AN APPLICATION OF SURVIVAL ANALYSIS IN AUTO INSURANCE CONTRACTS IN GHANA

BY

KWAKU OPOKU-AMEYAW

(10283545)

THIS THESIS IS SUBMITTED TO THE UNIVERSITY OF GHANA, LEGON IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF MPHIL STATISTICS DEGREE

DECLARATION

Candidate's Declaration

This is to certify that, this thesis is the result of my own research work and no part of it has been presented for another degree in this University or elsewhere.

SIGNATURE: ……...….. DATE: …….

KWAKU OPOKU-AMEYAW

(10283545)

Supervisors' Declaration

We hereby certify that this thesis was prepared from this candidate's own work and supervised in accordance with guidelines on supervision of the thesis laid down by the University of Ghana.

SIGNATURE: …… DATE: …….

DR. EZEKIEL N.N. NORTEY

(Principal Supervisor)

SIGNATURE: …… DATE: …….....

DR. ISAAC BAIDOO

(Co-Supervisor)

DEDICATION

This work is dedicated to my parents, Dr.Kwabena Opoku-Ameyaw and the late Mrs Gloria Opoku-Ameyaw.

ABSTRACT

Survival models for life-time data and other time-to-event data are widely used in many fields, including medicine, economics, agriculture and actuarial mathematics. In this study, survival analysis was applied to the Ghanaian insurance industry to model the time until first claim after policy inception and time until payment after reporting. The nonparametric Kaplan-Meier model is used in the analysis. Cumulative hazard functions for time until claim reporting and time until payment were calculated. Confidence intervals were also computed for the Kaplan-Meier estimates. The findings indicate that time until reporting claims and time until payment followed a polynomial of order 6. It was also observed that the log-transformed confidence interval is better than the linear confidence interval. The probability that claims will be reported within a shorter period (e.g. a week) was higher than that of a longer period (e.g. a month). It was concluded that survival analysis is an appropriate tool for studying the insurance industry.

ACKNOWLEDGEMENTS

I would like to thank the Almighty God for the gift of life and the strength given me throughout this research study. I would also like to thank my supervisors Dr. Ezekiel N.N Nortey and Dr. Isaac Baidoo for their excellent guidance and support throughout my study.

I render special thanks to Mr. Enoch Nii Boi Quaye for his immense help during this study.

I have received various help from members of the Department of Statistics particularly Dr. F.O.Mettle and I sincerely thank him.

Finally, I am deeply indebted to my father Dr. Kwabena Opoku-Ameyaw and the rest of my family for their love and support throughout my study.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

CI……. Confidence interval

DUC…….. Days until reporting claims

DUP……Days until payments are made to policyholders

KM…….. Kaplan- Meier

CHAPTER ONE

GENERAL INTRODUCTION

1.0 Introduction

Statistics is an important or inevitable tool in life. Its importance is derived from the fact that it has been applied in various fields of life such as economics, business, medicine, agriculture just to mention a few. There has been an overwhelming growth in the development and application of statistical methodology over the years. Its applications have been in the form of studying relationships between sets of variables, forecasting and estimation, determining contributory or underlying factors affecting various variables among others.

Modelling which is one of the most powerful techniques applied in statistics has been applied in the fields of banking, economics and other fields to make predictions and also estimations. For example, modelling has also been of great importance in the field of actuarial science. An example can be found in the area of insurance. General insurance consisting of health, property/personal and motor are perhaps the fastest growing areas for actuaries in recent times (Boland, 2006). As a result of this upcoming development, general actuaries have to fully understand the models they use for analysing risk, time until claim reporting and payment among others by a policyholder of an insurance company. Various models have been found to be very attractive for portfolio management. Some of the reasons modelling are preferred to other techniques in statistics in the area of insurance are firstly it helps in estimation of various parameters or coefficients. Secondly, it assists in making predictions, which facilitate the development and implementation of policies. Lastly, it helps in studying relationship between observed variables.

One of the widely used forms of modelling is survival models. Survival analysis which is sometimes referred to as failure time analysis was originally designed to study time until death. However it is currently applied in various disciplines such as event history, sociology,

engineering, economics to analyse the onset of disease, equipment failures, earthquakes, automobile accidents, stock market crashes, revolutions, job terminations, births, marriages, divorces, promotions, retirements, and arrests (Allison, 1995).

The main outcome variable of interest in survival analysis is the time taken for an event to occur (Kleinbaum, 1996). Survival analysis can be performed when the data consist only of the times of events. However the aim of survival analysis is to estimate causal or predictive models in which the risk of an event depends on covariates. According to Allison (1995), these covariates are mostly of two types mainly those constant with time such as sex and race and those time-varying covariates like income, marital status and blood pressure among others. Survival analysis has two features which make it more attractive than the conventional statistical methods like linear regression models. Some of the most commonly used models under survival analysis are the exponential regression, log-normal regression, proportional hazards regression, competing risks models, and discrete-time methods, to name only a few. Due to the importance of unique features of survival models such as making provisions for time-varying covariates, I intend applying survival models to the field of actuarial science particularly the motor insurance industry in Ghana to determine the duration it takes for policyholders to report claims after policy issue and get settled after reporting claims.

The insurance industry in Ghana dates as far back as in the 1920's. The National Insurance Commission (NIC) was established under Insurance Law 1989 (PNDC Law 227) but now operates under the new Insurance Act, 2006 (Act 724). The objective of the Commission, as detailed in Act 724 was to ensure effective administration, supervision, regulation and control of the business of Insurance in Ghana. The Commission was authorized to perform a wide range of functions comprising licensing of entities, setting of standards and facilitating the setting of codes for practitioners. The insurance companies in Ghana can be listed under two

main headings, namely life insurance and non-life insurance. As of 2011, there were 24 non-life companies, 18 life companies, 2 reinsurance companies, 53 broking companies comprising 1 loss adjuster and 1 reinsurance broker (NIC, 2011). A vibrant sector of the non-life insurance in Ghana is the motor insurance industry. In Ghana, every car owner by law is a consumer of insurance services because the law stipulates that driving without insurance is an offense punishable by fines, disqualification and in some cases imprisonment. It is a criminal offence not to insure your motor vehicle. Hence, other than car insurance, in the non-life category no other insurance is obligatory for the customers (Goswami, 2007).

At the moment there are three types of motor insurance contracts in Ghana. These are (i) the third party, (ii) third party, fire and theft and (iii) the comprehensive motor insurance contract. The third party is the minimum amount of insurance cover that one must have by law for a vehicle. It only covers for damage to someone else's vehicle or property, or injury to someone else in an accident, which involves a car. It also includes accidents caused by a passenger. If a vehicle is damaged in the accident one will have to pay for the repairs himself/herself. The third party, fire and theft type of insurance covers in addition to the compensation under third party, damage to or loss of a car by fire or theft. In the case of the comprehensive type the compensation includes third party, fire and theft insurance covers as well as repairs to a car. In some policies under comprehensive insurance other benefits include:

i. cover for one's own death or injury, or that of a partner or other members of a family, up to a limited amount.
ii. cover for personal belongings if they are stolen from one's vehicle or damaged
iii. cover for medical and legal expenses
iv. hiring a replacement vehicle.

1.1 Problem statement

The use of vehicles has played an integral role in mortality risk in several countries through motor accidents (Maddala, 2005). As a result of this, the motor vehicle or automobile insurance was designed to take care of the possible loss of life and property that might result from the use of vehicles (Onafalujo *et al*., 2011 & Amoo, 2002). The few studies that have been carried out in the motor insurance industry have focused mainly on developed economies. In Ghana, the only study conducted in the motor insurance industry investigated factors affecting the choice of a particular type of insurance policy (Awunyo-Vitor, 2012). However one of the most important areas of concern to the insurance industry is the duration of claims by claimants and duration until payment. The relevance of this study is that it will help insurance companies have knowledge on how long it takes for policyholders to report claims which will in turn enable them make preparations towards settlement of claims.

1.2 Objectives of the study

The general purpose of survival analytic techniques is to estimate time to event of a group of individuals, compare event between two or more groups and also to assess the relationship between explanatory variables and time to event. However, the objective of this study is to apply survival analytic techniques to motor insurance to model the time until policyholders report claims after purchasing a policy and time until settlement. In furtherance of this objective, these specific objectives will guide the conduct of the study

I. Determine the best method for constructing confidence intervals for survival estimates for both time until claims are reported and time until payments are made

II. Determine the probabilities that the time until claims are reported and time until claims are settled fall within a specific period.

1.3 Research Questions

1. What is the best method for estimating confidence intervals for the survival estimates for both time until claim reporting and time until claim settlement?
2. What is the Kaplan Meier estimate for time until claims are reported from the date of policy issue and time until payments are made from date claims were reported?

1.4 Significance of the study

The importance of this study is to determine the time until policyholders report claims and time until they get paid, which will in turn help the insurance companies use these estimations of survival functions to make preparations to pay policyholders on time since they already know the estimate within which claims will be made. The findings of this study will impact positively on the productivity of the motor industry in the Ghanaian economy

1.5 Limitations of the study

The main limitation of the study had to do with the nature of the data. The data lacked covariates i.e. age of policyholder, gender of policyholder, age of vehicle, model of vehicle, and sum assured. These would have improved the study. Also there was difficulty in obtaining data.

1.6 Outline of the study

The rest of the study is organized as follows; Chapter two which is the literature review looks at other works in the field of survival analysis conducted by people. Chapter three discusses the methodology employed in carrying out the study. Chapter four contains the well detailed analysis, results and discussion of the study. Chapter five concludes the study with the summary of findings, conclusions and recommendations of the study.

CHAPTER TWO

REVIEW OF LITERATURE

2.0 What is Survival Analysis?

Survival analysis, which can be described as the modelling of time to event data (Lawless, 2003) is extensively applied in many fields such as agriculture (Zavadilova *et al.*, 2009), economics (Gamerman and West, 1987) and many financial sectors (Czado and Rudolph, 2002). The length of time taken for the event of interest to occur is defined as the survival time. The survival function is employed to describe the probability that an individual survives beyond a specified time (Lawless, 2003).

In survival analysis, the subjects experience the event of interest at the end of the study. However, there are instances where some subjects in the study have been observed not to experience the event of interest at the end of the study or time of analysis. Examples of such instances occurs when in a particular study some patients may still be alive or disease-free at the end of the study period or when people are lost to follow-up after a period of study (Lee & Wang, 2003). In such instances, the exact survival times of these subjects are unknown. These are called censored observations or times.

In survival analysis, the dataset can be exact, censored or truncated. There are three types of censoring namely, right, left and interval. Right censoring arises when the event of interest occurs after the observed survival time. In this case the end point is not observed. Left censoring is most likely to occur when the researcher begins observing a sample at a time when some of the individuals may have already experienced the event. In interval censoring, the individual or the material is known to have experienced an event within an interval of time but the actual survival time is not known. In survival analysis, the distribution of survival time may be denoted by T and can be characterized by three equivalent functions namely, survival, probability density and hazard functions. They are as follows:

2.0.1. Survivorship Function (or Survival Function)

This function, denoted by $S(t)$ is defined as the probability that an individual survives longer than t (time until event occurs):

$$S(t) = P(an\ individual\ survives\ longer\ than\ t)$$

$$S(t) = P(T > t)$$

From the definition of the cumulative distribution function $F(t)$ of T,

$$S(t) = 1 - P(an\ individual\ fails\ before\ t)$$

$$S(t) = 1 - F(t)$$

In this case $S(t)$is a non-increasing function of time t with the properties

$$S(1) = 0\ and\ S(0) = \infty$$

That is, the probability of surviving at least at the time zero is 1 and that of surviving an infinite time is zero.

2.0.2 Probability Density Function (or Density Function)

For a continuous random variable, the survival time T has a probability density function defined as the limit of the probability that an individual fails in the short interval t to $t + \Delta t$ per unit width Δt, in other words the probability of failure in a small interval per unit time. It can be expressed as

$$f(t) = \lim_{\Delta t \to 0} \frac{P(an\ individual\ dying\ in\ the\ interval\ (t, t + \Delta t)}{\Delta t}$$

The graph of $f(t)$ is referred to as the density curve. Two main properties of the density function are:

i. $f(t)$is a nonnegative function:

$f(t) \geq 0 \; for \; all \; t \geq 0$

ii. The area between the density curve and the t axis is equal to 1.

2.0.3 Hazard Function

The hazard function $h(t)$ of survival time T gives the conditional failure rate. It is defined as the probability of failure during a very small time interval, assuming that the individual has survived to the beginning of the interval, or as the limit of the probability that an individual fails in a very short interval, $t + \Delta t$, given that the individual has survived to time t :

It can be expressed as

$$\lim_{\Delta t \to 0} \frac{[P(individual \; fails \; in \; the \; interval \; (t, t+\Delta t)] given \; the \; individual \; has \; survived \; to \; t]}{\Delta t}$$

The hazard function can also be defined in terms of the cumulative distribution function $F(t)$ and the probability density function $f(t)$

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

The cumulative hazard function is defined as:

$H(t) = \int_0^t h(x)dx$ or $H(t) = -logS(t)$

There are two general approaches used in survival data which are nonparametric and parametric methods. The two common nonparametric methods used are the Kaplan-Meier estimator derived by Kaplan and Meier (1958) and the Nelson Aalen (1978). In the case of parametric methods theoretical distributions are used for analysis and these include exponential, Weibull, lognormal and log-logistic distributions. Zelen (1966) used the exponential distribution

successfully as the model for survival time in a study of new anti-cancer drugs in the L1210 animal leukemia system.

The Weibull distribution, which was proposed by Weibull (1939) has also been widely applied. An example is that of Pike (1966) who applied the Weibull distribution to a two-group experiment on vaginal cancer in rats exposed to the carcinogen DMBA. The two groups were distinguished by pre-treatment regime. It was realized that the Weibull distribution fitted the data. The lognormal distribution was applied to study chronic lymphocytic and myelocytic leukemia by Feinleib and MacMahon (1960) to analyse survival data of 649 white residents of Brooklyn diagnosed from 1943 to 1952. It was concluded from the study that the three-parameter lognormal distribution adequately describes the distribution of survival times for each subgroup except women with chronic lymphocytic leukemia. The gamma distribution was found to have fitted the data of a study by Brown and Flood (1947) who attempted to describe the life of glass tumblers circulating in a cafeteria. Byers *et al.*, (1988) also employed the log-logistic distribution to describe the rate of spread of HIV between 1978 and 1986 and found based on the Akaike information criterion (Akaike, 1974) that the log-logistic distribution fitted well compared to the Weibull distribution.

The models mostly used in survival data are the semi-parametric Cox proportional hazard model (Cox, 1972), which is a survival analysis regression model, used to express the relation between the event incidence, as expressed by the hazard function and covariates that influence survival time and the Accelerated Failure Time (AFT) model, which is a parametric model that serves as an alternative to the Cox proportional hazards model (Newby, 1988). AFT is the logarithm of the failure time as linear function of covariates incorporated in the model (Kalbfleisch & Prentice, 1980).These models are very useful because they take into account time-varying covariates.

2.1 Survival Analysis and its application

Due its importance to research, many researchers have applied survival analysis in carrying out their studies. Examples of such applications are as follows:

2.2 Survival analysis and economics

Survival analysis has been used by several authors to produce the best predictive models for several economic activities. Xie and Giles (2007) modelled the length of time that it takes for a patent application to be granted by the U.S. Patent and Trademark Office using two major survival analysis techniques namely the nonparametric Kaplan-Meier and parametric accelerated failure time models. The number of claims a patent makes, the number of citations a patent makes, the patent's technological category, and the type of applicant were all found to have significant effects on the duration that a patent takes to be granted. Applying the Accelerated Failure Time framework to estimate parametric hazard functions based on the Exponential, Weibull, Log-Logistic and Log-Normal distributions, it was found that the Log-Normal model was the most preferred among all these distributions.

Using survival analysis to predict retention rates in the US, Sadler and Lang (2006) also found that survival models really fitted the data as compared to using the average rates, which do not make use of information provided by censored cases. The authors observed that the survival model provided the option of using Lower Confidence Limit (LCL), which results in a more precise estimate of the survival rates. In their attempt to investigate credit scoring using macroeconomic variables, Belloti and Crook (2009) discovered that survival analysis was more effective in predicting defaults than logistic regression. This was attributed to the fact that those macroeconomic variables being studied could not be readily inserted in the logistic regression models but were considered as time-varying covariates in the survival models. The model fit was improved by the inclusion of these time- varying covariates such as interest rate and

unemployment which affected the probability of default and as well as provided a statistically significant improvement in the prediction of default on independent test set.

