Interval-Censored Data and Time-Dependent Covariates |
| GRCF | Gompertz Model with Right-Censored Data and Fixed Covariates |
| GRCTD | Gompertz Model with Right-Censored Data and Time-Dependent Covariates |
| IC | Interval-Censored |
| LC | Left-Censored |
| MLE | Maximum Likelihood Estimate |
| OE | Observed Exactly |
| RC | Right-Censored |
| RMSE | Root Mean Square Error |
| SE | Standard Error |
| TD | Time-Dependent |

# CHAPTER 1

# INTRODUCTION

## 1.1 Background Study

In this section, survival time data analysis along with some basic definitions, models and concepts that will be used in subsequent chapters are introduced. These topics will not be discussed in detail.

### 1.1.1 Survival Time Data Analysis

Survival time or in some literature, life time, time-to-event and failure time, $T$, is the time from an initiating event (birth, marriage, affecting with a virus, enroll for a PhD. program) for a subject to some final event (death, divorce, onset of the diseases, graduation). $T$ is the non-negative random variable and usually continuous, unless stated otherwise.

Survival time data analysis, briefly, survival analysis is a collection of statistical methods for analyzing survival data where, the focus of these methods is to describe the distribution of $T$ on a population and relationship between $T$ and some covariates. A serious analytical problem in analyzing survival time data arises when a portion or even all $t_i$, $i = 1, 2, ..., n$, are censored. The $i^{th}$ subject's survival time, $t_i$, is censored when we do not know its exact value due to one of these four main reasons, the subject is lost to follow-up, the subject withdraws from the study, the subject does not experience the event before the study ends and the subject is not under continuous observation. However, we may have some partial information regarding the subject's survival time.

In survival analysis it is essential to have clear and explicit definitions of the time origin and endpoint. Following Kalbfleisch and Prentice (2002), time origin is the time from which survival is measured, $t = 0$. In some instances, time origin and endpoint could be the birth and death of the subject. Another example is infection with a virus and onset of the disease. In clinical trials time origin could also be assumed to be the time when all of the subjects in the trial had been event free.

The statistical analysis and modeling of survival time data are usually done by applying various kinds of non-parametric, semi-parametric or parametric models. Non-parametric models do not assume any parametric assumption about the form of the survivor function, $S(t)$.

On the other hand, a semi-parametric model is partially attached to some parametric assumption for $S(t)$. As Sasieni (2005) explains, "There is no widely accepted rigorous definition of a semi-parametric model. Informally we will call a model semi-parametric if it is not fully parametric but has a finite dimensional parameter of interest". In contrast, parametric models completely assume a fully parametric form for $S(t)$. More discussions about these three models are given in Sections 1.1.4 to 1.1.6.

### 1.1.2 Different Censoring Mechanisms

The $i^{th}$ subject is observed exactly (OE) at $t_i$ if the event of interest occurs at this time and this information is available to the observer. This subject is left-censored (LC) when $t_i \in (0, l_i]$ or, the subject has met the event at unknown time prior to $l_i$ and after time origin and it is right-censored (RC) when $t_i \in (r_i, \infty)$ or, the subject has been event free at the last known time $r_i$. Finally, the $i^{th}$ subject is interval-censored (IC) if instead of observing $t_i$ only an interval $(t_{L_i}, t_{R_i}]$ is observed where $t_i \in (t_{L_i}, t_{R_i}]$ and $t_{L_i} \leq t_{R_i}$.

Following Lawless (2003), there are two types of RC data, type-I and type-II. To describe type-I right censoring scheme let us assume that each subject has a potential censoring time $c_i$ and an event time $t_i$. The $i^{th}$ subject is RC if $c_i < t_i$; otherwise, subject is OE. This scheme usually occurs when study period is determined in advance. Type-II right censoring scheme arises when the study is terminated after first $k$ failures where $k \leq n$ and $n$ is number of subjects. Type-I RC data is referred to as RC data in this study.

IC data occur when subjects are under discontinuous observations/inspections. In this case, $T$ is not always OE or RC. IC data are more common in fields such as, sociology, economic, biology and epidemiology. Occurrence of IC data in medical studies is also very common when subjects are under scheduled follow-ups according to a predetermined calendar time for example weekly, monthly or yearly. In this study the term "actual inspection time (AC)" is used to denote the time when the subject was inspected.

Following Sun (2006) and Schick and Yu (2000) the six types of IC data are, case-I, case-II, case-k, mixed case and doubly IC data and panel count data.

Case-I IC data or current status data is referred to the IC data when all the $(T_L, T_R]$ intervals include either 0 or $\infty$ (Groeneboom and Wellner, 1992; Huang, 1996 and Schick and Yu, 2000). In this scheme, the $i^{th}$ subject is inspected only once at $ac_{i1}$ where $ac_{i1}$ will be left or right censoring time. Case-I IC data differ from RC data and LC data because there is not any OE data in this case. This type of IC data could be represented by

$$\left\{ac_{i1}, \delta_{i1} = I(t_i \leq ac_{i1}), \delta_{i2} = I(t_i > ac_{i1}), i = 1, ..., n\right\},$$

where $I$ is an indicator variable and $ac_{i1}$ is the actual inspection time for the $i^{th}$ subject. In this case

$$(t_{L_i}, t_{R_i}] = \begin{cases} (0, ac_{i1}], & \delta_{i1} = 1, \\ (ac_{i1}, \infty), & \delta_{i2} = 1. \end{cases} \tag{1.1}$$