Lobos and Szewcyzk (2012) also applied survival analysis with the objective of finding the factors that determine the survival rate of Polish micro and small enterprises. The pilot study consisted of a sample of 147 entities and the determinants of a company's survival were evaluated to ascertain whether firms with different characteristics showed different performances in terms of the duration of survival. The results indicated significant differences in the survival rates and also revealed that while larger firms are significantly less likely to close than smaller ones, firms were more prone to closure in highly competitive markets.

Gamerman and West (1987) employed dynamic Bayesian models for survival data analysis in their study of factors contributing to unemployment and found that the models can be applied in the prediction of unemployment.

Factors influencing the survival of ski lift companies over the period of 1996-2011 was studied by Falk (2011) using the Cox proportional hazard and competing risk survival models to distinguish between temporary closures and permanent exits. The findings indicated that early adoption of snowmaking at later periods can result in a significantly lower hazard rate.

Cameron and Hall (2003) used survival analysis techniques to investigate the patterns and determinants of mutual fund survivorship in Australia and found that the response to performance seemed asymmetrical, with positive shocks having a larger impact on the hazard rate than negative shocks. The role political, institutional and economic factors played in the exchange rate regime duration of 49 developing countries between 1974 and 2000 was assessed by Setzer (2004) using the Cox model. It was observed that exchange rate regime depended on partisan and institutional incentives such as the political colour of the government in power, the number of veto players and the degree of central bank independence. In addition, the model

was found to be useful since it enabled the influence of various time-varying covariates on the duration of exchange rate regimes to be analysed taking into consideration previous period's regime. Fixed exchange rates were also observed to have a lower hazard than floating or intermediate regimes.

Doghonadze (2012) analysed the determinants of survival of a sample of the Georgian firms on particular export markets. A major finding of his study was that fixed cost of exporting decreases with years of exporting.

Johannsen (2013) conducted a survival analysis in the form of proportional hazard regression on 65 Research and Development (R&D) units in 15 Swedish multinationals between 1991 and 2012. Among the findings of study were evidences suggesting positive associations between age, a high degree of local autonomy over hiring and firing of the unit's R&D professionals, training programs for R&D personnel, salary level for R&D employees, and lastly the ability to produce high quality output of a foreign R&D unit and survival. It was also found that acquisition as mode of establishment, and cumbersome communication between local and parent R&D were negatively associated with survival.

Fu and Wu (2013) explored the patterns and determinants of survival of exports in foreign markets using the survey data of the Chinese manufacturing firms for the period 1998-2007. The methods used by the authors included non-parametric techniques and the estimation of a discrete-time duration model. The findings of their work suggested that the probability of exit was higher for the exporters at the starting period and that large, highly productive and more export-oriented firms were more likely to survive. In addition, foreign ownership was found to be an important determinant of export survival, while state ownership increased the risk of export failure.

Nunes and Sarmento (2010) assessed the post-entry performance of new Portuguese firms by investigating the structural characteristics of the hazard and survival functions, using non-

parametric methods. In order to approach prevalence of some stylized facts and determinants of new firm survival, they produced a new entrepreneurship database, using the administrative data of *Quadros de Pessoal,* following the Eurostat/OECD´s internationally comparable business demography methodology. This allowed the computation of a comprehensive array of entrepreneurship indicators on employer enterprise and survival dynamics in Portugal over a period of 18 years, disaggregated in dimensions such as sectors, regions and size classes. The analysis showed that around 25 per cent of enterprises entering the market failed within the first 2 years of activity and that more than 50 percent failed within a period of six years. Also the instantaneous probability of exit was monotonically decreasing with age. The conditional probability of failure increased continuously up to the sixth year of activity after entry.

Kelly *et al.,* (2014) applied survival analysis to determine the role of credit and the macroeconomy in the distress of Small and Medium scale Enterprises (SME) during a prolonged economic downturn in Ireland which began in 2007/2008. The authors used insolvency as a measure of distress and captured both bank and non-bank forms of credit. A survival analysis of insolvent liquidations was conducted and it was realized that when location and economic sector are controlled, variables which captured a build-up of stress in the macroeconomy and those capturing bank credit standards and availability throughout the cycle were determinants of firm survival.

Besedes and Blyde (2010) applying survival analysis found that although export relationships were in general short-lived, significant differences across regions existed with Latin America exhibiting lower export survival rates than the US, the EU and East Asia, among others. In reanalysing the Pennsylvania Reemployment Bonus Experiments, which were conducted in 1988-89 to examine the effect of different types of reemployment bonus offers on the unemployment spell, Schunk (2003) fitted a Cox-proportional-hazards survival-model to the data and the results were compared to the results of a linear regression approach and to the

results of a quantile regression approach. The Cox-proportional-hazards model provided for a remarkable goodness of fit and yielded less effective treatment responses therefore lower expectations concerning the overall implications of the Pennsylvania experiment. An influence analysis was also proposed for obtaining qualitative information on the influence of the covariates at different quantiles. The results of the quantile regression and of the influence analysis showed that both the linear regression and the Cox-model still imposed stringent restrictions on the way covariates influence the duration distribution. However, due to its flexibility, the author found the Cox-proportional hazards model to be more appropriate for analysing the data.

Although survival analysis has proved a useful tool in analysing economic data, there have been occasions where it has been inappropriate. For instance Etzioni *et al.,* (1999) used survival analysis to analyse medical costs and observed that survival analysis approaches are not generally appropriate for the analysis of medical costs.

2.3 Survival Analysis and actuarial science

Survival analysis is inevitably part of the actuarial science. As a result of this, there have been researches on the application of survival analysis in actuarial science. They include:

2.3.1 Survival Analysis and Business/Bank failure prediction

The application of survival analysis in bank failure prediction was pioneered by Lane *et al.,* (1986) who found survival analysis to be a better technique than the discriminant analysis for predicting insolvency of banks two years in advance. Similar results were also arrived at by Crapp and Stevenson (1987) who applied cox models to Australian Credit Unions. Halling and Hayden (2006) compared the efficiency of a two-step survival time approach in predicting bank failure in Austria. The models used were discrete logit model with survival time dummies and multi-period logit model. The study revealed that the two-step approach outperformed the one-

step logit models in predicting failure and concluded that the two-step approach might be of importance in assessing the health status of a bank. It was also realized that the performance advantage of the two-step approach was mainly due to the estimation procedure. In the study it was also found that macroeconomic variables played no role in predicting default in the at-risk samples. The determinants of bank fragility of Islamic and conventional banks were established by Pappas (2010) who observed Islamic banks were highly fragile than commercial banks and that the hazard functions of both banks were influenced by different factors in different degrees or proportions with equity-to-liabilities being identified as important. Again the study showed that the inclusion of macroeconomic variables led to models with higher statistical significance.

Janot (2001) compared survival models with other models and found out that, the model estimated by analysis of survival obtained a better result in classifying a bank as a solvent or insolvent at a time frame of six months prior to bankruptcy. Overall, there have been several studies on the application of survival analysis in predicting bank failure using cox models. An example can be found in Wheelock and Wilson (2000) who used the Cox proportional hazard model with time-varying covariates estimated by partial likelihood, to identify specific factors that explain time to bank failures during 1984-1993. The use of cox proportional hazard models in survival analysis for development of models for the prediction of bank insolvency was also adopted by Whalen (1991) and Rocha (1999).

Lee (2014) studied business bankruptcy prediction based on survival analysis approach of some companies listed on the Taiwan Stock Exchange between the years 2003-2009. Cox proportional hazard model was used to assess the usefulness of the traditional financial ratios and market variables as predictors of the probability of business failure to a given time. The study revealed that, many ratios were not needed to be able to anticipate potential business bankruptcy. Luoma and Laitinen (1991) also compared survival analysis with discriminant

analysis and logistic regression in predicting business failure using a sample of 36 failed companies (24 from industrials and 12 retailing firms) each paired with a not failed company belonging to the same business and of similar size. From the results, the percentage of correct classifications was 61.8%, 70.6%, and 72.1%, for survival analysis, discriminant analysis and logistic regression, respectively.

Glennon and Nigro (2005) measured the default risk of small business loans using a survival approach. Employing a discrete hazard model the study revealed that the likelihood of default is conditional on borrower, lender, loan characteristics and changes in economic conditions. Klos (2008) used survival methods in analysing the negative performance and probability of failure of a sample of Ukrainian joint companies over a seven-year period i.e. 1999-2006. Using the discrete nonparametric proportional hazard models, it was observed that the length of the spell of negative performance as well as firm's characteristics should be taken into consideration when assessing the performance of firms. Henebry (1997) used both cash flow and non-cash flow proportional hazards models to test for stability of the models over time. Different time horizons and start dates were used to test stability over a five-year period. The results from the study indicated that, none of the specific formulations were stable across different starting dates nor across different horizons for the same starting date. In addition, forecast models were further used to test stability and only three variables namely, Primary Capital to Total Assets (PCTA), Nonperforming Loans to Total Loans (NPLTL) and Total Loans to Total Assets (TLTA) were found to be consistently useful in predicting bank failure.

Mannaso and Mayes (2009) used the discrete survival model to explain bank distress in Eastern European transition economies and established that it was possible to find bank specific, bank sector structure and macroeconomic variables that were able to predict vulnerabilities in the European transition countries' banking sector.

Pereira (2014) proposed a survival model for the prediction of corporate bankruptcy based on survival analysis. In his research, the hazard rate is the probability of ‘‘bankruptcy’’ as of time *t*, conditional upon having survived until time *t*. Based on the results obtained from the sample used, the method offered a good perspective when used for the development of forecasting models in the bankruptcy research field. Cabo and Rebelo (2010) evaluated risk of insolvency of Agricultural Credit Co-operatives, CCAM (Caixas de Crédito Agrícola Mútuo). To achieve their goal, the authors adopted a Cox proportional hazards model in analysing the CCAM failures in the period between 1995 and 2009. The model showed that transformation ratio and other structural costs ratio were important indicators to evaluate the relative risk of insolvency for CCAM.

Notwithstanding the application of survival analysis to bank and business failure, it can be applied in various sections in the banking and business field. An example of such applications is the work of Stefancic (2014) who applied survival analysis to examine the turnover of top managers in both commercial and non-commercial banks in Italy and found that the judicial system of banks had a significant relationship with management turnover since top managers in cooperate banks showed a higher survival probability. Banks’ history and institutional legacy also had a significant effect on both management turnover and on the disciplinary mechanisms for top managers. Several models including the survival model were used by Musakwa (2013) to measure bank funding liquidity risk. It was realized that it is possible to use the survival model to assign cash-flows to future time horizons.

In his study, Caree (2003) used hazard rate analysis which is another survival technique to investigate the determinants of hazard rates of banks active on the Moscovian deposits market between the periods of 1994- 1997. The study showed that market share and duration have negative effects on the hazard rate whereas the deposit interest had a positive effect.

Dabos and Escudero (2004) also examined the Argentinian banking system using survival analysis and accounting data and found evidence that increased profitability and liquidity reduces the hazard of the bank.

2.3.2 Survival analysis and loan data

Survival analysis has also been applied to determine rate of default on loans. In a study on personal loans in the United Kingdom, Stepanova and Thomas (2010) used three extensions of cox proportional hazard models. Survival analysis in this study was found to be a very useful technique because the diagnostic methods used to check the adequacy of the models proved that they fit the data.

Cao *et al.*, (2009) in Spain applied the regression model, the regression linear models under censoring and a nonparametric kernel estimation using product-limit conditional distribution function estimator by Beran to carry out survival analysis on loan. Their study concluded that survival analysis methods can be used to model credit quality in terms of lifetime of loans. In the study of Atsmegiorgis (2014), loan repayment rate of customers of Hawassa district commercial bank was studied using survival analysis. The author used a sample size of 182 customers, who took loan from October, 2005 to April, 2012 from the bank record. The Kaplan-Meier estimation method and Cox proportional hazard model were applied to model the survival repayment time as well as examine the association between the survival time with different demographic and loan characteristics variables respectively. Results from Kaplan-Meier estimation showed that the loan repayment rate is significantly related with loan size, loan type, and previous loan experience, purpose of loan, educational level and type of collateral offered.Survival analysis was also applied by Witzany *et al.*, (2012) to model Loss Given Default (LGD). The study compared four models namely linear regression, logistic regression, survival and pseudo survival to estimate the future recovery rates and LGD's. It

was found that, the pseudo cox models based on minimization of squared differences on best known recovery rates outperformed all other models.

On the other hand, Zhang and Thomas (2012) employed survival analysis to compare between single and mixed distribution models for modelling Loss Given Default of credit systems. The major techniques employed in the study were linear models and survival analysis. In the comparison of single distribution models, the study revealed that linear regression was better than survival analysis in most cases. Linear regression models was found to have higher R-square and Spearman rank coefficient as compared to survival analysis models for recovery rate modelling.

2.3.3 Survival Analysis and insurance

Survival analysis is inevitably part of the actuarial science particularly insurance. As a result of this, there have been researches on the application of survival analysis in actuarial science particularly insurance. An example is that of Brockett *et al.,* (2008) who used logistic regression and survival analysis techniques to assess the probability of total customer withdrawal, and the length of time between first cancellation and subsequent customer withdrawal. The results showed that, cancellation of one policy was a very strong indicator that other household policies will be cancelled. It was further revealed that the insurer can have time to react to retain the customer after the first cancellation. However the time to react to retain customers was significantly dependent on the method used to contact the company, household demographics, and the nature of the household's insurance policy portfolio.

Czado and Rudolph (2002) used the cox proportional hazard model to estimate intensities in large claim portfolio. They found this technique to be more reliable than the Poisson approach followed by Renshaw and Haberman (2005).

Conley (2013) employed Kaplan-Meier method of estimation in carrying out his study and found it to be an accurate and efficient method to model abandonment in a discrete event

simulation model. The inclusion of both the answered and abandoned observations in the analysis also gave a full picture of the actual patience demonstrated by callers.

Chow *et al.,* (2003) used survival analysis in carrying out their study of Home Equity Conversion Mortgage loans. The authors found that the use of survival analysis techniques helps in the understanding of Home Equity Conversion Mortgage (HECM) loans and their cash flows, enabling the program to grow and to be attractive to lenders and future investors. Beirlant *et al.,* (1991) applied an Accelerated Time Failure (AFT) model to explain the claim size in the statistical risk evaluation of the Belgian car insurance and found it to be a very useful model.

2.4 Survival Analysis in agriculture and forestry

Survival analysis has also been utilized to analyse events in studies involving agriculture and forestry**.** Zavadilova *et al.,* (2009) applied survival analysis in their study of 47786 Czech Fleckvieh cows first calved from 1994 to 2003 and observed that a relationship existed between phenotypic type traits and the functional survival of cows. In a study aimed at determining whether survival analysis resulted in a better genetic evaluation of female fertility and mastitis traits and also to assess the effects of mastitis and pregnancy status as risk factors for culling, Schneider (2006) realized that survival analysis was a useful method for such purposes.

Carlén *et al.,* (2004) compared the Weibull proportional hazard which is a survival model with linear model based on binary data in their study of genetic evaluation of mastitis of Swedish cows. The findings indicated that survival analysis is advantageous than linear models in clinical mastitis. This was attributed to the fact that more information was used in survival models therefore increasing precision and also the culled cows were treated appropriately hence reducing the bias.

Kamleh *et al.,* (2012) in an attempt to estimate the shelf-life of stored Halloumi cheese used survival analysis and considered consumer rejection as a failure index. The study concluded

that there was the need for adherence to good manufacturing practices and maintenance of low temperatures during the storage and distribution of the packaged Halloumi cheese. Vance and Geoghegan (2002) estimated a spatially explicit model of the forest clearance process among smallholder farmers in an agricultural frontier of southern Mexico. Their analysis took as its point of departure a simple utility-maximizing model that suggested many possible determinants of deforestation in an economic environment characterized by missing or thin markets. Hypotheses from the model were tested on a dataset that combined a time series of satellite imagery with data collected from a survey of farm households whose agricultural plots were geo-referenced using a Global Positioning System (GPS). Survival analysis was used to identify the effect of household level explanatory variables on the probability of deforestation. This approach allowed for the introduction of a measure of the time until clearance as a covariate, thereby affording a control for the effect of potentially important explanatory variables that varied through time but were not directly observable. The results suggested that the deforestation process is characterized by non-linear duration dependence, with the probability of forest clearance first decreasing and then increasing with the passage of time.

Wang *et al.*, (2010) explored the appropriateness of survival models for crop insurance program design. The results of the study indicated that the estimated premium rates for each crop were consistent with the currently prevailed crop insurance premium rate in Panjin.

In the area of forestry a couple of studies have also been carried out using survival analysis. Woodall *et al.*, (2005) employed survival analysis in analysing tree mortality in forest inventories. The authors found out that survival analysis techniques facilitated by diameter at breast height and diameter growth classes of the tree may provide foresters with the ability to test tree mortality hypotheses and summarize regional tree mortality trend.

Greenberg *et al,* (2005) also observed that survival analysis can be used in assessing current and future trends in deforestation rates and to investigate the impact of spatial, cultural and economic factors on deforestation in Neotropical rainforest using multi temporal satellite imagery. In their study it was concluded that with the rate of deforestation as at 2005 being 0.11% with an increasing rate of with time, 50% of the forest within 2 km of an access road will be lost by the year 2063 due to unhindered colonization and anthropogenic conversion.

2.5 Survival analysis and medicine

In a study with 176 patients with heamato-oncological diagnoses who had undergone bone marrow blood transplant, Langova (2008) found survival analysis methods to be useful in estimating survivor functions, comparing survivor functions and assessing the relationship of the explanatory variables and time. Abada *et al.,* (2001) examined modern and traditional factors that lengthen or shorten the duration of breastfeeding in a sample of women from Philippines. Some of the factor in the study included health sector, socio-economic and supplementary foods. The Cox Proportional Hazard model was employed in the study for the analysis of breastfeeding. The results showed that traditional factors associated with breastfeeding do not play any significant role whereas factors associated with modernity do. The findings of the study further suggested that the health institutions and medical professionals can play a significant role in promoting breastfeeding in the Philippines. In carrying out a clinical trial of 30 cervical cancer patients, Zaman and Pfeiffer (2012) compared the log-rank test to the Wilcoxon and other tests. It was found from the study that, although the log-rank test was very good, it was better to use the weighted test in certain instances because they give more satisfactory results.

Even though survival analysis has widely been applied in the field of medicine with positive results, there are however negative results. An example is the study by Franco *et al.*,(2005) who compared the performances of Cox proportional hazard model and approach based on

artificial neural networks constructed for the prognosis of outcome in patients with primary breast cancer. The data used was from 32 hospitals from Spain. The results from the study indicated that the neural network predictions were much more accurate particularly in the early months after surgical intervention

CHAPTER THREE

METHODOLOGY

3.0 Introduction

This chapter gives a thorough report of the data and the statistical approach employed. A detailed consideration of the theoretical background of the Kaplan-Meier method of survival functions is presented followed by the derivation of the Kaplan-Meier estimates of the dataset as well as construction of confidence intervals (linear and log transformation) of essential survival estimates.

3.1 Materials and Methods

In order to accomplish the objectives set, secondary data was collected from an insurance company in Ghana. The data consists of time until reporting a claim after policy becomes effective, time until settlement as well as the various types of classes in the policies underwritten. Thus the data gives a vivid description of the entire portfolio of policies underwritten by the insurer. The data collected covered the period January, 2010 to December, 2012. The data analysis was divided into mainly two parts; exploratory data analysis and further analysis. This analysis was done using Microsoft Excel software.

3.2 Data modifications

The data obtained for the study is inherently incomplete due to censoring and truncation. These observed characteristic is however expected for any actuarial dataset on claim settlement. The data contained information on claims reported and claims settlement observed between the periods of January 2010 to December 2012. The study primarily encountered censoring from below with some few notable data elements being censored form above. Regarding our study, censoring from below (left censoring) at d (i.e. January 2010) covers claims reported prior to d and settled after d but before u i.e. December 2012. All claims outstanding as at December 2012 constitute truncation from above (right truncation). If the claim is outstanding at the end

of the observation data, the date of settlement is not known, but it is known that it is at least as large as the sum insured or policy limit allowable on the policy. Due to the nature of the data, nonparametric techniques were used instead of the parametric. The reasons for using the nonparametric techniques for the analysis are because: (i) the population followed suitable distribution (ii) the scale of the data was ordinal (i.e. dataset was ordered). The nonparametric method used is the Kaplan Meier estimation.

3.3 Nonparametric Methods

Nonparametric methods have turned very popular within survival analysis (Hougaard, 2000). One of such reasons is due to censoring. Nonparametric methods are very efficient when the population under study follows no distribution. Nonparametric methods are suggested to analyse survival data before attempting to fit a distribution (Siegel, 1957). Estimates from the nonparametric methods and graphs are very helpful in an attempt to find a model to fit the data. One of the most widely used nonparametric methods for handling incomplete dataset due to censoring and truncation is the Kaplan Meier method.

3.4 The Kaplan-Meier Method

The Kaplan-Meier (KM) estimator is the most extensively used technique for estimating survival functions in the area of biomedicine (Allison, 1995). The Kaplan-Meier (KM) estimator is also known as the product limit estimator. Survival data analysis is considered by many researchers as the application of two orthodox statistical methods to a special type of problem: parametric if the distribution of survival times is known to be normal and nonparametric if the distribution is unknown. This assumption would be true if the survival times of all the subjects were exact and known; however, some survival times are not. Additionally, the survival distribution is often skewed, or far from being normal hence the need for new statistical techniques. One of the most important developments is due to a special feature of survival data in the life sciences that occurs when some subjects in the study have

not experienced the event of interest at the end of the study or time of analysis. For example, some patients may still be alive or disease-free at the end of the study period. The exact survival times of these subjects are unknown. These are called censored observations or censored times and can also occur when people are lost to follow-up after a period of study. When these are not censored observations, the set of survival times is complete. Therefore the Kaplan Meier estimate is a major technique which makes room for estimating censored data or observations. If the data were not censored, the obvious estimate would be the empirical survival function

$$\hat{S}(t) = \frac{1}{n}\sum_{i=1}^{n} I\{t_i > t\} \tag{3.1}$$

Where I is the indicator function that takes the value 1 if the condition in braces is true and 0 if otherwise. The estimator is simply the proportion alive at t. In cases of single right censoring, that is to say if all censored cases are censored at the same time say c, and all the observed events time are less than time c, the situation is pretty clear-cut. Kaplan and Meier (1958) extended the estimate to censored data. Let $t_{(1)} < t_{(2)} < \ldots < t_{(m)}$ denote the distinct ordered times of death (not counting censoring times). Let d_i be the number of deaths at $t_{(1)}$, and let n_i be the number alive just before $t_{(1)}$. This is the number exposed to risk at time $t_{(1)}$. Then the Kaplan- Meier or product limit estimate of the survivor function is

$$\widehat{S}(t) = \prod_{i:t_i<t}\left(1 - \frac{d_{(i)}}{n_{(i)}}\right) \tag{3.2}$$

A heuristic justification of the estimate is as follows. To survive to time t you must first survive to $t_{(1)}$. You must then survive from $t_{(1)}$ to $t_{(2)}$ given that you have already survived to $t_{(1)}$ and so on. Because there are no deaths between $t_{(i-1)}$ and $t_{(2)}$, we take the probability of dying between these times to be zero. The conditional probability of dying at $t_{(1)}$ given that the

subject was alive just before can be estimated by $\frac{d_{(i)}}{n_{(i)}}$. The conditional probability of surviving time $t_{(1)}$ is the complement $1 - \frac{d_{(i)}}{n_{(i)}}$. The overall unconditional probability of surviving to t is obtained by multiplying the conditional probabilities for all relevant times up to t. The Kaplan-Meier estimate is a step function with discontinuities or jumps at the observed death times. The Kaplan Meier estimator is a step function with jumps at the observed event times. The size of these jumps depends not only on the number of events observed at each event time t_i, but also on the pattern of the censored observations prior to t_i. The main functions of survival analysis used in the Kaplan Meier methods are the survivorship function, probability density function and hazard function.

3.5 Model Specification

For the available individual data we outline three required parameters: First, the truncation point for the observation. Suppose we let the value of this parameter be d_i for the i^{th} observation then if there is no truncation, then $d_i = 0$. The second required parameter is the observation itself. Two notations are outlined to completely describe this parameter depending on whether or not that observation was censored. The observation value is reported as λ_i if not censored (where rejected claims constitute censored claims with respect to the study) and γ_i if censored. Placing more emphasis on the uncensored observations we let $y_1 < y_2 < \cdots < y_k$ be the k unique values of the $x_i's$ that appear in the sample where k must be less than or equal to the number of uncensored observations. $\boldsymbol{y}$ is a of vector respective times (dates) until payment after reporting. Further, we let s_i be the number of times the uncensored observation y_i appears in the sample. The third parameter of interest is the risk set at the i^{th} ordered observation date y_i and is denoted r_i. The risk set comprises policies under observation at any respective data. As at the observation date, it includes all outstanding claims, all paid/settled claims at that data

or later and all censored observations at the date or later. Hence we specify the risk set r_i retrospectively as

$$r_i = (number\ of\ \lambda_i's < y_i) - (number\ of\ d_i's \leq y_i) - (number\ of\ c_i's > y_i) \qquad (3.3)$$

Equation (3.3) conceptualizes that the model includes all claims which enter the study prior to the given date less those already paid and/or settled claims $(d_i's)$ and those reported after y_i. Prospectively, the risk set can be determined as follows

$$r_i = (number\ of\ d_i's \geq y_i) + (number\ of\ c_i's \geq y_i) - (number\ \lambda's \geq y_i) \qquad (3.4)$$

The risk set is the number of claims observed to be outstanding (unpaid) at the date y_i with some observed loss amount.

3.5.1 Formulating the Kaplan-Meier Product-Limit Estimator for time until claim

To formulate the Kaplan-Meier product-limit estimator, we begin with a simplified assumption that $S(0) = 1$ to indicate that all policies within the study are active prior to y_1. Suppose X is a continuous random variable with density function defined as the number of days until a claim is reported after a policy has become effective. Then X can be modeled as a probability function of the form

$${}_k^tP = P(k \leq X \leq k + t) = \int_k^{k+t} f(x > k | x \leq k + t)dx,\ t = 1, 2, \ldots \qquad (3.5)$$

the probability that the claim is reported by the t^{th} day upon becoming effective or issuing on the k^{th} date. It captures the probability that a claim is reported after some terminal number of days (t) based on empirical evidence. Thus, t is a random variable which is modelled as a discrete case stochastic process using the Kaplan-Meier Product limit estimator. Suppose r_i is the number of effective policies available just before y_i and of these, suppose λ_i of them have

already reported claims, then, the probability that an effective claim remains outstanding past y_i is given by

$$S(y_i) = \frac{r_i - s_i}{r_i}, i = 1,2, \dots ,92 \tag{3.6}$$

where $S(y_i)$ is a survival function which remains until y_{i+1} at which time

$$S(y_{i+1}) = S(y_i)\left(\frac{r_{i+1} - s_{i+1}}{r_{i+1}}\right) \tag{3.7}$$

we obtain a generalized formulation of the above stochastic process as

$$\hat{S}(t) = \begin{cases} 1 & 0 \leq t < y_1 \\ \prod_{i=1}^{j-1}\left(\frac{r_i - s_i}{r_i}\right), & y_{j-1} \leq t < y_j \ , j = 2, \dots, k \\ \prod_{i=1}^{k}\left(\frac{r_i - s_i}{r_i}\right) or\ 0 & t \geq y_k \end{cases} \tag{3.8}$$

if $s_k = r_k$, then $S(t) = 0 \ \forall \ t \geq y_k$.

3.5.2 Formulating the Kaplan-Meier Product-Limit Estimator for time until payment

To formulate the Kaplan-Meier product-limit estimator, we begin with a simplified assumption that $S(0) = 1$ to indicate that all policies within the study are active prior to y_1. Suppose X is a continuous random variable with density function defined as the number of days until a claim is paid after reporting a claim event. Then X can be modelled as a probability function of the form

$${}_k^t P = P(k \leq X \leq k + t) = \int_k^{k+t} f(x > k | x \leq k + t)dx, \ = \ 1, 2, \dots \tag{3.9}$$

the probability that the claim is paid by the t^{th} day upon report on the k^{th} date. It captures the probability that a claim is paid after some terminal number of days (t) based on empirical evidence. Thus, t is a random variable which makes X inherently random. $X \ and \ t$ are

modelled as a discrete case stochastic process using the Kaplan-Meier product- limit Estimator. Suppose r_i is the number of reported claims (outstanding) available just before y_i and of these, suppose λ_i of them have already been settled, then, the probability that an effective claim remains outstanding past y_i is given by

$$S(y_i) = \frac{r_i - s_i}{r_i}, i = 1,2 \dots ,92 \tag{3.10}$$

a survival function which remains until y_{i+1} at which time

$$S(y_{i+1}) = S(y_i)\left(\frac{r_{i+1} - s_{i+1}}{r_{i+1}}\right) \tag{3.11}$$

we obtain a generalized formulation of the above stochastic process as

$$\hat{S}(t) = \begin{cases} 1 & 0 \le t < y_1 \\ \prod_{i=1}^{j-1}\left(\frac{r_i - s_i}{r_i}\right), & y_{j-1} \le t < y_j \ , j = 2, \dots, k \\ \prod_{i=1}^{k}\left(\frac{r_i - s_i}{r_i}\right) or\ 0 & t \ge y_k \end{cases} \tag{3.12}$$

if $s_k = r_k$, then $S(t) = 0 \ \forall \ t \ge y_k$.

3.5.3 Derivation of the survival curves

After deriving the company specific survival functions in equations 3.8 and 3.12, survival probabilities were used to obtain the overall survival function. The hazard functions for time until payment and time claims were reported and time until payments were also computed.

3.5.4 Estimating the mean, variances and interval estimation

When working with complete data, calculating the empirical estimates is pretty straightforward. Since the data in this study is incomplete due to censoring, counts no longer have binomial distribution and therefore the distribution of the estimator is difficult to obtain.

Consider the Kaplan-Meier estimate $S(t)$. It is the product of the number of terms of the form $\frac{r_i - s_i}{r_i}$ where r_i is viewed as the number of policies available to be paid or number of claims available to made at date y_i and the number who actually did so. The mean or Kaplan-Meier estimate is thereby given as

$$\hat{S}(t) = \prod_{t_i<t}\left(\frac{r_i - s_i}{r_i}\right) \tag{3.13}$$

Since the value of $S(t)$ is conditional on being alive at y_0 and also $\frac{r_i - s_i}{r_i}$ is an estimate of $\frac{S(t_i)}{S(t_{i-1})}$

then the variance of the Kaplan-Meier estimate is given by

$$Var\left[\hat{S}(t)\right] = \hat{S}(t)^2 \sum_{i=1}^{j} \frac{s_i}{r_i(r_i - s_i)} \tag{3.14}$$

The Product-Limit estimator which is an efficient means of estimating the survival function for right-censored data can also be used to estimate the cumulative hazard function. It is given by

$$H(t) = -In[S(t)] \tag{3.15}$$

3.5.5 Constructing Pointwise confidence intervals for survival function

The Product-Limit estimator provides a summary estimate of the mortality experience of a given population. The corresponding standard error provides some limited information about the precision of the estimate. The intervals are constructed to assure, with a given confidence level $1-\alpha$ that the true value of the survival function, at a predetermined time t, falls in the interval we shall construct.

3.5.5.1 Linear confidence interval

The most commonly used $100\times(1-\alpha)\%$ confidence interval for the survival function at time to, termed the linear confidence interval, is defined by

$$\hat{S}(t) \pm Z_{1-\propto/2}\sqrt{Var[\hat{S}(t)]} \tag{3.16}$$

where the $Z_{1-\propto/2}$ is the $1-{}^{\propto}/_{2}$ percentile of a standard normal distribution. Construction of the linear confidence intervals follows directly from the asymptotic normality of the Product-Limit.

3.5.5.2 Log-transformed confidence interval

Let $\theta = In\left[In\,\hat{S}(t)\right]$. This implies that $\hat{S}(t) = \exp(-\exp(\theta))$. Then the $100\times(1-\alpha)\%$ confidence interval for θ is given by

$$In[-In\,\hat{S}(t)] \pm Z_{1-\propto/2}\frac{\sqrt{Var[\hat{S}(t)]}}{\hat{S}(t)In\,\hat{S}(t)} \tag{3.17}$$

Since $\hat{S}(t) = \exp(-\exp(\theta))$ the $100\times(1-\alpha)\%$ log-transformed confidence interval for the survival function at t is given by

$$\left[\hat{S}(t)^{1/u}, \hat{S}(t)^{u}\right] \text{ where } u = Z_{1-\propto/2}\frac{\sqrt{Var[\hat{S}(t)]}}{\hat{S}(t)In\,\hat{S}(t)} \tag{3.18}$$

3.5.6 Interval estimates of percentiles

The Kaplan-Meier estimator can also be used to provide estimates of quantiles and percentiles of the distribution of the time-to-event distribution. Recall that the pth quantile of the distribution of X is the smallest x_p so that $S(x_p) \le 1-p$ i.e. $x_p = \inf\{t : S(t) \le 1-p\}$.