In case-II IC data each subject is inspected twice. As a result, there is at least one finite interval $(T_L, T_R]$ belonging to $(0, \infty)$ (Groeneboom and Wellner, 1992; Huang and Wellner, 1997; Sun, 2005; Schick and Yu, 2000). This type of IC data could be represented by

$$\Big\{ac_{i1}, ac_{i2}, \delta_{i1} = I(t_i \le ac_{i1}), \delta_{i2} = I(ac_{i1} < t_i \le ac_{i2}), \delta_{i3} = I(t_i > ac_{i2}), i = 1, ..., n\Big\},$$

where $ac_{i1}$ and $ac_{i2}$ are two actual inspection times and $ac_{i1} \le ac_{i2}$. In this case

$$(t_{L_i}, t_{R_i}] = \begin{cases} (0, ac_{i1}], & \delta_{\mathrm{i1}} = 1, \\ (ac_{i1}, ac_{i2}], & \delta_{\mathrm{i2}} = 1, \\ (ac_{i2}, \infty), & \delta_{\mathrm{i3}} = 1. \end{cases} \tag{1.2}$$

Case-k IC data occurs when there are $k$ actual inspection times for subjects, $ac_{i1} \le ac_{i2} \le ... \le ac_{ik}$ where, $k$ is a fixed number (Schick and Yu, 2000; Wellner, 1995). This type of IC data could be represented by

$$\Big\{ac_{ij}, k, \delta_{ij} = I(ac_{i(j-1)} < t_i \le ac_{ij}), \delta_{i(k+1)} = I(t_i > ac_{ik}), i = 1, ..., n, j = 1, ..., k\Big\},$$

where $ac_{i0} = 0$. In this case

$$(t_{L_i}, t_{R_i}] = \begin{cases} (0, ac_{i1}], & \delta_{\mathrm{i1}} = 1, \\ (ac_{i(j-1)}, ac_{ij}], & \delta_{\mathrm{ij}} = 1 \text{ and } 1 < \mathrm{j} \le \mathrm{k}, \\ (ac_{ik}, \infty), & \delta_{\mathrm{i(k+1)}} = 1. \end{cases} \tag{1.3}$$

Mixed case IC data occurs when there is $k_i$ actual inspection times for the $i^{th}$ subject or $k$ is a random number, $ac_{i1} \le ac_{i2} \le ... \le ac_{ik_i}$ (Schick and Yu, 2000;

Wellner, 1995). This type of IC data could be represented by

$$\Big\{ac_{ij}, k_i, \delta_{ij} = I(ac_{i(j-1)} < t_i \leq ac_{ij}), \delta_{i(k_i+1)} = I(t_i > ac_{ik_i}), j = 1, ..., k_i, i = 1, ..., n\Big\},$$

where $ac_{i0} = 0$. In this case

$$(t_{L_i}, t_{R_i}] = \begin{cases} (0, ac_{i1}], & \delta_{\mathrm{i1}} = 1, \\ (ac_{i(j-1)}, ac_{ij}], & \delta_{\mathrm{ij}} = 1 \text{ and } 1 < \mathrm{j} \leq \mathrm{k}, \\ (ac_{ik_i}, \infty), & \delta_{\mathrm{i(k_i+1)}} = 1. \end{cases} \quad (1.4)$$

Mixed case IC data is very common in medical trials because the number of actual inspection times varies from patient to patient. It is clear that representation (1.4) is the general form of the representations (1.1), (1.2) and (1.3). In this study the term "IC" data is used to refer to "mixed case IC" data. More details about IC data is presented in Chapters 2 and 4.

Doubly interval-censored (DIC) data is the fifth type of IC data. DIC data often arises in the follow-up studies where $T$ is the elapsed time between two related events where both events are IC (Gruttola and Lagakos, 1989; Sun, 2004). If $V$ is the time to the first event and $W$ is the time to the second event then

$$V \in (V_L, V_R], \; W \in (W_L, W_R] \text{ and } T = W - V,$$

where $V_L \leq V_R$ and $W_L \leq W_R$. DIC data include usual RC and IC survival time data as special cases. For example DIC data reduce to IC data if $V$ is OE and $W$ is IC and DIC data reduce to RC data if $V$ is OE and $W$ is RC. DIC data is discussed more in Chapters 2 and 5.

Finally, panel count data is another form of IC data where in the presence of interval censoring, event of interest occurs more than one time during follow-ups and researcher is interested to know the exact number of these occurrences

(Lawless and Zhan, 1998; Thall, 1988). This type of IC data is not explored in this study. For more discussions see Sun (2006).

### 1.1.3 Independent Censoring

Independent censoring is a realistic assumption in survival analysis. Independent or random censoring is a censoring mechanism that is independent of the subjects' survival time distribution. For instance, consider two subjects that are under trial at time $t$ where one of them has a higher failure risk. Under independent censoring assumption, both subjects have a same probability to be RC at this time.

An independent censoring mechanism is also equivalent to the non-informative censoring. However, the reverse may not always true (Sun, 2006; Betensky, 2000; Oller et al., 2004). In contrast, dependent censoring and informative censoring are always the same.

For RC data assume $S(t)$, $f(t)$, $G(c)$ and $g(c)$ represent survivor function and probability density function of survival time $T$ and right censoring time $C$. If censoring process produces a potential censoring time $c_i$ for the $i^{th}$ subject and $t_i$ is the survival time of this subject then observed survival time would be $ot_i$ =min$(t_i, c_i)$. If $ot_i = c_i$ the indicator variable is $\delta_{R_i} = 1$ and $\delta_{R_i} = 0$ if $ot_i = t_i$. Dependent (informative) censoring for RC data scheme could be represented by

$$Pr(ot_i = t, \delta_{R_i} = 1) = Pr(c_i = t, t_i > t) = Pr(c_i = t)Pr(t_i > t|c_i) = g(t)S(t|c_i),$$
$$Pr(ot_i = t, \delta_{R_i} = 0) = Pr(t_i = t, c_i \geq t) = Pr(c_i \geq t|t_i)Pr(t_i = t) = G(t|t_i)f(t),$$

and for independent (non-informative) censoring is

$$Pr(ot_i = t, \delta_{R_i} = 1) = Pr(c_i = t, t_i > t) = Pr(c_i = t)Pr(t_i \geq t) = g(t)S(t),$$
$$Pr(ot_i = t, \delta_{R_i} = 0) = Pr(t_i = t, c_i \geq t) = Pr(c_i > t)Pr(t_i = t) = G(t)f(t).$$