A $100\times(1-\alpha)\%$ confidence interval for x_p, based on the linear confidence interval, is the set of all time points t which satisfy the following condition:

$$-Z_{1-\propto/2} \le \frac{\hat{S}(t)-1-p}{\sqrt{Var[\hat{S}(t)]}} \le Z_{1-\propto/2} \tag{3.19}$$

A $100\times(1-\alpha)\%$ confidence interval for x_p, based on the log transformed confidence interval, is the set of all time points t which satisfy the following condition:

$$-Z_{1-\propto/2} \leq \frac{\left[In\{-In[\hat{S}(t)]\} - In\{-In[1-p]\}\right][\hat{S}(t)In[\hat{S}(t)]}{\sqrt{Var[\hat{S}(t)]}} \leq Z_{1-\propto/2} \qquad (3.20)$$

CHAPTER FOUR

DATA ANALYSIS AND DISCUSSION

4.0 Introduction

This chapter gives a detailed description of the analysis carried out in the study. The Kaplan-Meier estimate was applied in this study to analyse various auto insurance contracts in Ghana from January 2010 to December 2012. The Kaplan-Meier estimates for time until claims are reported by insurers and time it takes for payment to be settled are obtained. In addition to that, linear and log transformed confidence intervals for the various percentiles of the Kaplan-Meier estimate for time until claims are reported by insurers and the time it takes for payment to be settled are also obtained. The probabilities of insurers reporting claims and being settled within a specific time period are also obtained.

4.1 Descriptive

Table 4.1: Descriptive statistics for time until reporting and settlement of claims

Variable	Number	Minimum	Maximum	Mean	Standard Deviation
DUC	92	1	532	190.52	124.53
DUP	92	2	500	132.91	141.40

*****DUC =days until claims are reported and DUP= days until claims are settled/paid

The data analysis process begun with computation of summary statistics of days until claims were reported and settled. This summary helped in identifying relevant characteristics of the data. It was observed from the table that, the mean number of days it took for claims to be reported was relatively higher than that settlement of payments (i.e. 190.52 and 132.91). The maximum time it takes for claims to be settled was lower (500) than the time took for claims to be reported (532) indicating that the time from policy issue to reporting claims was relatively longer than time from reporting claims to settlement of claims.

Table 4.2: Distribution of policyholders by time settlement period

Duration(Months)	Frequency	Percentages
less than 3 months	59	64
3-6 months	13	14
6-12 months	10	11
More than 12 months	10	11
Total	**92**	**100**

With regards to the time it took until payments were made to policyholders, 64% of the insurers had their claims settled within the first three months of issue, 11% had their payments honoured more than year with the remaining 25% being settled between three to twelve months

Table 4.3: Distribution of policyholders by claim reporting period

Duration(Months)	Frequency	Percentages
less than 3 months	28	30.4
3-6 months	15	16.4
6-12 months	44	47.8
More than 12 months	5	5.4
Total	**92**	**100**

From the data presented in Table 4.3, it was observed that, more than fifty percent of the policyholders reported claims from six months after the policy was issued.

4.2 Analysis on time until claims were reported

This section addresses all the analysis conducted on time until claims were reported by policyholders.

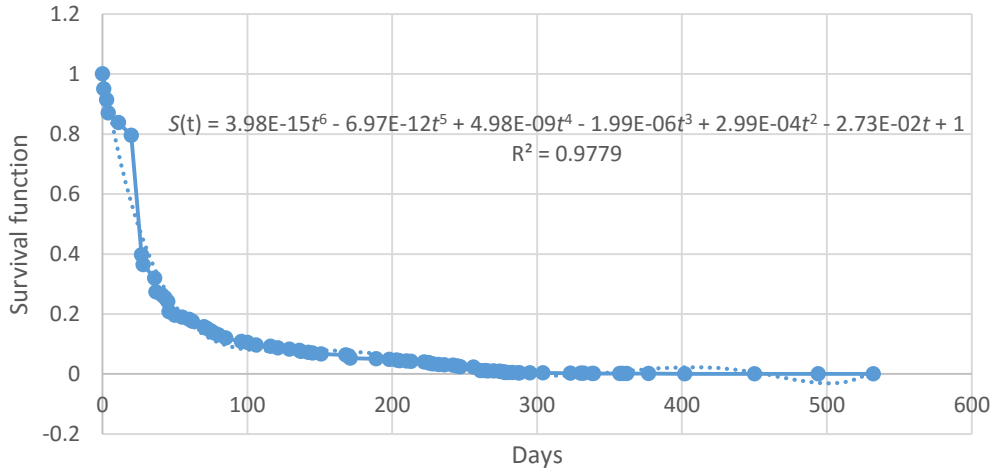

Figure 4.1: Kaplan Meier curve for time until claims were reported

The survival function in the above diagram is given as

$S(t) = 3.98E\text{-}15t^6 - 6.97E\text{-}12t^5 + 4.98E\text{-}09t^4 - 1.99E\text{-}06t^3 + 2.99E\text{-}04t^2 - 2.73E\text{-}02t + 1$

This functions is in conformity with the property $S(0)=1$ and $S(\infty)=0$ (Klein and Moeschberger,2003) because $S(t) = 1\ when\ t = 0$. The *r-squared* value of 0.9779 shows how well the model fitted the data. The Kaplan-Meier estimate for time until claims were reported by policyholders was 0.0009 (See Appendix II).

4.2.1 Estimating the cumulative hazard function for time until claim

In determining the cumulative hazard functions graphs of the various hazard functions were constructed. The following figures give an overview if the hazard functions at various time frames.

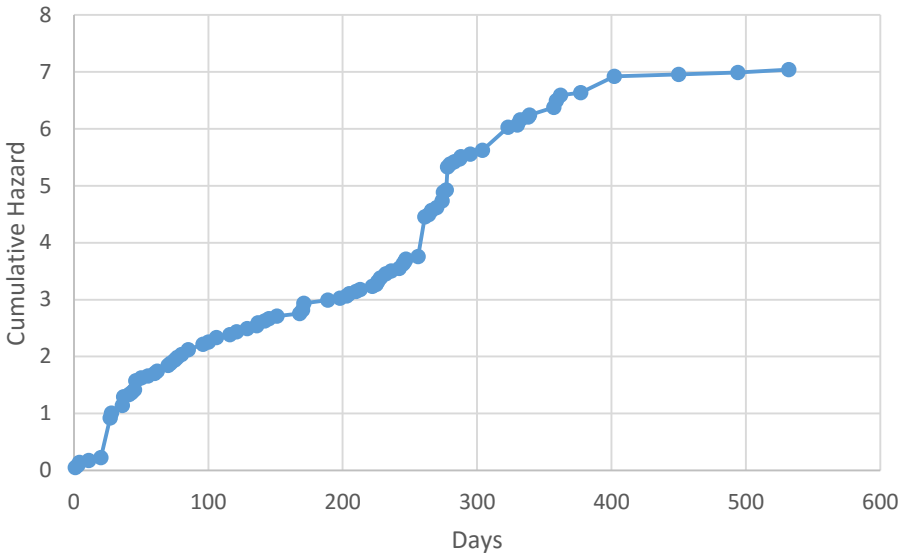

Figure 4.2: Cumulative hazard function for days until claims were reported

The hazard function is a monotonic increasing function with breaks at various intervals. It was observed from the figure above that the cumulative hazard function for time until payment has breaks at various time periods. The time periods are 1-4 days, 4-80 days, 80-256 days and 256-532 days. In attempt to determine the cumulative hazard functions for time until payment, the graphs of the various breaks were drawn to determine the function at interval (break). Figures 4.3, 4.4, 4.5 and 4.6 are constructed.

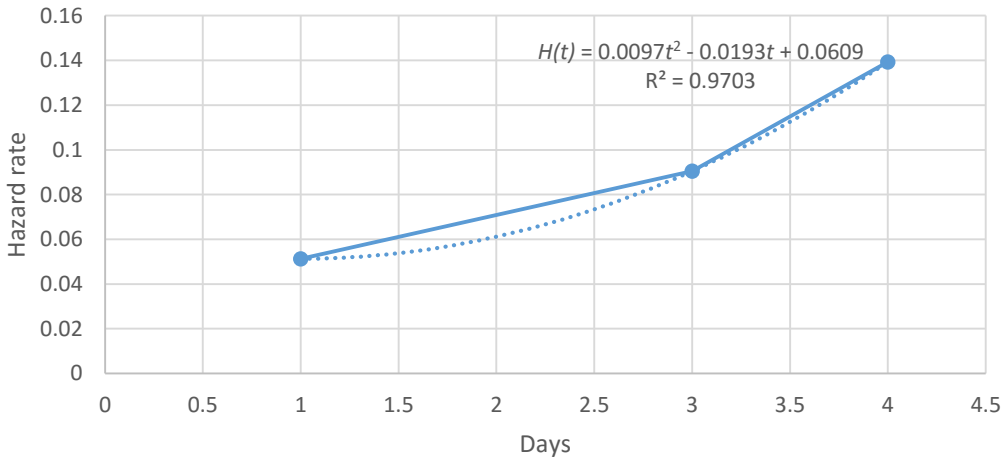

Figure 4.3: Hazard function for specific time period

Figure 4.3 shows the first interval of the cumulative hazard function. This functions ranges from 1-4 days. This function is a quadratic function with an *r-squared* value of 0.9703 indicating how well the model fits the data

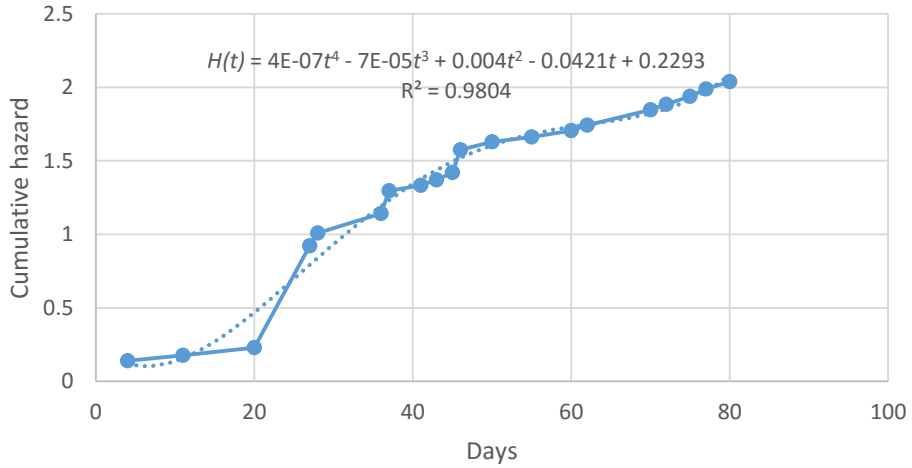

Figure 4.4: Cumulative graph function for specific days

The figure above is the second interval of the cumulative hazard function. It ranges from 4-80 days. This is a polynomial function of order 4. Its *r-squared* value of 0.9804 shows indicates a good fit.

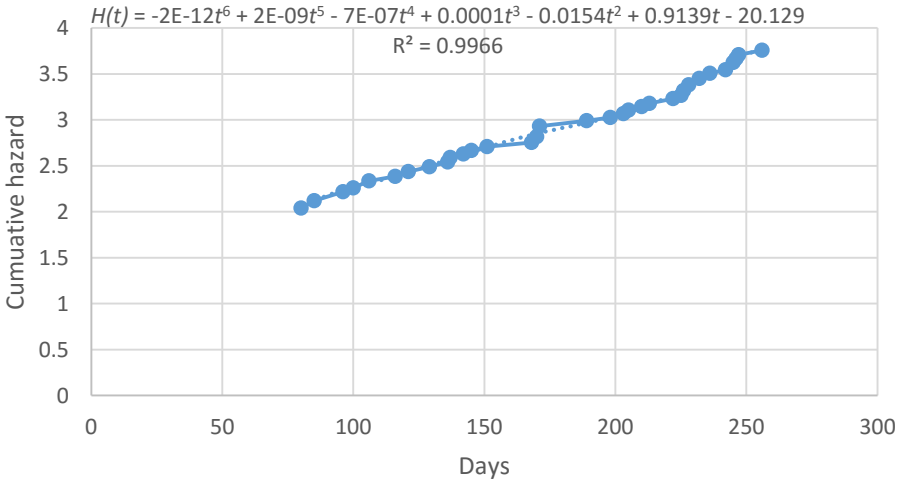

Figure 4.5: Cumulative hazard graph for specific days

The third range of the cumulative function is shown in Figure 4.5. It is shown from the graph that a polynomial of order 6 fit the data well since it has an *r-squared* value of 0.9966

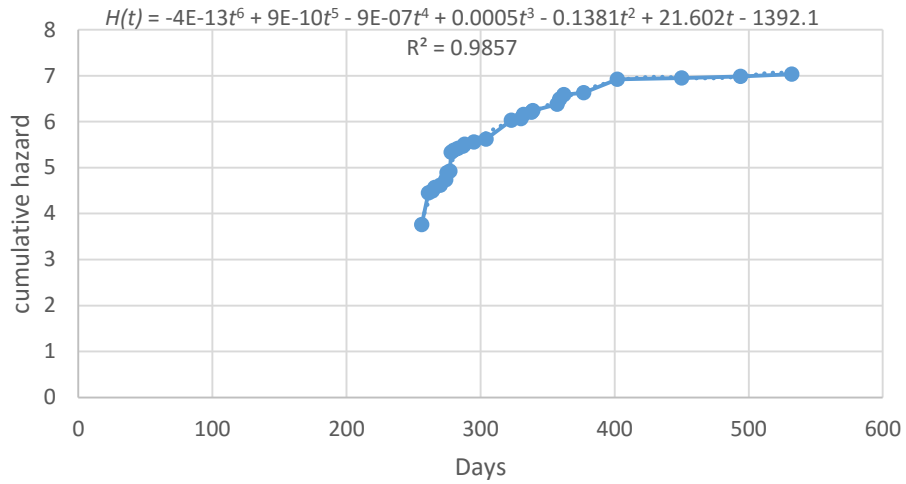

Figure 4.6: Cumulative hazard graph for specific days

The last interval for the cumulative hazard function is shown the figure above. The model fitted to the curve is a polynomial of order 6 which is also very good fit considering the *r-squared* value of 0.9857

Figs 4.3, 4.4, 4.5, and 4.6 give the hazard rates for 1-4 days, 8-80 days, 80-256 days and 256-532 days respectively. From the diagrams above, the hazard function $[H(t)]$ for time until claims are reported is given as

$$H(t) = \begin{cases} 0.0097t^2 - 0.0193t + 0.0609 & 1 \le t < 4 \\ 4E-07t^4 - 7E-05t^3 + 0.004t^2 - 0.0421t + 0.2293, & 4 \le t < 80 \\ -2E-12t^6 + 2E-09t^5 - 7E-07t^4 + 0.0001t^3 - 0.0154t^2 + 0.9139t - 20.129, & 80 \le t < 256 \\ -4E-13t^6 + 9E-10t^5 - 9E-07t^4 + 0.0005t^3 - 0.1381t^2 + 21.602t - 1392.1, & 256 \le t < 532 \end{cases}$$

where $(E - t = 10^{-t})$

4.2.2 Confidence intervals for time until reporting claim

To determine the reliability and region of the Kaplan-Meier estimates for time until claims were reported, a 95% confidence interval was calculated using the linear and log-transformed methods. Both methods were constructed in Appendix III to determine which one was better in constructing the confidence interval. Since the Kaplan-Meier estimate ranges between zero and one and its confidence interval should also be bounded by zero and one, the log-transformed confidence was adjudged the better of the two methods due to the fact that its confidence intervals were within the range of zero and one whereas the linear confidence had intervals which exceeded one. This confirms the findings of Borgan and Liestøl (1990)

Table 4.4: Percentiles and their corresponding confidence interval for time until claims are reported and settled

Time period	Percentile	95% Linear C.I	95%Log-transformed C.I
DUC	50^{th} percentile	27-45 days	27-45 days
DUP	50^{th} percentile	10-18 days	7-18days

*DUC=days until claims were reported and DUP=days until claims were settled /paid

The Kaplan-Meier estimate of the median, which is relatively unbiased as compared to the Kaplan-Meier estimate of the mean (Zhong and Hess, 2009) was employed in the study. Table 4.4 presents the confidence intervals for the median survival time which were calculated for policies underwritten from January 2010 to December 2012. It was shown from the table that the linear and log-transformed median confidence intervals for the time it took for policyholders to report a claim was between 27-45 days. The median day for time until payments will be settled by the insurers is 36 days. (See Appendix V for details)

4.2.3 Estimating probabilities and hazard rates for time until claims were reported

In Appendix II, the probability of reporting a claim by four weeks (approximately one month) after the policy takes effect given by $S(28)$ is 0.3648. The rate at which policies had been in effect for four weeks (approximately one month) and had claims reported by then is 1.008.

The probability that claims will be reported at least two months is 0.1818 and the rate at which at which policies in effect for two months were paid on the last day of the second month (i.e. $S(60)$)is 1.7049. The probability that claims will be reported at least one year was 0.001319. This indicates that, it was very unlikely for claims to be reported a year or more after policies took effect. A general finding from the table was that, the longer it took for claims to be reported the lower its probability of survival.

4.3 Analysis for time until payments were made to policyholders

All analysis conducted on time until payments were settled to policyholders is being addressed in this section

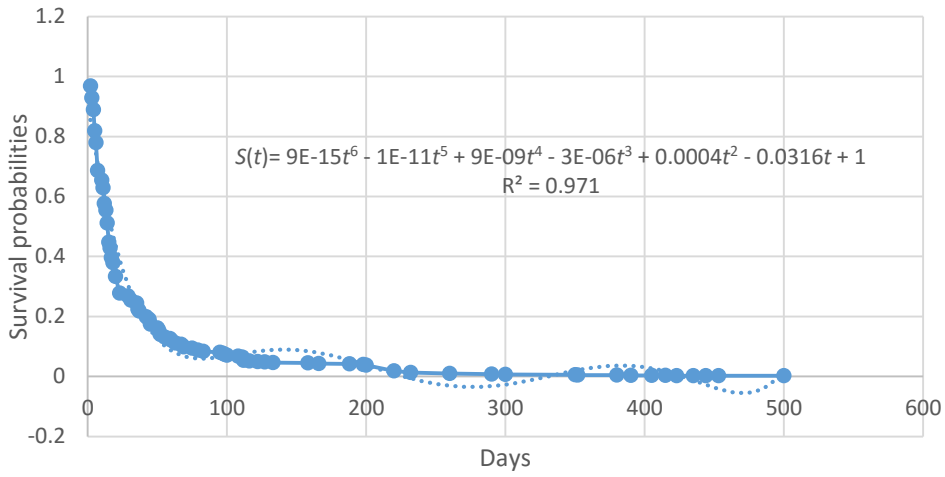

Figure 4.7: Kaplan-Meier curve for time until payment are made

Figure 4.7 above gives the survival curve for the length of time it took for policyholders to get paid. The survival function for the time it took for policyholders to get paid is given as

$S(t)$= 9.8E-15t^6 - 1E-11t^5 + 9E-09t^4 – 3E-06t^3 + 0.0004t^2 - 0.0345t + 1

This function is in conformity wih the properties $S(0)=1$ and $S(\infty)=0$. The r-squared value of 0.971 shows how well the model fts the curve. From Appendix I, the Kaplan-Meier estimate for the length it takes for payment to be made to policyholder was 0.0021.

4.3.1 Estimating the cumulative hazard function for time until payment

Due to the nature of the cumulative hazard function for time until payments were made to policyholders, the cumulative hazard functions for the various time periods are constructed in order to determine the general cumulative hazard function for the length of time it takes for policyholders to be settled.

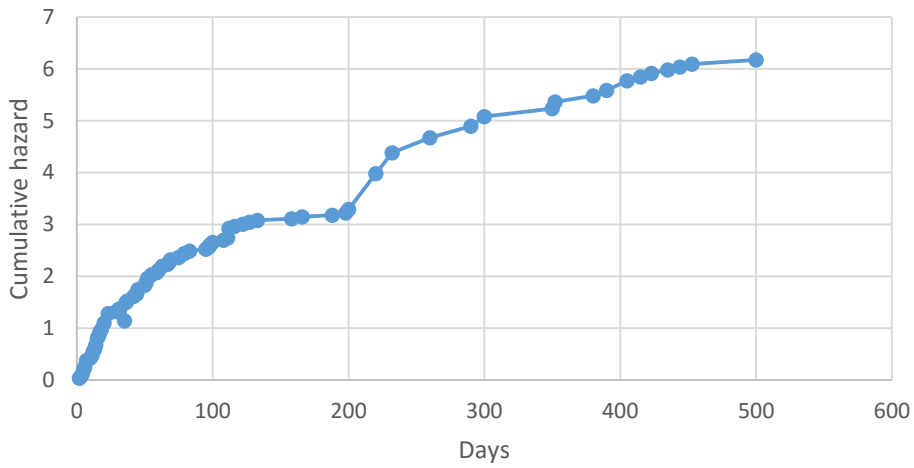

Figure 4.8: Cumulative hazard graph for time until payment was made to policyholders

Figure 4.8 above gives the hazard graph and function for the length of time it takes for policyholders to get settled.

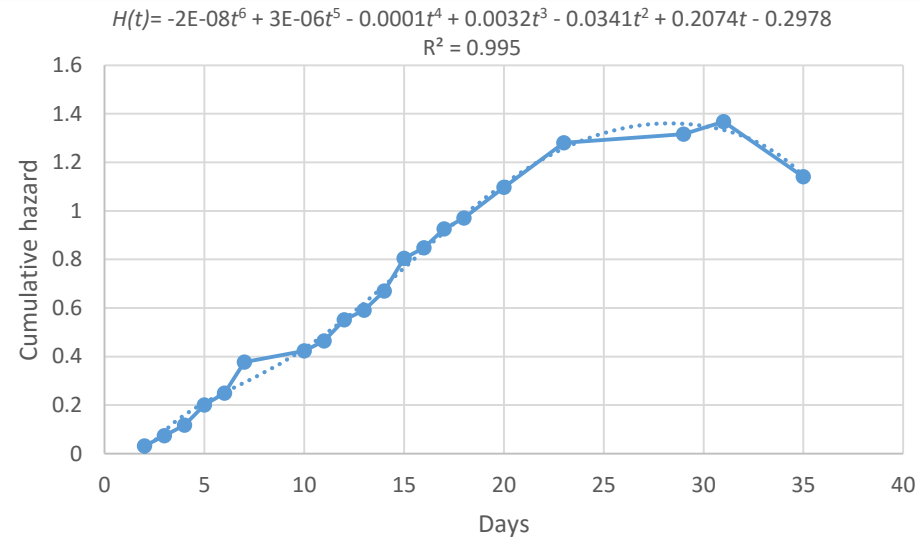

Figure 4.9: Cumulative graph for a specific period

Figure 4.9 gives the cumulative hazard function for the first interval. The first interval was modelled using a polynomial of 6. Its *r-squared* value was 0.995, which is very high indicating how well the model suits the curve.

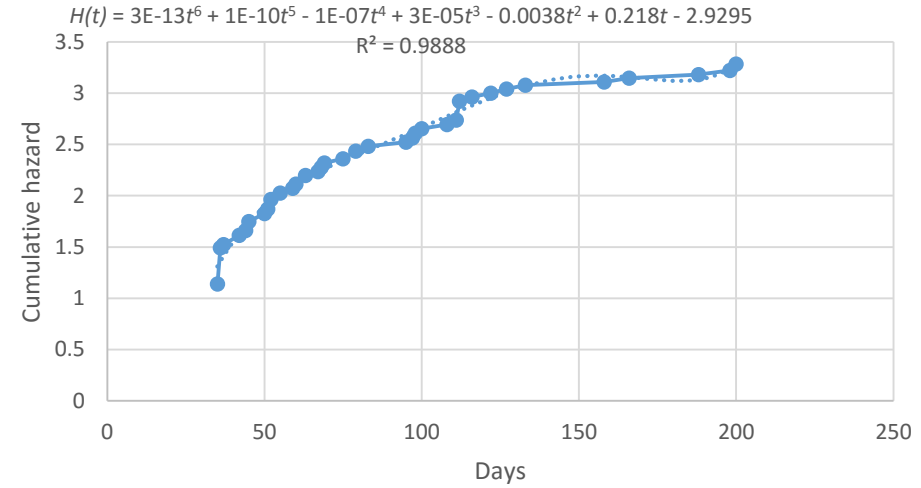

Figure 4.10: Cumulative graph for specific time period

Figure 4.10 represents the second interval for the cumulative hazard function. This interval is from 35-200 days. It was observed that the polynomial model of order 6 fitted the curve had a high *r-squared* value of 0.9888 indicating a good fit.

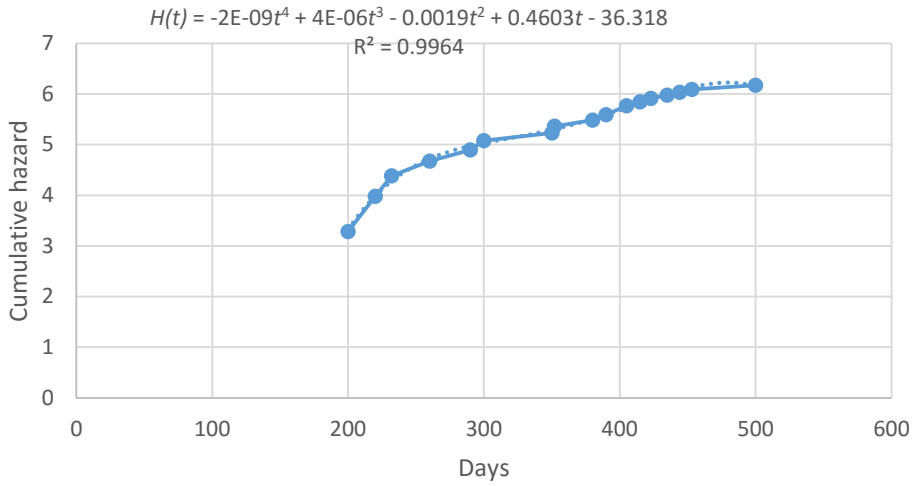

Figure 4.11: Cumulative hazard for specific days

The third region ranging from 200-500 days is drawn in the graph above. A polynomial model of order 4 was fitted to the curve. The *r-squared* value of the model is 0.9964 indicating a good fit

Figures 4.9, 4.10 and 4.11 give the cumulative hazard functions rates for 2-35 days, 35-200 days and 200-500 days respectively. The cumulative hazard function for time until payments are settled to policyholders is given by

$$H(t) = \begin{cases} -2E-08t^6 + 3E-06t^5 - 0.0001t^4 + 0.0032t^3 - 0.034t^2 + 0.2074t - 0.2978, & 2 \le t < 35 \\ 3E-13t^6 + 1E-10t^5 - 1E-07t^4 + 3E-05t^3 - 0.0038t^2 + 0.218t + 2.9295, & 35 \le t < 200 \\ -2E-09t^4 + 4E-06t^3 - 0.0019t^2 + 0.4603t - 36.318, & 200 \le t < 500 \end{cases}$$

where $(E - t = 10^{-t})$

satisfying the properties if a cumulative hazard function

4.3.2 Confidence intervals for time until payment

Similar to the observation found in determining the confidence interval for time until claims were reported by policyholders, the log-transformed confidence interval was found to be the better method for estimating the confidence interval of the Kaplan-Meier estimate for time until payments were made (Appendix IV)

The linear and log-transformed median confidence interval for time it took for payments to be made to policyholders was 10-18 days and 7-18 days respectively. (See Appendix VI for details).

4.3.3 Estimating probabilities and hazard rates for time until payment

In Appendix I, the probability of a claim being settled by seven days weeks denoted by $S(7)$ is 0.6860. The rate at which policies had been in effect for seven days (one week) and paid by the end of the week is 0.3769. Also, the probability that payments will be settled in 31 days (approximately) a month is 0.2547. It was found from the table that, weekly settlements of claims had higher probabilities than monthly settlements.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.0 Introduction

This study was carried out to investigate the application of survival analysis in auto- insurance contracts in Ghana. It covered a dataset consisting of effective date of issued polices, time until claims were reported by policyholders and time until payments were made to policyholders ranging from January 2010 to December 2012. The main objective was to model the length of time it takes for policyholder to make claims and also be settled. This chapter presents a summary of the findings and conclusions drawn from those findings. Some recommendations are made including directions for future research in this field.

5.1 Summary of findings

The Kaplan-Meier estimates for time until claims were reported and time until claims were paid were 0.0009 and 0.0021. The survival estimates for both time until payments as well as time until claims were calculated. However, the intercepts of the survival functions slightly exceeded one violating the properties of survival functions. The survival functions were therefore standardized in order for them to satisfy the properties of survival functions.

Also, the log-transformed confidence interval was found to be better in providing the range for the Kaplan-Meier estimate compared to the linear confidence interval. This was due to the fact that the log- transformed confidence intervals range from zero to one. The study also showed that the probability that policyholders will report claims within a week was higher than in a month. The survival estimates from the survival function estimated is very close to the Kaplan-Meier estimate computed. This indicated that, the models computed were efficient for predicting future survival estimates. It was found from the study that, shorter duration of claim settlements had higher probabilities than monthly settlements

5.2 Conclusions

In order to model time it takes for claims to be reported and payments to be settled, the nonparametric Kaplan-Meier estimate was used to obtain estimators of the hazard functions associated with the length of time it takes for claims to be reported and paid.

The log-transformed interval was found to be better than the linear confidence interval. It can therefore be concluded from the study that survival analysis is an appropriate tool for studying the insurance industry

5.3 Recommendations

It has been established in the study that the survival functions give the probability that at least a policyholder will make a claim or get paid at least at a certain period of time. It is recommended that insurance firms use this in calculating the probabilities that claims will fall within a certain range so that they can make the necessary preparations to meet the demands of their customers.

5.4 Areas for future research

This study has demonstrated the use of nonparametric methods of survival analysis applied to the motor insurance industry of Ghana. However other areas which may need further research include:

1. Parametric modeling for the length of time it takes for policyholders to make claims and also get settled taking into consideration covariates such as age of vehicle, type of vehicle, type of insurance policy among others
2. Using survival analysis to estimate the lifespan of auto insurance contracts in Ghana

REFERENCES

Aalen, O. O. (1978). Nonparametric inference for a family of counting processes. *Annals of Statistics, 6,* 701-726.

Abada, T. S., Trovato, F., & Lalu, N. (2001). Determinants of breastfeeding in the Philippines: a survival analysis. *Social Science and Medicine 52,* 71-81.

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control 19/6,* 716-723.

Allison, P. D. (1995). *Survival Analysis Using SAS: A Practical Guide.* North Carolina: SAS Institute Inc.

Amoo, G. K. (2002). *Going the extra mile: The challenge of providing insurance cover for loss of use of motor vehicle in a developing economy.* A dissertation summated to Chartered Insurance Institute.

Atsmegiorgis, C. (2014). Survival Analysis of bank loan repayment rate of customers of Hawassa commercial bank of Ethiopia. *Journal of the Korean data and information science society,* 1591-1598.

Awunyo-Vitor, D. (2012). Comprehensive Motor Insurance Demand In Ghana: Evidence from Kumasi Metropolis. *Management 2(4),* 80-86.

Beirlant, J., Derveaux, V., De Meyer, A. M., Goovaerts, M. J., Labie, E., & Maenhoudt, B. (1991). Statistical risk evaluation applied to (Belgian) car insurance. *Insurance: Mathematics and Economics 10,* 289-302.

Belloti , T., & Crook, J. (2009). Credit Scoring With Macroeconomic Variables Using Survival Analysis. *Journal of the Operational Research Society, 60,* 1699-1707.

Besedes, T., & Blyde, J. (2010). What Drives Export Survival? An Analysis of Export in Latin America. *Journal of Development Economics, Vol.75*, 417-450.

Boland, P. J. (2006). Statistical methods in general insurance. *International Conference on Teaching Statistics, ICOTS-7* (pp. 1-6). Salvador: IASE.

Borgan , O., & Liestol, K. (1990). A note on confidence intervals and bands for the survival curve based on transformations. *Scandinavian Journal of Statistics, 17*, 35-41.

Brockett, P. L., Golden, L., Guillen, M., Nielson, J. P., Parner, J., & Perez-Marin, A. M. (2008). Survival Analysis of household insurance policies: How Much Time Do You Have to Stop Total Customer Defection? *Journal of Risk and Insurance,75 (3)*, pp.550-563.

Brown, G. W., & Flood , M. M. (1947). Tumbler Mortality. *Journal of the American Statistical Association, 42*, 562-574.

Byers, R. H., Morgan, W. M., Darrow, W. W., Doll, L., Jaffe, H. W., Rutherford, G., . . . O'Malley, P. M. (1988). Estimating AIDS Infection Rates in the San Fransico Cohort . *AIDS 2(3)*, 207-210.

Cabo, P., & Rebelo, J. (2010). "Co-operatives contributions to a plural economy" The survival of Portuguese Credit Co-operatives: An Econometric Approach. *ICA European Research Conference.* Lyon. 2-4 September.