For IC data assume $S(t)$ and $f(t)$ represent survivor function and probability density function of survival time $T$ and $G(t_L, t_R)$ represents joint survivor function of left and right censoring time. If censoring process produces $t_{L_i}$ and $t_{R_i}$ as a left and right censoring times and $t_i$ be the survival time of the subjects then dependent (informative) censoring for IC data scheme could be represent by

$$Pr(t_{L_i} = l, t_{R_i} = r | t_i = t) = Pr(l < t_i \leq r | t_{L_i} = l, t_{R_i} = r) G(l, r),$$

and for independent (non-informative) censoring is

$$Pr(t_{L_i} = l, t_{R_i} = r | t_i = t) = Pr(l < t_i \leq r).$$

For DIC data independent censoring assumption means that survival time, $T$, and time to the first event, $V$, are independent. Independent censoring for RC, IC and DIC data schemes are assumed throughout the thesis.

### 1.1.4 Parametric Survival Models

In parametric survival models $T$ usually follows a continuous distribution function, $f(t)$ where the cumulative distribution function is $F(t)$ =Pr($T \leq t$), survivor function is $S(t) = 1 - F(t)$ =Pr($T > t$), hazard function is $h(t) = f(t)/S(t)$ and cumulative hazard function is $H(t) = \int_0^t h(v)dv$. When one of these five functions is known other four functions could be specified. As a result, a parametric survival model is a model with specified $S_{\boldsymbol{\theta}}(t)$ function where, $\boldsymbol{\theta}$ is the vector of parameters.

Some widely used distributions to model $T$ are the exponential, Weibull, Gompertz, log-normal, log-logistic, generalized gamma and generalized $F$. The existences of different censoring mechanisms and different type of covariates have motivated researchers to introduce new parametric models and methods or extend existing parametric models to accommodate these components. Some of the most important advantages of the parametric models are existence of straightforward methods to obtain the maximum likelihood estimation (MLE) of the parameters, confidence intervals (CI) and hypothesis testing procedures.

The exponential and Gompertz are two models which are explored in this research. The exponential model has a constant hazard rate

$$h(t) = \lambda,$$

where $\lambda > 0$ is a parameter. The survivor function of the model is

$$S(t) = \exp(-\lambda t),$$

and the probability density function is

$$f(t) = \lambda \exp(-\lambda t).$$

The hazard rate of this distribution, $\lambda$, is constant or independent of time. Constant hazard rate or lack of aging or memory less property of the exponential distribution implies that

$$Pr(T \geq k) = Pr(T \geq k + t | T \geq k).$$

This implies that the chance for a new born subject to survive at least up to time $k$ is equal to chance of a $k$ age subject to survive additional $t$ units. Because of the memory less property this distribution has found limited use in studies where subject's aging process play an important role in its survival. This distribution is frequently used in the simulation studies in order to explore new and complicated models.

The Gompertz model was introduced by Gompertz (1825) as a model for human mortality. Recently, it has found more application in fields such as biology and demography. The hazard function of the Gompertz model is

$$h(t) = \lambda \exp(\gamma t),$$

The scale parameter is $\lambda > 0$ and the shape parameter is $\gamma$. The survivor function of the model is

$$S(t) = \exp\left[\frac{\lambda}{\gamma}\left(1 - e^{\gamma t}\right)\right]$$

and the probability density function is

$$f(t) = \lambda \exp(\gamma t) \times \exp\left[\frac{\lambda}{\gamma}\left(1 - e^{\gamma t}\right)\right].$$

The properties of the Gompertz distribution is presented in Johnson et al. (1994) and recently many authors have done studies on different characteristics and statistical methodology of Gompertz distribution, for instance, Makany (1991) and Chen (1997). The exponential model is the special case of the Gompertz model. In other words, Gompertz model could be reduced to the exponential model when the shape parameter $\gamma$ is assumed to be zero.

### 1.1.5 Non-Parametric Survival Models

Non-parametric methods have been discussed the most in the analysis of IC data. Kaplan and Meier (1958) derived an estimation for the $S(t)$ in the presence of RC data which is often referred to as the Kaplan-Meier or product-limit estimate. Thus, a non-parametric estimate for the cumulative hazard function has been proposed by Nelson (1969, 1972). Ayer et al. (1955) and van Eeden (1956) were the first to introduce non-parametric maximum likelihood estimates (MLE) of a distribution function based on case-I IC data. Non-parametric likelihood functions with other types of IC data are more analytically and practically complicated than that RC and case-I IC data and MLE of a $S(t)$ function does not have a closed form and should be calculate via an iterative algorithm.

Peto (1973) and Turnbull (1976) studied the estimations based on the case-II IC data. Suppose that $t_i \in (t_{Li}, t_{Ri}]$ where, $i = 1, 2, ..., n$ and $t_i$'s are iid with survivor function $S(t)$. Then the likelihood function is

$$L = \prod_{i=1}^{n} \Big\{ S(t_{Li}) - S(t_{Ri}) \Big\}. \tag{1.5}$$

If, $0 < s_0 < s_1 < s_2 < ... < s_{m-1} < s_m = \infty$ denote the ordered elements of the set $\Big\{0, t_{Li}, t_{Ri}, \infty, i = 1, 2, ..., n\Big\}$ then define an indicator variable $\alpha_{ij} = I(S_j \in (t_{Li}, t_{Ri}])$ and $p_j = S(s_{j-1}) - S(s_j)$, $j = 1, 2, ..., m$. Expression (1.5) can be rewritten as

$$L(P) = \prod_{i=1}^{n} \sum_{j=1}^{m} \alpha_{ij} p_j.$$

MLE of this function could be determine by applying one of the iterative procedures like, self-consistency algorithm of the EM algorithm by Dempster (1977), ICM algorithm by Groeneboom and Wellner (1992) and later modified by Jongbloed (1998) and EM-ICM algorithm proposed by Wellner and Zhan (1997).