Cameron, A. C., & Hall, A. D. (2003). A Survival Analysis of Australian Equity Mutual Fund. *Australian Journal of Management, Vol.28,No.2*, 209-226.

Cao, R., Vilar, J. M., & Devia, A. (2009). Modelling consumer credit risk via survival analysis. *SORT 33(1)*, 3-30.

Caree, M. A. (2003). A hazard rate analysis of Russian commercial banks in the period 1994-1997. *Economic Systems 27*, 255-269.

Carlen, E., Schneider, M. d., & Strandberg, E. (2004). Genetic Evaluation for Mastitis using Survival Analysis. *55th Annual Meeting of the European Association for Animal Producton,September 5th-9th,Commission on Animal Genetics,Session G2.8*, (pp. 1-3). Bled, Slovenia.

Chow, M. H., Szymanoski, E. J., & DiVenti, T. R. (2003). Applying Survival Analysis Techniques to Loan Terminations for HUD's Reverse Mortgage Insurance Program-HECM. *Manuscript*, 8p.

Conley, Q. D. (2013). Simulating abondonment using kaplan-meier survival analysis in a shared billing and claims call center. *Winter Simulation Conference.* Westfield.

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of Royal Statistical Society. Series B (Methodological), Vol.34, No.2*, 187-220.

Crapp, H. R., & Stevenson, M. (1987). Development of a method to assess the relevant variables and the probability of financial distress. *Australia Journal of Management, 12 (2)*, 221-236.

Czado , C., & Rudolph, F. (2002). Application of survival analysis methods to long-term care insurance. *Insurance: Mathematics and Economics 31*, 396-413.

Dabos, M., & Escudero, S. W. (2004). Explaining and predicting bank failure using duration models: The case of Argentina after the Mexican crisis. *Revista de analisis economico. Vol.19, No.1*, 31-49.

Doghonadze, N. (2012). *Survival Analysis of Export Spells-Empirical Evidence from Georgia. Master's Thesis.* International School of Economics at Tbilisi State University (ISET) 33p.

Etzioni, R. D., Feuer, E. J., Sullivan, S. D., Lin, D., Hu, C., & Ramsey, S. D. (1999). On the use of survival analysis techniques to estimate medical care costs. *Journal of Health Economics, Vol.18,* 365-380.

Falk, M. (2011). A Survival Analysis of Ski Lift Companies. *Tourism Management. Vol. 36,* 377-390.

Feinleib, M., & MacMahon, B. (1960). Variation in the Duration of Survival of patients with chronic Leukemia. *Blood,17,* 322-349.

Finan, M. B. (2007). *A probability course for the Actuaries. A preparation for Exam P/1. Arkansas Tech University.*

Franco, L., Jerez, J. M., & Alba, E. (2005). Artificial neural networks and prognosis in medicine. Survival analysis in breast cancer patients. *ESANN2005 proceedings-European Symposium on Artificial Neural Networks Bruges (Belgium), 27-29 April* (pp. 91-102). Bruges: D-side.

Fu, D., & Wu, Y. (2013). *Export survival pattern and determinants of Chinese manufacturing firms. Discussion paper 13.18.* University of Western Australia,37p.

Gamerman, D., & West, M. (1987). An application of dynamic survival models in unemployment studies. *The Statistician, Vol.36,* 269-274.

Glennon, D. C., & Nigro, P. (2005). Measuring the Default Risk of Small Business Loans: A Survival Analysis Approach. *Journal of Money,Credit, and Banking, Volume 37, Number 5, October,* 923-947.

Goswami, P. (2007). Customer Satisfaction with Quality in the life Insurance Industry in India. *The Icfai Journal of Services Marketing, Vol.V, No. 1*, 2.

Greenberg, J. A., Kefauver, S. C., Stimson, H. C., Yeaton, C. J., & Ustin, S. L. (2005). Survival Analysis of a neotropical rainforest using multitemporal satellite imagery. *Remote Sensing of Environment 96*, 202-211.

Halling, M., & Hayden, E. (2006). Bank failure prediction:a two-step survival time approach. *IFC Bulletin No.28*, 48-73.

Henebry, K. L. (1997). A test of the temporal stability of proportional hazards models for predicting bank failure. *Journal of Financial And Strategic Decisions Vol.10(3)*, 1-11.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data.* Springer-Verlag New York.

Janot, M. (2001). Modelos de previsao de insolvencia bancaria no Brasil. *Trabalhos para Discussao n.13 Brasilia: Banco Central do Brasil*, 1-41.

Johannsen, E. K. (2013). *Survival Analysis of foreign R&D units in Swedish Multinational Enterprises.* Master's Thesis. Copenhagen Business School.

Kalbfleisch, J. D., & Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* New York: John Wiley & Sons Inc.

Kamleh, R., Toufeili, I., Ajib, R., Kanso, B., & Haddad, J. (2012). Estimation of the Shelf-Life of Halloumi Cheese Using Survival Analysis. *Czech Journal of Food Science. Vol.30, No.6*, 512-519.

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

Kelly, R., Brien, E. O., & Stuart, R. (2014). *A long-run analysis of corporate liquidations in Ireland.* Research Technical Paper. Central Bank of Ireland. 23p.

Klein, J. P., & Moeschberger, M. L. (2003). *Survival Analysis Techniques for Censored and Truncated Data, Second Edition.* New York: Springer-Verlag .

Kleinbaum, D. G. (1996). *Survival analysis: a self-learning text. First Edition.* New York: Springer-Verlag.

Klos, V. (2008). *Firm's Performance Analysis Using Survival Methods.* Master's Thesis. National University "Kyiv-Mohyla Academy", Kyiv. 41p.

Klugman , S. A., Panjer, H. H., & Willmot, G. E. (2004). *Loss Models: From Data to Decisions.Second Edition.* New Jersey: John Wiley & Sons Inc.

Lane, W. R., Looney, S. W., & Wansley, J. W. (1986). An Application of the Cox Proportional Hazards Model to Bank Failure. *Journal of Banking and Finance 10*, 511-531.

Langova, K. (2008). Survival Analysis for clinical studies. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub.152(2)*, 303-307.

Lawless, J. (2003). *Statistical models and methods for lifetime data.* New York: Wiley.

Lee, E. T., & Wang, J. W. (2003). *Statistical Methods for Survival Data Analysis.* New Jersey: John Wiley & Sons.

Lee, M.-C. (2014). Business Bankruptcy Prediction Based on Survival Analysis Approach. *International Journal of Computer Science & Information Technology(IJCSIT) Vol.6, No2*, 103-119.

Lobos, K., & Szewczyk, M. (2012). Survival Analysis: A case study of micro and small enterprises in Dolnoslaskie and Opolskie Voivodship(Poland). *Ekonomicka Revue-Central European Review of Economic Issues, Vol.15*, 207-216.

Luoma, M., & Laitinen, E. (1991). Survival Analysis as a Tool for Company Failure Prediction. *Omega 19(6)*, 673-678.

Maddala, G. S. (2005). *"Introduction to Economics" 3rd Edition.* John Wiley &Sons Ltd. The Atrium, Southern Gat, England: 318-323.

Mannasoo, K., & Mayes, D. G. (2009). Explaining bank distress in Eastern European transition. *Journal of banking and finance Vol.33, No.2*, 244-254.

Musakwa, F. T. (2013). Measuring Bank Funding Liquidity Risk. *IAA COLLOQUIUM & SUMMER SCHOOL*, (pp. 1-26). Lyon.

Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics, 14*, 945-965.

Newby, M. (1988). Accelerated failure time models for reliability analysis. *Reliability Engineering & System Safety, 20 (3)*, 187-197.

NIC. (2011). *Annual Report of National Insurance Commission, Ghana.*

Nunes, A., & Sarmento, E. (2010). Business Demography Dynamics in Portugal: A Non-parametric Survival Analysis. *GEE papers*, 24p.

Onafalujo, A. K., Abass, O. A., & Dansu, S. F. (2011). "Effect of Risk Premium Perception on the Demand for Insurance: Implication on Nigerian Road Users". *Journal of Emerging Trends in Economics and Management Sciences,2(4)*, 285-290.

Pappas, V. (2010). *Determinants of banking fragility: Comparing Islamic and conventional banks. An application of survival analysis.* A report of Gulf one Lancaster Centre for Economic Research. Lancaster University Management School.8p.

Pereira, J. (2014). Survival Analysis Employed in Predicting Corporate Failure: A Forecasting Model Proposal. *International Business Research; Vol.7, No.5*, 9-20.

Pike, M. C. (1966). A Method of Analysis of a Certain Class of Expriments in Carcinogenesis. *Biometrics,22*, 142-161.

Renshaw, A., & Haberman, S. (2005). Mortality reduction factors incorporating cohort effects. *Actuarial Research Paper No.160. Cass Business School*, 33p.

Rocha, F. (1999). Previsao de Falencia Bancaria: un modelo de risco proporcional. *Pesquisa e Planejamento Economico, 29(1)*, 137-152.

Sadler, A., & Lang, L. (2006). Using Survival Analysis to Predict Sample Retention Rates. *Proceedings-American Statistical Association-CD-ROM Edition-:1457-1464.*

Schneider, M. P. (2006). *Mastitis and Longetivity in Dairy Cattle Analyzed using Survival Models.Doctoral Dissertation.* Swedish University of Agricultural Science, Uppsala. 33p.

Schunk, D. (2003). The Pennsylvania Reemployment Bonus Experiments: How a survival model helps in the analysis of the data. *SONDERFORSHUNGSBEREICH504. NO.03-35*.

Setzer, R. (2004). The Political Economy of Exchange Rate Regime Duration: A Survival Analysis. Discussion Paper, University of Hohenheim, Stuttgart. 36p.

Siegel, S. (1957). Nonparametric Statistics. *The American Statistician Vol.11, No.3*, 13-19.

Stefancic, M. (2014). Investigating Management Turnover in Italian Cooperative Banks. *Journal of Entrepreneurial and Organizational Diversity, Vol.3, Issue 1*, 131-163.

Stepanova, M., & Thomas , L. (2002). Survival Analysis Methods For Personal Loan Data. *Operations Research, Vol.50, No.2*, 277-289.

Vance, C., & Geoghegan, J. (2002). Temporal and Spatial modelling of tropical deforestation : A survival analysis linking satellite and household survey data. *Agricultural economics*, 317-322.

Wang, E., Yu, Y., Little, B. B., & Li, Z. (2010). Crop Insurance Premium Design Based on Survival Analysis Model. *Agriculture and Agricultural Science Procedia 1*, 67-75.

Weibull, W. (1939). A Statistical Theory of the Strength of Materials. *Ingenioers vetenskaps akakemien Handlingar,151*, 293-297.

Whalen, G. (1991). A Proportional Hazard Model of Bank Failure: An examination of its usefulness as an early tool. *Federal Reserve Bank of Cleveland, Economic Review*, First Quarter: 21-31.

Wheelock, D. C., & Wilson, P. W. (2000). Why do banks disappear? The determinants of US bank failures and acquisitions. *Rev.Econ.Stat. 82*, 127-138.

Witzany, J., Rychnovsky, M., & Charamza, P. (2012). Survival Analysis in LGD Modelling. *European Financial and Accounting Journal Vol.7, No.1*, 6-27.

Woodall, C. W., Grambsch, P. L., & Thomas, W. (2005). Applying survival analysis to a large-scale forest inventory for assessment of tree mortality in Minnesota. *Ecological Modelling 189*, 199-208.

Xie, Y., & Giles, D. E. (2007). *A Survival Analysis of the Approval of U.S Patent Applications. Econometrics Working Paper EWP0707.* Department of Economics, Univeristy of Victoria.23p.

Zaman, Q., & Pfeiffer, K. P. (2012). Does log-rank test give satisfactory results? *Journal of applied and quantitative methods. Vol.7, No 1*, 1-8.

Zavadilova, L., Stipkova, M., Nemcova, E., Bouska, J., & Matejickova, J. (2009). Analysis of the phenotype relationships betwen type traits and functional survival in Czech Fleckvieh cows. *Czech Journal of Animal Science, Vol.54,No.12*, 521-531.

Zelen, M. (1966). Applications of Exponential Models to Problems in Cancer Research. *Journal of the Royal Statistical Society,Series A,129*, 368-398.

Zhang, J., & Thomas, L. C. (2012). Comparisons of linear regression and survival analysis using single and mixture distributions approaches in modelling LGD. *International Journal of Forecasting 28*, 204-215.

Others (Laws)

Ghana Insurance Law P.N.D.C Law 227, 1989

National Insurance Law, Act 724, 2006

APPENDICES

Appendix I: Kaplan-Meier estimate and Cumulative hazard estimates for time until payments were made to policyholders

t	s_i	r_i	$\frac{s_i}{r_i(r_i - s_i)}$	$\frac{r_i - s_i}{r_i}$	$S(t) = \prod_{i=1}^{k}\left(1 - \frac{s_i}{r_i}\right)$	$Var[\hat{S}(t)] = \hat{S}(t)^2 \sum_{i=1}^{j} \frac{s_i}{r_i(r_i - s_i)}$	$H(t) = -In[S(t)]$
2	1	32	0.0010	0.9688	0.96885	0.0009	0.0317
3	1	24	0.0018	0.9583	0.9284	0.0024	0.0743
4	1	24	0.0018	0.9583	0.8897	0.0037	0.1169
5	2	25	0.0035	0.9184	0.8185	0.0054	0.2003
6	1	21	0.0024	0.9524	0.7795	0.0064	0.2490
7	3	25	0.0055	0.8784	0.6860	0.0075	0.3769
10	1	22	0.0022	0.9545	0.6548	0.0078	0.4234
11	1	25	0.0017	0.9600	0.6286	0.0078	0.4642
12	2	24	0.0038	0.9167	0.5762	0.0078	0.5512
13	1	26	0.0015	0.9615	0.5541	0.00771	0.5904
14	2	26	0.0032	0.9216	0.5115	0.0074	0.6705
15	3	24	0.006	0.8732	0.4475	0.0069	0.8040
16	1	23	0.002	0.9565	0.4281	0.0066	0.8485
17	2	27	0.003	0.9245	0.3964	0.0062	0.9254
18	1	23	0.002	0.9565	0.3791	0.0059	0.9699
20	3	25	0.0055	0.8800	0.3336	0.0052	1.0977
23	4	24	0.0083	0.8351	0.2780	0.0042	1.2800
29	1	28	0.0013	0.9643	0.2681	0.0038	1.3164
31	1	20	0.0026	0.9500	0.2547	0.0038	1.3677
35	1	24	0.0018	0.9583	0.2441	0.0036	1.1403
36	2	25	0.0036	0.9200	0.2246	0.0032	1.4936
37	1	33	0.0009	0.9697	0.2177	0.0031	1.5244
42	2	23	0.0041	0.9130	0.1988	0.0027	1.6154
44	1	23	0.002	0.9565	0.1902	0.0026	1.6598

Appendix I: Continued

t	s_i	r_i	$\frac{s_i}{r_i(r_i-s_i)}$	$\frac{r_i-s_i}{r_i}$	$S(t)=\prod_{i=1}^{k}\left(1-\frac{s_i}{r_i}\right)$	$Var[\hat{S}(t)]=\hat{S}(t)^2\sum_{i=1}^{j}\frac{s_i}{r_i(r_i-s_i)}$	$H(t)=-In[S(t)]$
45	2	24	0.0038	0.9167	0.1743	0.0023	1.7469
50	2	26	0.0032	0.9231	0.1609	0.0020	1.8269
51	1	23	0.002	0.9565	0.1539	0.0019	1.8713
52	2	23	0.0041	0.9130	0.1405	0.0017	1.9623
55	2	32	0.0021	0.9375	0.1317	0.0015	2.0269
59	1	22	0.0022	0.9545	0.1258	0.0014	2.0734
60	1	25	0.0017	0.9600	0.1207	0.0013	2.1142
63	2	25	0.0035	0.9200	0.1111	0.0012	2.1976
67	1	28	0.0013	0.9643	0.1071	0.0011	2.2339
68	1	24	0.0018	0.9583	0.1026	0.0010	2.2765
69	1	24	0.0018	0.9583	0.0984	0.001	2.3191
75	1	25	0.0017	0.9600	0.0944	0.0009	2.3599
79	2	28	0.0027	0.9286	0.0877	0.0008	2.4340
83	1	21	0.0024	0.9524	0.0835	0.0007	2.4828
95	1	25	0.0017	0.9600	0.0802	0.0007	2.5236
97	1	26	0.0015	0.9615	0.0771	0.0006	2.5628
98	1	22	0.0022	0.9545	0.0736	0.0006	2.6093
100	1	23	0.002	0.9565	0.0704	0.0006	2.6538
108	1	26	0.0015	0.9615	0.0677	0.0005	2.6930
111	1	22	0.0022	0.9545	0.0646	0.0005	2.7395
112	1	6	0.0333	0.8333	0.0538	0.0004	2.9219
116	1	25	0.0017	0.9600	0.0517	0.0004	2.9627
122	1	26	0.0015	0.9615	0.0497	0.0004	3.0019

Appendix I: Continued

t	s_i	r_i	$\frac{s_i}{r_i(r_i-s_i)}$	$\frac{r_i-s_i}{r_i}$	$S(t)=\prod_{i=1}^{k}\left(1-\frac{s_i}{r_i}\right)$	$Var[\hat{S}(t)]=\hat{S}(t)^2\sum_{i=1}^{j}\frac{s_i}{r_i(r_i-s_i)}$	$H(t)=-In[S(t)]$
127	1	26	0.0015	0.9615	0.0478	0.0004	3.0411
133	1	27	0.0014	0.9630	0.0460	0.0003	3.0789
158	1	33	0.0009	0.9697	0.0446	0.0003	3.1096
166	1	27	0.0014	0.9630	0.0430	0.0003	3.1474
188	1	31	0.0011	0.9677	0.0416	0.0003	3.1802
198	1	23	0.002	0.9565	0.0398	0.0003	3.2225
200	1	17	0.0037	0.9412	0.0374	0.0002	3.2853
220	1	2	0.5	0.500	0.0187	0.0002	3.9784
232	1	3	0.1667	0.6667	0.0128	0.0001	4.3839
260	1	4	0.0833	0.7500	0.0094	8.01407E-05	4.6715
290	1	5	0.05	0.800	0.0075	5.40923E-05	4.8947
300	1	6	0.0333	0.8333	0.0062	3.88614E-05	5.0770
350	1	7	0.0238	0.8571	0.0053	2.92321E-05	5.2312
352	1	8	0.0179	0.8750	0.0047	2.27717E-05	5.3647
380	1	9	0.0139	0.8889	0.0042	1.82327E-05	5.4825
390	1	10	0.0111	0.900	0.0037	1.49242E-05	5.5879
405	2	12	0.0167	0.8333	0.0031	1.05262E-05	5.7702
415	1	14	0.0055	0.9286	0.0029	9.12224E-06	5.8444
423	1	15	0.0048	0.9333	0.0027	7.98129E-06	5.9134
435	1	16	0.0042	0.9375	0.0025	7.04157E-06	5.9780
444	1	18	0.0033	0.9444	0.0024	6.29963E-06	6.0348
453	1	19	0.0029	0.9474	0.0023	5.669E-06	6.0889
500	1	13	0.0064	0.9231	0.0021	4.85847E-06	6.1692

Appendix II: Kaplan-Meier estimate and Cumulative hazard estimates for time until claims were reported by policyholders

t	s_i	r_i	$\frac{s_i}{r_i(r_i - s_i)}$	$\frac{r_i - s_i}{r_i}$	$S(t) = \prod_{i=1}^{k}\left(1 - \frac{s_i}{r_i}\right)$	$Var[\hat{S}(t)] = \hat{S}(t)^2 \sum_{i=1}^{j} \frac{s_i}{r_i(r_i - s_i)}$	$H(t) = -In[S(t)]$
1	1	20	0.0026	0.95	0.95	0.0024	0.0513
3	1	26	0.0015	0.9615	0.9135	0.0035	0.0905
4	1	21	0.0024	0.9524	0.8700	0.0050	0.1393
11	1	27	0.0014	0.9630	0.8377	0.0056	0.1770
20	1	20	0.0026	0.95	0.7959	0.0067	0.2283
27	2	4	0.25	0.5	0.3979	0.0413	0.9215
28	2	24	0.0038	0.9167	0.3648	0.0352	1.0085
36	1	8	0.0179	0.875	0.3192	0.0288	1.1420
37	2	14	0.0119	0.8571	0.2736	0.0220	1.2962
41	1	28	0.0013	0.9643	0.2638	0.0206	1.3325
43	1	28	0.0013	0.9643	0.2544	0.0192	1.3689
45	1	20	0.0026	0.95	0.2417	0.0175	1.4202
46	1	7	0.0238	0.8571	0.2071	0.0139	1.5744
50	1	19	0.0029	0.9474	0.1962	0.0126	1.6284
55	1	30	0.0011	0.9667	0.1897	0.0118	1.6623
60	1	24	0.0018	0.9167	0.1818	0.0109	1.7049
62	1	28	0.0013	0.9643	0.1753	0.0102	1.7413
70	1	10	0.0111	0.9	0.1578	0.0085	1.8466
72	1	27	0.0014	0.9630	0.1519	0.0079	1.8844
75	1	19	0.0029	0.9474	0.1439	0.0072	1.9384
77	1	21	0.0024	0.9524	0.1371	0.0065	1.9872
80	1	20	0.0026	0.95	0.13023	0.0060	2.0385
85	2	26	0.0032	0.9231	0.1202	0.0051	2.1185
96	2	21	0.0050	0.9048	0.1088	0.0042	2.2186
100	1	26	0.0015	0.9615	0.1046	0.0039	2.2579
106	2	27	0.0029	0.9259	0.0968	0.0034	2.3348

Appendix II: Continued

t	s_i	r_i	$\frac{s_i}{r_i(r_i - s_i)}$	$\frac{r_i - s_i}{r_i}$	$S(t) = \prod_{i=1}^{k}\left(1 - \frac{s_i}{r_i}\right)$	$Var[\hat{S}(t)] = \hat{S}(t)^2 \sum_{i=1}^{j} \frac{s_i}{r_i(r_i - s_i)}$	$H(t) = -In[S(t)]$
116	1	21	0.0024	0.9524	0.0922	0.0031	2.3836
121	1	19	0.0029	0.9474	0.0874	0.0028	2.4377
129	1	20	0.0026	0.95	0.0830	0.0026	2.4890
136	1	20	0.0026	0.95	0.0788	0.0023	2.5403
137	1	20	0.0026	0.95	0.0749	0.0021	2.5915
142	1	28	0.0013	0.9643	0.0722	0.0020	2.6279
145	1	26	0.0015	0.9615	0.0695	0.0018	2.6671
151	1	25	0.0017	0.96	0.0667	0.0017	2.7080
168	1	22	0.0022	0.9545	0.0636	0.0016	2.7545
170	1	17	0.0037	0.9412	0.0599	0.0014	2.8151
171	1	9	0.0139	0.8889	0.0532	0.0011	2.9329
189	1	18	0.0033	0.9444	0.0503	0.0010	2.9900
198	1	29	0.0012	0.9655	0.0486	0.0010	3.0251
203	1	25	0.0017	0.96	0.0466	0.0009	3.0660
205	1	25	0.0017	0.96	0.0447	0.0008	3.1068
210	1	28	0.0013	0.9643	0.0431	0.0008	3.1431
213	1	28	0.0013	0.9643	0.0416	0.0007	3.1795
222	1	20	0.0026	0.95	0.0395	0.0006	3.2308
225	1	29	0.0012	0.9655	0.0382	0.0006	3.2659
226	1	20	0.0026	0.95	0.0363	0.0005	3.3172
228	1	16	0.0042	0.9375	0.0340	0.0005	3.3817
232	1	15	0.0048	0.9333	0.0317	0.0004	3.4507
236	1	19	0.0029	0.9474	0.0301	0.0004	3.5048
242	1	25	0.0017	0.96	0.0289	0.0004	3.5456
245	2	26	0.0032	0.9231	0.0266	0.0003	3.6256

Appendix II: Continued

t	s_i	r_i	$\frac{s_i}{r_i(r_i - s_i)}$	$\frac{r_i - s_i}{r_i}$	$S(t) = \prod_{i=1}^{k}\left(1 - \frac{s_i}{r_i}\right)$	$Var[\hat{S}(t)] = \hat{S}(t)^2 \sum_{i=1}^{j} \frac{s_i}{r_i(r_i - s_i)}$	$H(t) = -In[S(t)]$
246	1	25	0.0017	0.96	0.0256	0.0003	3.6665
247	1	24	0.0018	0.9583	0.0245	0.0003	3.7090
256	1	21	0.0024	0.9524	0.0233	0.0002	3.7578
261	1	2	0.5	0.5	0.0117	0.0001	4.4510
264	1	25	0.0017	0.96	0.0112	0.0001	4.4918
266	1	14	0.0055	0.9286	0.0104	0.0001	4.5659
270	1	21	0.0024	0.9524	0.0099	9.32264E-05	4.6147
274	2	18	0.0069	0.8889	0.0088	7.14529E-05	4.7324
275	1	7	0.0238	0.8571	0.0075	5.46489E-05	4.8866
277	1	26	0.0015	0.9615	0.0073	5.17226E-05	4.9258
278	2	6	0.0833	0.6667	0.0048	2.49438E-05	5.3313
280	1	22	0.00225	0.9545	0.0046	2.27739E-05	5.3778
283	1	24	0.0018	0.9583	0.0044	2.0951E-05	5.4205
287	1	21	0.00241	0.9524	0.0042	1.90455E-05	5.4691
288	1	27	0.0014	0.9630	0.0041	1.76843E-05	5.5068
295	1	22	0.0022	0.9545	0.0039	1.61457E-05	5.5535
304	1	15	0.0048	0.9333	0.0036	1.41269E-05	5.6224
323	1	3	0.1667	0.6667	0.0024	7.24709E-06	6.0277
330	1	26	0.0015	0.9615	0.0023	6.70861E-06	6.0671
332	1	12	0.0076	0.9167	0.0021	5.67129E-06	6.1540
338	1	20	0.0026	0.95	0.0020	5.12906E-06	6.2056
339	1	28	0.0013	0.9643	0.0019	4.77425E-06	6.2420
357	2	16	0.0089	0.875	0.0017	3.68118E-06	6.3754
359	2	18	0.0069	0.8889	0.0015	2.9245E-06	6.493
362	1	11	0.0091	0.9091	0.0014	2.43416E-06	6.5886

Appendix II: Continued

t	s_i	r_i	$\frac{s_i}{r_i(r_i - s_i)}$	$\frac{r_i - s_i}{r_i}$	$S(t) = \prod_{i=1}^{k}\left(1 - \frac{s_i}{r_i}\right)$	$Var[\hat{S}(t)] = \hat{S}(t)^2 \sum_{i=1}^{j} \frac{s_i}{r_i(r_i - s_i)}$	$H(t) = -In[S(t)]$
377	1	24	0.0018	0.9583	0.0013	2.23869E-06	6.6309
402	1	4	0.0833	0.75	0.0019	1.3408E-06	6.9188
450	1	30	0.0011	0.9667	0.0001	1.25395E-06	6.9528
494	1	29	0.0012	0.9655	0.0009	1.17002E-06	6.9879
532	1	21	0.0024	0.9524	0.0009	1.06307E-06	7.0367

Appendix III: 95% Linear and Log-transformed Confidence Intervals for days until claim

DUC	S(t)	SE	Linear C.I	Log-transformed C.I
1	0.95	0.0487	(0.859257,1.040743)	(0.694734,0.992802)
3	0.9135	0.0590	(0.822719,1.004204)	(0.693195,0.977891)
4	0.8700	0.0704	(0.7499,0.990027)	(0.647235,0.956374)
11	0.8377	0.0748	(0.714898,0.960587)	(0.621364,0.93625)
20	0.7959	0.0820	(0.667999,0.923712)	(0.575376,0.909984)
27	0.3979	0.2031	(0.239490,0.556365)	(0.065391,0.732466)
28	0.3648	0.1876	(0.230672,0.498862)	(0.0646,0.689868)
36	0.3192	0.1696	(0.213094,0.425248)	(0.058293,0.631998)
37	0.2736	0.1484	(0.194014,0.353136)	(0.052689,0.565070)
41	0.2638	0.1434	(0.189659,0.337950)	(0.050261,0.552243)
43	0.2544	0.1386	(0.185285,0.323481)	(0.050471,0.533927)
45	0.2417	0.1322	(0.179027,0.304301)	(0.048693,0.513047)
46	0.2071	0.1178	(0.159327,0.254954)	(0.040957,0.460373)
50	0.1962	0.1121	(0.153132,0.239345)	(0.039235,0.440913)
55	0.1897	0.1085	(0.149345,0.230049)	(0.038256,0.428805)
60	0.1818	0.1043	(0.144632,0.218955)	(0.03699,0.41413)
62	0.1753	0.1008	(0.140677,0.209924)	(0.035949,0.401846)
70	0.1578	0.0922	(0.129257,0.186283)	(0.032264,0.370441)
72	0.1519	0.0890	(0.125432,0.178422)	(0.031269,0.358895)
75	0.1439	0.0847	(0.120052,0.167812)	(0.029798,0.343182)
77	0.1371	0.0809	(0.115342,0.158812)	(0.028535,0.329452)
80	0.1302	0.0771	(0.110534,0.149913)	(0.027240,0.315588)
85	0.1202	0.0715	(0.103353,0.137060)	(0.025375,0.294752)
96	0.1088	0.0652	(0.094864,0.122651)	(0.023118,0.270725)
100	0.1046	0.0628	(0.091702,0.117448)	(0.022307,0.261704)
106	0.0968	0.0584	(0.085747,0.107910)	(0.020786,0.244791)
116	0.0922	0.0558	(0.082134,0.102302)	(0.019841,0.234717)
121	0.0874	0.0531	(0.078277,0.096451)	(0.018824,0.224071)
129	0.0830	0.0506	(0.074766,0.091226)	(0.017909,0.214361)
136	0.0788	0.0482	(0.071392,0.0863)	(0.017035,0.205049)
137	0.0749	0.0461	(0.068142,0.081665)	(0.016150,0.1963587)
142	0.0722	0.0444	(0.065941,0.078517)	(0.015651,0.189912)
145	0.0695	0.0428	(0.063625,0.075276)	(0.015074,0.183445)
151	0.0667	0.0412	(0.061292,0.072053)	(0.014493,0.176949)
168	0.0636	0.0394	(0.058726,0.068558)	(0.013846,0.169859)
170	0.0599	0.0373	(0.055523,0.064274)	(0.013016,0.161166)
171	0.0532	0.0337	(0.049724,0.056762)	(0.011353,0.146488)
189	0.0503	0.0320	(0.047134,0.053437)	(0.010712,0.139342)
198	0.0486	0.0309	(0.045609,0.051494)	(0.0103546,0.135016)
203	0.0466	0.0297	(0.043892,0.049326)	(0.009947,0.130177)
205	0.0447	0.0286	(0.042235,0.047254)	(0.009555,0.1255062)
210	0.0431	0.0276	(0.04081,0.045484)	(0.009221,0.121461)
213	0.0416	0.0267	(0.039429,0.043782)	(0.008898,0.117543)
222	0.0395	0.0254	(0.037555,0.041496)	(0.008447,0.112311)

Appendix III: continued

DUC	S(t)	SE	Linear C.I	Log-transformed C.I
225	0.0382	0.0246	(0.036323,0.040002)	(0.008161,0.108799)
226	0.0363	0.0234	(0.034589,0.037920)	(0.007746,0.103947)
228	0.0340	0.0221	(0.032517,0.03546)	(0.007239,0.098225)
232	0.0317	0.0207	(0.030434,0.033011)	(0.006728,0.092479)
236	0.0301	0.0197	(0.028891,0.031215)	(0.006348,0.088231)
242	0.0289	0.0190	(0.027779,0.029923)	(0.006108,0.084934)
245	0.0266	0.0176	(0.025715,0.027548)	(0.005637,0.078985)
246	0.0256	0.0169	(0.02472,0.026413)	(0.005410,0.076111)
247	0.0245	0.0162	(0.023722,0.02528)	(0.005182,0.073231)
256	0.0233	0.0155	(0.022626,0.024043)	(0.004929,0.070082)
261	0.0117	0.0113	(0.011408,0.011926)	(0.001089,0.054812)
264	0.0112	0.0109	(0.010962,0.011439)	(0.001048,0.052824)
266	0.0104	0.0101	(0.010194,0.010607)	(0.000975,0.049432)
270	0.0099	0.0097	(0.009718,0.010093)	(0.000929,0.047346)
274	0.0088	0.0085	(0.008659,0.008950)	(0.000873,0.041593)
275	0.0075	0.0074	(0.007437,0.007656)	(0.000718,0.036921)
277	0.0073	0.0072	(0.007154,0.007358)	(0.000670,0.036132)
278	0.0048	0.0050	(0.004790,0.004885)	(0.000413,0.026056)
280	0.0046	0.0048	(0.004575,0.004661)	(0.000395,0.024971)
283	0.0044	0.0046	(0.004386,0.004465)	(0.000379,0.024014)
287	0.0042	0.0044	(0.004179,0.004251)	(0.0003661,0.022968)
288	0.00419	0.0042	(0.004025,0.004092)	(0.000348,0.022183)
295	0.0039	0.0040	(0.003844,0.003905)	(0.000333,0.021256)
304	0.0036	0.0038	(0.003589,0.003642)	(0.000310,0.019973)
323	0.0024	0.0027	(0.002398,0.002423)	(0.000172,10.01511)
330	0.0023	0.0026	(0.002306,0.002330)	(0.000166,0.014572)
332	0.0021	0.0024	(0.002115,0.002135)	(0.000152,0.013480)
338	0.0020	0.0023	(0.002009,0.002027)	(0.000144,0.012858)
339	0.0019	0.0022	(0.001938,0.001955)	(0.000139,0.012432)
357	0.0017	0.0019	(0.001697,0.001709)	(0.000122,0.011009)
359	0.0015	0.0017	(0.001507,0.001520)	(0.000108,0.009883)
362	0.0014	0.0016	(0.001372,0.001380)	(0.000098,0.009074)
377	0.0013	0.0015	(0.001315,0.001323)	(0.000094,0.008723)
402	0.001	0.0012	(0.000987,0.000991)	(0.000065,0.006971)
450	0.001	0.0011	(0.000954,0.000958)	(0.000063,0.006754)
494	0.0009	0.0011	(0.000921,0.000925)	(0.000061,0.006536)
532	0.0009	0.0010	(0.000877,0.000881)	(0.000058,0.006246)