Wellner (1995), Groeneboom (1996) and Huang and Wellner (1997) studied the non-parametric MLE of the case-k IC data and Schick and Yu (2000) and Van der Vaart and Wellner (2000) discussed non-parametric MLE of the mixed case IC data.

Non-parametric approaches to estimate $S(t)$ in the presence of DIC data usually use the self-consistency idea, see Sun (2006). There are three estimation procedures at the base of this idea. Firstly, a method based on the ML approach has been proposed by Gruttola and Lagakos (1989). Secondly, a two-step procedure as a simplification of the first method has been proposed by Gómez and Lagakos (1994) and finally, a conditional likelihood-based approach proposed by

Sun (1997). All these three methods can also be regarded as a generalization of the self-consistency algorithm.

### 1.1.6 Semi-Parametric Survival Models

Two major semi-parametric models are Cox proportional hazard model (PH) and accelerated failure time (AFT) or log-rank model. Hazard function of the PH model is

$$h(t, \boldsymbol{X}) = h_0(t)\exp(\boldsymbol{\beta'}\mathrm{X}),$$

and for the AFT model is

$$h(t, \boldsymbol{X}) = h_0(te^{-\beta' X})\exp(-\boldsymbol{\beta'}\mathrm{X}),$$

where $\boldsymbol{X}$ is the vector of the $p$ covariates, $\boldsymbol{X}$=$(x_1, x_2, ..., x_p)$, and $\boldsymbol{\beta}$ is the vesctor of $p$ parameters, $\boldsymbol{\beta}$=$(\beta_1, \beta_2, ..., \beta_p)$ and $h_0(t)$ is an unspecified baseline hazard function.

Three types of methods have discussed in literature for analyzing IC data by the PH model. Firstly, use an imputation method to reduce IC data to RC data and then use specified methods for the RC data for example see, Satten (1996), Goggins et al. (1998) and Pan (2000a). Secondly, estimating the unspecified baseline hazard function, $h_0(t)$, non-parametrically and then estimate covariate's parameters, Finkelstein (1986) and Goetghebeur and Ryan (2000). Thirdly and finally, using regression splines or local likelihood smoothing to estimate baseline hazard function, $h_0(t)$, and then estimate rest of the parameters where, Kooperberg and Clarkson (1997) and Betensky and Finkelstein (1999) are pioneers of these two methods.

Analyzing IC data with AFT model is less developed compared to the PH model and some of the main works were given by Rabinowitz et al. (1995), Betensky et al.(2001) and Komárek et al. (2005).

### 1.1.7 Time-Dependent Covariate

Many parametric regression models have been developed to analyze relationship between $T$ as an outcome variable and covariates in the presence of censored data. Covariates can be either fixed or time-dependent (TD). Fixed covariates are measured at the start of study and stay constant over the study period, for instance, gender or race. TD covariates on the other hand vary over time, for example, age of subjects, blood pressure and cholesterol level. Cox (1972) has introduced idea of TD covariates and he has suggested the use of TD covariate in the PH regression models. Idea behind using the TD covariates implies that the history of a TD covariate process up to time $t$ should be incorporated into the model to assess the full effect of the covariate on the $T$, because this history may influence the rate of survival over time.

Following Kalbfleisch and Prentice (2002), TD covariates are categorized as either external (exogenous) or internal (endogenous). External TD covariate may influence the observed survival time of a subject, $t$, but the covariate's path after $t$ and $t$ are independent. For instance, the effect of air pollution on subjects heart attack. In contrast, the path of the internal TD covariate is dependent on the survival status of a subject or the TD covariate process is generated by the subject for example, blood pressure of a subject. This covariate is observable as long as the subject is under observation.

## 1.2 Objectives

The aim of this research is to obtain parametric survival models for survival time data with LC, RC, IC, DIC and OE data in presence of TD covariates and obtain various CI estimates for the parameters of these models. In this research we will utilize and extend two important parametric survival models, the Gompertz and the exponential to accommodate these censoring mechanisms and TD covariates. As mentioned before the exponential model is the special case of the Gompertz model. The exponential model is used in Chapter 5 to establish a base for future extensions with other parametric models include Gompertz.

Three main models will be explored,

- Gompertz model with RC data and TD covariates (GRCTD),
- Gompertz model with IC data and TD covariates (GICTD),
- Exponential model with DIC data and TD covariates (EDICTD).

In order to achieve these final models other models must be explored in advance. The models are

- Gompertz model with RC data and fixed covariates (GRCF),
- Gompertz model with IC data and fixed covariates (GICF),
- Exponential model with DIC data (EDIC),
- Exponential model with DIC data and fixed covariates, Case 1 (EDICF1),
- Exponential model with DIC data and fixed covariates, Case 2 (EDICF2).

The main objectives of this research are as follows:

- To extend the Gompertz model to incorporate TD covariates in the presence of RC data or obtaining GRCTD model and conduct simulation to study the bias, standard error (SE) and root mean square error (RMSE) of the parameter estimates.

- To extend the Gompertz model to incorporate TD covariates in the presence of IC data or obtaining GICTD model and conduct simulation to study the bias, SE and RMSE of the parameter estimates.

- To extend the exponential model to incorporate TD covariates in the presence of DIC data or obtaining EDICTD model and conduct simulation to study the bias, SE and RMSE of the parameter estimates.

- Obtain GRCF, GICF, EDIC, EDICF1 and EDICF2 models and study performance of the models.

- Evaluate the performance of the the Wald, jackknife and likelihood ratio CI estimate methods for these models and draw the results based on the coverage probability study.

## 1.3 Outline of Thesis

This thesis is organized into six chapters. Chapter 2 provides a review of related literature to the current work. Special consideration in this literature review is the research conducted on parametric models with RC, IC and DIC data and parametric models with TD covariate.

Chapter 3 starts with the extension of the Gompertz model to incorporate TD covariates in the presence of RC data (GRCTD model). Then, the performance of the model is compared with the fixed covariate model (GRCF model). Following that, comparisons are made when a fixed covariate model is used instead of the true TD covariate model. In addition, two methods of constructing CIs, the Wald and jackknife are explored for the parameters of the model. Conclusions are drawn based on the coverage probability study.

Chapter 4 concentrates on the extension of the Gompertz model to incorporate TD covariates with LC, RC, IC data as well as OE data (GICTD model). The model is then investigated thoroughly at dependent and independent covariate levels through a comprehensive simulation study. Following that, the model is compared with a fixed covariate model (GICF model). Then, two methods of constructing CIs the Wald and likelihood ratio are investigated for the parameters of the model and conclusions are drawn based on the coverage probability study.

Chapter 5 investigates a parametric survival model that accommodates DIC data with TD covariates (EDICTD model). In order to achieve this final model three models are explored in advance. Firstly, a simple model consisting DIC data without any covariate is studied (EDIC model). Following that, a model with DIC data and fixed covariates is considered where all the covariates affect $T$ (EDICF1 model). Lastly, a model with fixed covariates is studied where some of the covariates affect $T$ and the others affect $V$ (EDICF2 model). All these models are studied by the simulation study and two methods of constructing confidence intervals, the Wald and jackknife are explored for the parameters of these models.

Finally, Chapter 6 summarizes the study and offers some recommendations for future research.

It should be mentioned all simulation studies were carried out by using the FORTRAN®, (FTN95), programming language.

## BIBLIOGRAPHY


Aitkin, M., & Clayton, D. (1980). The Fitting of Exponential, Weibull and Extreme Value Distributions to Complex Censored Survival Data Using GLIM. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 29(2), 156-163.

Arasan, J. (2006). *Lifetime of Parallel Component Systems with Dependent Failures and Multiple Covariates*. Ph.D. Thesis. Oxford University, UK.

Arasan, J., & Lunn, M. (2008). Alternative interval estimation for parameters of bivariate exponential model with time varying covariate. *Computational Statistics*, 23(4), 605-622.

Arasan, J., & Lunn, M. (2009). Survival model of a parallel system with dependent failures and time varying covariates. *Journal of Statistical Planning and Inference*, 139(3), 944-951.

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., & Silverman, E. (1955). An Empirical Distribution Function for Sampling with Incomplete Information. *The Annals of Mathematical Statistics*, 26(4), 641-647.

Bacchetti, P. (1990). Estimating the Incubation Period of AIDS by Comparing Population Infection and Diagnosis Patterns. *Journal of the American Statistical Association*, 85(412), 1002-1008.

Bacchetti, P., & Jewell, N. P. (1991). Nonparametric Estimation of the Incubation Period of AIDS Based on a Prevalent Cohort with Unknown Infection Times. *Biometrics*, 47(3), 947-960.

Bagdonavicius, V. B., & Nikulin, M. S. (1997). Transfer functions and semiparametric regression models. *Biometrika*, 84(2), 365-378.

Banerjee, T., Grover, G., Pensi, T., Banerjee, D. (2011). Estimation of hazard of death in vertically transmitted HIV-1-infected children for doubly censored failure times and fixed covariates. *Journal of Applied Statistics*, 38(2), 273-285.

Bebchuk, J. D., & Betensky, R. A. (2000). Multiple imputation for simple estimation of the hazard function based on interval censored data. *Statistics in medicine*, 19(3), 405-419.

Betensky, R. A. (2000). Miscellanea. On nonidentifiability and noninformative censoring for current status data. Biometrika, 87(1), 218-221.

Betensky, R. A., & Finkelstein, D. M. (1999). An extension of Kendall's coefficient of concordance to bivariate interval censored data. *Statistics in medicine*, 18(22), 3101-3109.

Betensky, R. A., Rabinowitz, D., & Tsiatis, A. A. (2001). Computationally simple accelerated failure time regression for interval censored data. *Biometrika*, 88(3), 703-711.

Brookmeyer, R., & Goedert, J. J. (1989). Censoring in an epidemic with an application to hemophilia-associated AIDS. *Biometrics*, 45(1), 325-335.

Brown, C. C. (1975). On the use of indicator variables for studying the time-dependence of parameters in a response-time model. *Biometrics*, 31(4), 863-872.

Cai, Z., & Sun, Y. (2003). Local Linear Estimation for Time-Dependent Coefficients in Cox's Regression Models. *Scandinavian Journal of Statistics*, 30(1), 93-111.

Chen, Z. (1997). Parameter Estimation of the Gompertz Population. *Biometrical Journal*, 39(1), 117-124.

Chen, D. G., Lio, Y. L. (2010). Parameter estimations for generalized exponential distribution under progressive type-I interval censoring. *Computational Statistics & Data Analysis*, 54(6), 1581-1591.

Cheng, C., Chen, J., & Li, Z. (2010). A New Algorithm for Maximum Likelihood Estimation with Progressive Type-I Interval Censored Data. *Communications in Statistics, Simulation and Computation*, 39(4), 750-766.

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2), 187-220.

Cox, D. R., & Hinkley, D. V. (1974). *Theoretical statistics.* London: Chapman & Hall.

Cox, D. R., & Oakes, D. (1984). *Analysis of survival data.* London: Chapman & Hall.

Crowley, J., & Hu, M. (1977). Covariance Analysis of Heart Transplant Survival Data. *Journal of the American Statistical Association*, 72(357), 27-36.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), 1-38.

Doganaksoy, N., & Schmee, J. (1993). Comparisons of Approximate Confidence Intervals for Distributions Used in Life-Data Analysis. *Technometrics*, 35(2), 175-184.

Dorey, F. J., Little, R. J. A., & Schenker, N. (1993). Multiple imputation for threshold-crossing data with interval censoring. *Statistics in medicine*, 12(17), 1589-1603.

Dümbgen, L., Freitag-Wolf, S., & Jongbloed, G. (2006). Estimating a unimodal distribution from interval-censored data. *Journal of the American Statistical*

*Association*, 101(475), 1094-1106.

Fang, H. B., & Sun, J. (2001). Consistency of nonparametric maximum likelihood estimation of a distribution function based on doubly interval-censored failure time data. *Statistics & probability letters*, 55(3), 311-318.

Finkelstein, D. M. (1986). A Proportional Hazards Model for Interval-Censored Failure Time Data. *Biometrics*, 42(4), 845-854.

Goetghebeur, E., & Ryan, L. (2000). Semiparametric Regression Analysis of Interval-Censored Data. *Biometrics*, 56(4), 1139-1144.

Goggins, W. B., Finkelstein, D. M., Schoenfeld, D. A., & Zaslavsky, A. M. (1998). A Markov Chain Monte Carlo EM Algorithm for Analyzing Interval-Censored Data under the Cox Proportional Hazards Model. *Biometrics*, 54(4), 1498-1507.

Goggins, W. B., Finkelstein, D. M., & Zaslavsky, A. M. (1999a). Applying the Cox proportional hazards model for analysis of latency data with interval censoring. Statistics in medicine, 18(20), 2737-2747.

Goggins, W. B., Finkelstein, D. M., & Zaslavsky, A. M. (1999b). Applying the Cox Proportional Hazards Model When the Change Time of a Binary Time-Varying Covariate Is Interval Censored. *Biometrics*, 55(2), 445-451.

Gómez, G., & Lagakos, S. W. (1994). Estimation of the Infection Time and Latency Distribution of AIDS with Doubly Censored Data. *Biometrics*, 50(1), 204-212.

Gómez, G., & Calle, M. L. (1999). Non-parametric estimation with doubly censored data. *Journal of Applied Statistics*, 26(1), 45-58.

Gómez, G., Calle, M. L., & Oller, R. (2004). Frequentist and Bayesian approaches for interval-censored data. *Statistical Papers*, 45(2), 139-173.

Gómez, G., Calle, M. L., Oller, R., & Langohr, K. (2009). Tutorial on methods for interval-censored data and their implementation in R. *Statistical Modelling*, 9(4), 259.

Gompertz, B. (1825). On the Nature of the Function Expressive of the Law of Human Mortality and on the New Mode of Determining the Value of Life Contingencies. *Philosophical Transactions of Royal Society A*, 115, 513-580.

Groeneboom, P. (1996). *Lectures on inverse problems. In Lecture Notes in Mathematics* (Vol. 1648, pp. 67-164). Berlin: Springer.

Groeneboom, P., & Wellner, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation.* DMV Seminar. Band 19. Boston:Birkhäuser.

Gruttola, V. D., & Lagakos, S. W. (1989). Analysis of Doubly-Censored Survival Data, with Application to AIDS. *Biometrics*, 45(1), 1-11.

Hashimoto, E. M., Ortega, E. M. M., Cancho, V. G., & Cordeiro, G. M. (2010). The log-exponentiated Weibull regression model for interval-censored data. *Computational Statistics & Data Analysis*, 54(4), 1017-1035.

Hastie, T., & Tibshirani, R. (1993). Varying-Coefficient Models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55(4), 757-796.

Huang, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *The Annals of Statistics*, 24(2), 540-568.

Huang, J., & Wellner, J. A. (1997). Interval censored survival data: a review of recent progress. In *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis* (pp. 123-170). New York: Springer-Verlag.

Jewell, N. P. (1994). Non-parametric estimation and doubly-censored data: General ideas and applications to AIDS. *Statistics in medicine*, 13(19-20), 2081-2095.

Jewell, N. P., Malani, H. M., & Vittinghoff, E. (1994). Nonparametric Estimation for a Form of Doubly Censored Data, With Application to Two Problems in AIDS. *Journal of the American Statistical Association*, 89(425), 7-18.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous univariate distributions*, vol. 2. New York: John Wiley & Sons.

Jongbloed, G. (1998). The Iterative Convex Minorant Algorithm for Nonparametric Estimation. *Journal of Computational and Graphical Statistics*, 7(3), 310-321.

Kalbfleisch, J. D., & Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data.* New York: John Wiley & Sons.

Kaplan, E. L., & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, 53(282), 457-481.

Kim, D. K. (1997). Regression Analysis of Interval-Censored Survival Data with Covariates Using Log-Linear Models. *Biometrics*, 53(4), 1274-1283.

Kim, M. Y., Gruttola, V. G. D., & Lagakos, S. W. (1993). Analyzing Doubly Censored Data with Covariates, with Application to AIDS. *Biometrics*, 49(1), 13-22.

Komárek, A., & Lesaffre, E. (2008). Bayesian accelerated failure time model with multivariate doubly interval-censored data and flexible distributional assumptions. *Journal of the American Statistical Association*, 103(482), 523-533.

Komárek, A. t., Lesaffre, E., & Hilton, J. F. (2005). Accelerated Failure Time Model for Arbitrarily Censored Data With Smoothed Error Distribution. *Journal of Computational and Graphical Statistics*, 14(3), 726-745.

Kooperberg, C., & Clarkson, D. B. (1997). Hazard Regression with Interval-Censored Data. *Biometrics*, 53(4), 1485-1494.

Lawless, J. F. (2003). *Statistical models and methods for lifetime data* (2nd ed.). New York: John Wiley & Sons.

Lawless, J. F., & Babineau, D. (2006). Models for interval censoring and simulation-based inference for lifetime distributions. *Biometrika*, 93(3), 671-686.

Lawless, J. F., & Zhan, M. (1998). Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. *Canadian Journal of Statistics*, 26(4), 549-565.

Lee, C. (1999). An urn model in the simulation of interval censored failure time data. *Statistics & probability letters*, 45(2), 131-139.

Lesaffre, E., Komárek, A., & Declerck, D. (2005). An overview of methods for interval-censored data with an emphasis on applications in dentistry. *Statistical methods in medical research*, 14(6), 539-552.

Lim, H., Sun, J., & Matthews, D. E. (2002). Maximum likelihood estimation of a survival function with a change point for truncated and interval-censored data. *Statistics in medicine*, 21(5), 743-752.

Lindsey, J. C., & Ryan, L. M. (1998). Methods for interval-censored data. *Statistics in medicine*, 17(2), 219-238.

Lindsey, J. K. (1998). A study of interval censoring in parametric regression models. *Lifetime Data Analysis*, 4(4), 329-354.

Ma, S. (2010). Mixed case interval censored data with a cured subgroup. *Statistica Sinica*, 20, 1165-1181.

Makany, R. (1991). A Theoretical Basis for Gompertz'S Curve. *Biometrical Journal*, 33(1), 121-128.

Mantel, N., & Hankey, B. F. (1978). A logistic regression analysis of response-time data where the hazard function is time dependent. *Communications in Statistics - Theory and Methods*, 7(4), 333-347.

Martinussen, T., Scheike, T. H., & Skovgaard, I. M. (2002). Efficient Estimation of Fixed and Time-varying Covariate Effects in Multiplicative Intensity Models. *Scandinavian Journal of Statistics*, 29(1), 57-74.

Marzec, L., & Marzec, P. (1997). On fitting Cox's regression model with time-dependent coefficients. *Biometrika*, 84(4), 901-908.

McBryde, E. S., Gibson, G., Pettitt, A. N., Zhang, Y., Zhao, B., & McElwain, D. L. S. (2006). Bayesian modelling of an epidemic of severe acute respiratory syndrome. *Bulletin of mathematical biology*, 68(4), 889-917.

Murphy, S. A. (1995). Likelihood Ratio-Based Confidence Intervals in Survival Analysis. *Journal of the American Statistical Association*, 90(432), 1399-1405.

Murphy, S. A., & Sen, P. K. (1991). Time-dependent coefficients in a Cox-type regression model. *Stochastic Processes and their Applications*, 39(1), 153-180.

Nelson, W. (1969). Hazard plotting for incomplete failure data(Multiply censored data plotting on various type hazard papers for engineering information on time to failure distribution). *Journal of Quality Technology*, 1, 27-52.

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

Nelson, W. B. (1990). *Accelerated testing: statistical models, test plans and data analyses*. New York: John Wiley & Sons.

Ng, H. K. T., & Wang, Z. (2009). Statistical estimation for the parameters of Weibull distribution based on progressively type-I interval censored sample. *Journal of Statistical Computation and Simulation*, 79(2), 145-159.

Odell, P. M., Anderson, K. M., & D'Agostino, R. B. (1992). Maximum likelihood estimation for interval-censored data using a Weibull-based accelerated failure time model. *Biometrics*, 48(3), 951-959.

Oller, R., Gomez, G., & Calle, M. L. (2004). Interval censoring: Model characterizations for the validity of the simplified likelihood. *Canadian Journal of Statistics*, 32(3), 315-326.

Pan, W. (2000a). A Multiple Imputation Approach to Cox Regression with Interval-Censored Data. *Biometrics*, 56(1), 199-203.

Pan, W. (2000b). A two-sample test with interval censored data via multiple imputation. *Statistics in medicine*, 19(1), 1-11.

Pan, W. (2001). A multiple imputation approach to regression analysis for doubly censored data with application to AIDS studies. *Biometrics*, 57(4), 1245-1250.

Peng, X., & Yan, Z. (2011). *Parameter estimations with gamma distribution based on progressive type-I interval censoring*. Paper presented at the Proceedings - 2011 IEEE International Conference on Computer Science and Automation Engineering, CSAE 2011.

Petersen, T. (1986). Fitting Parametric Survival Models with Time-Dependent Covariates. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 35(3), 281-288.

Peto, R. (1973). Experimental Survival Curves for Interval-Censored Data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 22(1), 86-91.

Pons, O. (2000). Nonparametric estimation in a varying-coefficient Cox model. Mathematical Methods of Statistics, 9(4), 376-398. Rabinowitz, D., Tsiatis, A., & Aragon, J. (1995). Regression with interval-censored data. *Biometrika*, 82(3), 501-513.

Rabinowitz, D., Tsiatis, A., & Aragon, J. (1995). Regression with interval-censored data. *Biometrika*, 82(3), 501-513.

Reich, N. G., Lessler, J., Cummings, D. A. T., & Brookmeyer, R. (2009). Estimating incubation period distributions with coarse data. *Statistics in medicine*, 28(22), 2769-2784.

Robins, J., & Tsiatis, A. A. (1992). Semiparametric estimation of an accelerated failure time model with time-dependent covariates. *Biometrika*, 79(2), 311-319.

Sasieni, P. (2005). Semiparametric Regression. In *Encyclopedia of biostatistics*. John Wiley & Sons.

Satten, G. A. (1996). Rank-based inference in the proportional hazards model for interval censored data. *Biometrika*, 83(2), 355-370.

Satten, G. A., Datta, S., & Williamson, J. M. (1998). Inference Based on Imputed Failure Times for the Proportional Hazards Model with Interval-Censored Data. *Journal of the American Statistical Association*, 93(441), 318-327.

Scallan, A. J. (1999). Regression modelling of interval-censored failure time data using the Weibull distribution. *Journal of Applied Statistics*, 26(5), 613-618.

Scheike, T. H. (2004). Time-Varying Effects in Survival Analysis. In *Handbook of Statistics: Advances in survival analysis* (Vol. 23, pp. 61-85). Amsterdam: Elsevier.

Schick, A., & Yu, Q. (2000). Consistency of the GMLE with Mixed Case Interval-Censored Data. *Scandinavian Journal of Statistics*, 27(1), 45-55.

Sen, B., & Banerjee, M. (2007). A pseudolikelihood method for analyzing interval censored data. *Biometrika*, 94(1), 71-86.

Song, S. (2004). Estimation with univariate" mixed case" interval censored data. *Statistica Sinica*, 14(1), 269-282.

Sparling, Y. H., Younes, N., Lachin, J. M., & Bautista, O. M. (2006). Parametric survival models for interval-censored data with time-dependent covariates. *Biostatistics*, 7(4), 599-614.

Sun, J. (1995). Empirical Estimation of a Distribution Function with Truncated and Doubly Interval-Censored Data and Its Application to AIDS Studies. *Biometrics*, 51(3), 1096-1104.

Sun, J. (2001). Nonparametric Test for Doubly Interval-Censored Failure Time

Data. *Lifetime Data Analysis*, 7(4), 363-375.

Sun, J. (1997). Self-Consistency Estimation of Distributions Based on Truncated and Doubly Censored Survival Data with Applications to AIDS Cohort Studies. *Lifetime Data Analysis*, 3(4), 305-313.

Sun, J. (2004). Statistical analysis of doubly interval-censored failure time data. *Handbook of Statistics: Survival Analysis*, 23, 105-122.

Sun, J. (2005). Interval Censoring. In *Encyclopedia of biostatistics*: John Wiley & Sons.

Sun, J. (2006). *The statistical analysis of interval-censored failure time data*. New York: Springer-Verlag.

Sun, J., Liao, Q., & Pagano, M. (1999). Regression Analysis of Doubly Censored Failure Time Data with Applications to AIDS Studies. *Biometrics*, 55(3), 909-914.

Sun, J., Lim, H.-J., & Zhao, X. (2004a). An Independence Test for Doubly Censored Failure Time Data. *Biometrical Journal*, 46(5), 503-511.

Sun, L., Kim, Y.-j., & Sun, J. (2004b). Regression Analysis of Doubly Censored Failure Time Data Using the Additive Hazards Model. *Biometrics*, 60(3), 637-643.

Thall, P. F. (1988). Mixed Poisson Likelihood Regression Models for Longitudinal Interval Count Data. *Biometrics*, 44(1), 197-209.

Tu, X. M. (1995). Nonparametric Estimation of Survival Distributions with Censored Initiating Time, and Censored and Truncated Terminating Time: Application to Transfusion Data for Acquired Immune Deficiency Syndrome. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 44(1), 3-16.

Turnbull, B. W. (1974). Nonparametric estimation of survivorship function with doubly censored data. *Journal of the American Statistical Association*, 69, 169-173.

Turnbull, B. W. (1976). The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data. J*ournal of the Royal Statistical Society. Series B (Methodological)*, 38(3), 290-295.

van der Vaart, A., & Wellner, J. A. (2000). Preservation theorems for Glivenko-Cantelli and uniform Glivenko-Cantelli classes. In H*igh Dimensional Probability II* (pp. 115-134). Boston:Birkhäuser.

van Eeden, C. (1956). Maximum likelihood estimation of ordered probabilities. *Indagationes Mathematicae*, 18, 444-455.

Wellner, J. A. (1995). Interval Censoring, Case 2: Alternative Hypotheses. *Lec-*

*ture Notes-Monograph Series*, 27, 271-291.

Wellner, J. A., & Zhan, Y. (1997). A Hybrid Algorithm for Computation of the Nonparametric Maximum Likelihood Estimator From Censored Data. *Journal of the American Statistical Association*, 92(439), 945-959.

Wulfsohn, M. S., & Tsiatis, A. A. (1997). A Joint Model for Survival and Longitudinal Data Measured with Error. *Biometrics*, 53(1), 330-339.

Zhang, Z., & Sun, J. (2010). Interval censoring. *Statistical methods in medical research*, 19(1), 53-70.

Zucker, D. M., & Karr, A. F. (1990). Nonparametric survival analysis with time-dependent covariate effects: A penalized partial likelihood approach. *The Annals of Statistics*, 18(1), 329-353.

## BIODATA OF STUDENT

Kaveh Kiani was born on November 15 1976 in Zanjan, Iran. After graduating from the Ostad Reza Rozbeh high school in Zanjan in 1995, he started his study in Statistics at the Shahid Beheshti University (SBU), Tehran, Iran and he obtained his Bachelor's degree in 2000. He was then employed by Statistical Center of Iran (SCI) in 2001. He obtained his Master's degree in Social and Economic Statistics from Allameh Tabatabaee University (ATU), Tehran, Iran in 2004. In December 2008, he enrolled in the Doctor of Philosophy program in Applied Statistics at Universiti Putra Malaysia (UPM), Institute for Mathematical Research (INSPEM).

## LIST OF PUBLICATIONS

The following papers are extracted from the current thesis:

1. Kiani, K., Arasan, J. (2012). Gompertz model with time-dependent covariate in the presence of interval-, right- and left-censored data. *Journal of Statistical Computation and Simulation*, 1-19, doi: 10.1080/00949655.2012.662979, URL: http://dx.doi.org/10.1080/00949655.2012.662979.

2. Kiani, K., Arasan, J. and Midi, H. (2012). Interval Estimations for Parameters of Gompertz Model with Time-Dependent Covariate and Right Censored Data. *Sains Malaysiana*, 41(4)(2012), 471-480.

3. Kiani, K. and Arasan, J. (2012). Simulation of Interval Censored Data in Medical and Biological Studies. *International Journal of Modern Physics: Conference Series (IJMPCS), Proceedings of the International Conference on Mathematical and Computational Biology 2011*, 112-118, doi: 10.1142/S2010194512005168.

4. Kiani, K. and Arasan, J. (2012). Interval Estimations For Parameters Of Exponential Model With Doubly Interval-Censored Survival Time Data. In Proceedings of the 2nd Regional Conference on Applied and Engineering Mathematics 2012 (RCAEM-II), 30-31 May, Penang, MALAYSIA.

5. Kiani, K., Arasan, J. and Midi, H. (2010). Confidence Interval Estimations for Parameters of Gompertz Survival Time Model. In Proceedings of the Fundamental Science Congress: Mathematics Symposium (FSC 2010).

6. Kiani, K. and Arasan, J. (2009). Gompertz Model with Time-Dependent Covariate and Right Censored Data. In Proceedings of the Simposium Kebangsaan Sains Matematik (KE-17), 15-17 December 2009, Melaka, MALAYSIA, pp: 771-777.