Appendix IV: 95% Linear and Log-transformed Confidence Intervals for days until payment

DUP	S(t)	SE	Linear C.I	Log-transformed C.I
2	0.9688	0.0308	(0.910349,1.027151)	(0.798186,0.995538)
3	0.9284	0.0493	(0.838682,1.018089)	(0.739696,0.981854)
4	0.8897	0.0605	(0.784119,0.995286)	(0.693568,0.963361)
5	0.8185	0.0737	(0.700272,0.936781)	(0.616643,0.920404)
6	0.7795	0.0798	(0.657555,0.901544)	(0.572562,0.894741)
7	0.6860	0.0866	(0.569531,0.802475)	(0.483465,0.822481)
10	0.6548	0.0881	(0.541723,0.767919)	(0.454119,0.796851)
11	0.6286	0.0884	(0.519706,0.737550)	(0.431456,0.773861)
12	0.5762	0.0885	(0.476337,0.676149)	(0.386188,0.726613)
13	0.5541	0.0878	(0.458743,0.649416)	(0.368225,0.705425)
14	0.5115	0.0861	(0.425194,0.597722)	(0.334052,0.663645)
15	0.4475	0.0828	(0.374867,0.520184)	(0.282949,0.599269)
16	0.4281	0.0815	(0.3597,0.496436)	(0.267917,0.578919)
17	0.3964	0.0785	(0.335395,0.457323)	(0.244749,0.544186)
18	0.3791	0.0769	(0.339192,0.436293)	(0.231879,0.525386)
20	0.3336	0.0720	(0.286520,0.380742)	(0.199063,0.474012)
23	0.2780	0.0652	(0.242507,0.313545)	(0.159957,0.409030)
29	0.2681	0.0619	(0.235566,0.300627)	(0.156210,0.393209)
31	0.2547	0.0618	(0.223831,0.285553)	(0.144181,0.380648)
35	0.2441	0.0601	(0.215304,0.272855)	(0.137203,0.367409)
36	0.2246	0.0569	(0.199510,0.249596)	(0.124577,0.342623)
37	0.2177	0.0556	(0.194027,0.241470)	(0.120443,0.333563)
42	0.1988	0.0523	(0.178420,0.219208)	(0.108256,0.309215)
44	0.1902	0.0508	(0.171246,0.209093)	(0.102800,0.297890)
45	0.1743	0.0478	(0.158004,0.190640)	(0.092967,0.276773)
50	0.1609	0.0450	(0.146715,0.175111)	(0.083140,0.261357)
51	0.1539	0.0436	(0.140764,0.167070)	(0.080645,0.248845)
52	0.1405	0.0408	(0.129288,0.151777)	(0.072595,0.230357)
55	0.1317	0.0387	(0.121745,0.141753)	(0.067645,0.217575)
59	0.1258	0.0374	(0.116532,0.134989)	(0.064103,0.209134)
60	0.1207	0.0363	(0.112146,0.129315)	(0.061214,0.201867)
63	0.1110	0.0340	(0.103667,0.118477)	(0.055702,0.187804)
67	0.1071	0.0330	(0.100172,0.114038)	(0.053503,0.181882)
68	0.1026	0.0320	(0.096214,0.109070)	(0.050985,0.175294)
69	0.0984	0.0309	(0.092407,0.104324)	(0.048600,0.168910)
75	0.0944	0.0299	(0.088893,0.099968)	(0.046409,0.163021)
79	0.0877	0.0282	(0.082846,0.092526)	(0.042752,0.152690)
83	0.0835	0.0271	(0.079079,0.087950)	(0.040416,0.146432)
95	0.0802	0.0262	(0.076046,0.084294)	(0.038611,0.141279)
97	0.0771	0.0254	(0.073246,0.080927)	(0.036962,0.136474)
98	0.0736	0.0245	(0.070049,0.077116)	(0.035054,0.131083)
100	0.0704	0.0236	(0.067121,0.073645)	(0.033267,0.126257)
108	0.0677	0.0229	(0.064640,0.070712)	(0.031915,0.121799)
111	0.0646	0.0221	(0.061808,0.067393)	(0.030273,0.116970)

Appendix IV: continued

DUP	S(t)	SE	Linear C.I	Log-transformed C.I
112	0.0538	0.0208	(0.051634,0.056033)	(0.022633,0.105025)
116	0.0517	0.0201	(0.049642,0.053718)	(0.021644,0.101271)
122	0.0497	0.0194	(0.047799,0.051586)	(0.020741,0.097768)
127	0.0478	0.0188	(0.046021,0.049541)	(0.019874,0.094393)
133	0.0460	0.0182	(0.044372,0.047651)	(0.019078,0.091240)
158	0.0446	0.0177	(0.043071,0.046163)	(0.018464,0.088714)
166	0.0430	0.0171	(0.041524,0.044405)	(0.017725,0.085745)
188	0.0416	0.0166	(0.040225,0.042932)	(0.017115,0.083223)
198	0.0398	0.0161	(0.038525,0.041017)	(0.016295,0.079999)
200	0.0374	0.0152	(0.036315,0.038547)	(0.015198,0.075931)
220	0.0187	0.0153	(0.018156,0.019275)	(0.002619,0.069780)
232	0.0125	0.0114	(0.012199,0.012756)	(0.001369,0.054225)
260	0.0094	0.0090	(0.009194,0.009522)	(0.000932,0.043843)
290	0.0075	0.0074	(0.007378,0.007594)	(0.000707,0.036782)
300	0.0062	0.0062	(0.006162,0.006315)	(0.000572,0.031682)
350	0.0053	0.0054	(0.005291,0.005404)	(0.000481,0.027832)
352	0.0047	0.0048	(0.004635,0.004723)	(0.000415,0.024826)
380	0.0042	0.0043	(0.004124,0.004194)	(0.000366,0.022412)
390	0.0037	0.0039	(0.003715,0.003771)	(0.000327,0.020432)
405	0.0031	0.0032	(0.003099,0.003139)	(0.000270,0.017374)
415	0.0029	0.0030	(0.002879,0.002914)	(0.000251,0.016251)
423	0.0027	0.0028	(0.002688,0.002718)	(0.000234,0.015660)
435	0.0025	0.0027	(0.002521,0.002548)	(0.000219,0.014395)
444	0.0024	0.0025	(0.002382,0.002405)	(0.000207,0.013663)
453	0.0023	0.0024	(0.002257,0.002278)	(0.000196,0.013001)
500	0.0021	0.0022	(0.002084,0.002102)	(0.000180,0.012095)

APPENDIX V: Construction of a 95% Confidence Interval for the Median for DUC

t_i	$S(t_i)$	$\sqrt{\hat{V}[S(t_i)}$	Linear	Log
1	0.95	0.0487	9.2338	2.6034
3	0.9135	0.0590	7.0093	2.8534
4	0.8700	0.0704	5.2542	2.7617
11	0.8377	0.0748	4.5144	2.7057
20	0.7959	0.0820	3.6095	2.4619
27	0.3979	0.2031	-0.5025	-0.5140
28	0.3648	0.1876	-0.7210	-0.7354
36	0.3192	0.1696	-1.0664	-1.0733
37	0.2736	0.1484	-1.5260	-1.4959
41	0.2638	0.1434	-1.6471	-1.6023
43	0.2544	0.1386	-1.7723	-1.710
45	0.2417	0.1322	-1.9535	-1.8617
46	0.2071	0.1178	-2.4867	-2.2717
50	0.1962	0.1121	-2.7104	-2.4354
55	0.1897	0.1085	-2.8592	-2.5416
60	0.1818	0.1043	-3.0511	-2.6746
62	0.1753	0.1008	-3.2221	-2.7901
70	0.1578	0.0922	-3.7116	-3.0962
72	0.1519	0.0890	-3.9120	-3.2179
75	0.1439	0.0847	-4.2063	-3.3894
77	0.1371	0.0809	-4.4862	-3.5465
80	0.1302	0.0771	-4.7934	-3.7121
85	0.1202	0.0715	-5.3094	-3.9775
96	0.1088	0.0652	-6.0028	-4.3071
100	0.1046	0.0628	-6.2962	-4.4397
106	0.0968	0.0584	-6.9048	-4.7021
116	0.0922	0.0558	-7.3090	-4.8662
121	0.0874	0.0531	-7.7759	-5.0469
129	0.0830	0.0506	-8.24258	-5.2198
136	0.0788	0.0482	-8.7318	-5.3933
137	0.0749	0.0461	-9.2302	-5.5585
142	0.0722	0.0444	-9.6308	-5.6952
145	0.0695	0.0428	-10.0606	-5.8326
151	0.0667	0.0412	-10.5243	-5.9755
168	0.0636	0.0394	-11.0712	-6.1367
170	0.0599	0.0373	-11.8077	-6.3404
171	0.0532	0.0337	-13.249	-6.6801
189	0.0503	0.0320	-14.064	-6.8735
198	0.0486	0.0309	-14.6003	-6.9990
203	0.0466	0.02975	-15.2427	-7.1433
205	0.0447	0.0286	-15.9106	-7.2880
210	0.0431	0.0276	-16.5311	-7.4185
213	0.0416	0.0267	-17.1735	-7.5493
222	0.0395	0.0254	-18.1017	-7.7270
225	0.0382	0.0246	-18.7758	-7.8541
226	0.0363	0.0234	-19.7831	-8.0322

APPENDIX V: Continued

t_i	$S(t_i)$	$\sqrt{\hat{V}[S(t_i)}$	Linear	Log
228	0.0340	0.0221	-21.1002	-8.2483
232	0.0317	0.0207	-22.5902	-8.4761
236	0.0301	0.0197	-23.8184	-8.6517
242	0.0289	0.0190	-24.8578	-8.8092
245	0.0267	0.0176	-26.9563	-9.0975
246	0.0256	0.0169	-28.0888	-9.2445
247	0.0245	0.0162	-29.3151	-9.3971
256	0.0233	0.0155	-30.7729	-9.5690
261	0.0117	0.0113	-43.1553	-8.5343
264	0.0112	0.0109	-44.9567	-8.6472
266	0.0104	0.0101	-48.3873	-8.8473
270	0.0099	0.0097	-50.7587	-8.9747
274	0.0088	0.0085	-58.1091	-9.4691
275	0.0075	0.0074	-66.6153	-9.7429
277	0.0073	0.0072	-68.5142	-9.7465
278	0.0048	0.0050	-99.144	-10.5353
280	0.00462	0.0048	-103.806	-10.6616
283	0.0044	0.0046	-108.27	-10.7782
287	0.0042	0.0044	-113.605	-10.9105
288	0.0041	0.0042	-117.933	-11.015
295	0.0039	0.0040	-123.471	-11.1419
304	0.0036	0.0038	-132.067	-11.3223
323	0.0024	0.0027	-184.837	-11.6745
330	0.0023	0.0026	-192.148	-11.7784
332	0.0021	0.0024	-209.064	-11.9894
338	0.0020	0.0023	-219.884	-12.1226
339	0.0019	0.0022	-227.942	-12.2198
357	0.0017	0.0019	-259.714	-12.557
359	0.0015	0.0017	-291.492	-12.8594
362	0.0014	0.0017	-319.594	-13.0866
377	0.0013	0.0015	-333.293	-13.1995
402	0.001	0.0012	-430.951	-13.5978
450	0.001	0.0011	-445.655	-13.6878
494	0.0009	0.0011	-461.393	-13.7806
532	0.0009	0.0010	-484.088	-13.9067

APPENDIX VI: Construction of a 95% Confidence Interval for the Median for DUP

t_i	$S(t_i)$	$\sqrt{\hat{V}[S(t_i)}$	Linear	Log
2	0.9688	0.0308	15.2399	3.0832
3	0.9284	0.0493	8.6897	3.1249
4	0.8897	0.0605	6.4364	3.0571
5	0.8185	0.0737	4.3213	2.7611
6	0.7795	0.0798	3.5012	2.4889
7	0.6860	0.0866	2.1472	1.8186
10	0.6548	0.0881	1.7569	1.5509
11	0.6286	0.0884	1.4550	1.3234
12	0.5762	0.0885	0.8619	0.8227
13	0.5541	0.0878	0.6160	0.5976
14	0.5115	0.0861	0.1332	0.1324
15	0.4475	0.0828	-0.6335	-0.6446
16	0.4281	0.0815	-0.8828	-0.9012
17	0.3964	0.0785	-1.3207	-1.3510
18	0.3791	0.0769	-1.5712	-1.6057
20	0.3336	0.0720	-2.3093	-2.3371
23	0.2780	0.0652	-3.4055	-3.3492
29	0.2681	0.0619	-3.7460	-3.6566
31	0.2547	0.0618	-3.9680	-3.8296
35	0.2441	0.0601	-4.2548	-4.0648
36	0.2246	0.0569	-4.8409	-4.5255
37	0.2177	0.0556	-5.0782	-4.7068
42	0.1988	0.0523	-5.7549	-5.1920
44	0.1902	0.0508	-6.1027	-5.4292
45	0.1743	0.0478	-6.8192	-5.8936
50	0.1609	0.0450	-7.5324	-6.3286
51	0.1539	0.0436	-7.9379	-6.5612
52	0.1405	0.0408	-8.8057	-7.0299
55	0.1317	0.0387	-9.5055	-7.3960
59	0.1258	0.0374	-9.9957	-7.6309
60	0.1207	0.0363	-10.4543	-7.8461
63	0.1111	0.0340	-11.4347	-8.2806
67	0.1071	0.0331	-11.8958	-8.4780
68	0.1026	0.0320	-12.4361	-8.6963
69	0.0984	0.0309	-12.9949	-8.9135
75	0.0944	0.0299	-13.5556	-9.1251
79	0.0877	0.0282	-14.6418	-9.5197
83	0.0835	0.0271	-15.3533	-9.7520
95	0.0802	0.0262	-15.996	-9.9610
97	0.0771	0.0254	-16.639	-10.1638
98	0.0736	0.0245	-17.4034	-10.3878
100	0.0704	0.0236	-18.1687	-10.6046
108	0.0677	0.0229	-18.887	-10.806
111	0.0646	0.0221	-19.7416	-11.0277
112	0.0538	0.0208	-21.4071	-10.858
116	0.0517	0.0201	-22.2823	-11.0542

APPENDIX VI: Continued

t_i	$S(t_i)$	$\sqrt{\hat{V}[S(t_i)}$	Linear	Log
122	0.04992	0.0194	-23.1616	-11.2462
127	0.0478	0.0188	-24.0683	-11.4361
133	0.0460	0.0182	-24.976	-11.6207
158	0.0446	0.0177	-25.7555	-11.7785
166	0.04297	0.0171	-26.7241	-11.964
188	0.0416	0.0166	-27.6041	-12.1299
198	0.0398	0.0160	-28.7931	-12.3346
200	0.0374	0.0152	-30.4082	-12.578
220	0.0187	0.0153	-31.5431	-8.52738
232	0.0125	0.0114	-42.7989	-8.85725
260	0.0094	0.0090	-54.8081	-9.3175
290	0.0075	0.0074	-66.9632	-9.7380
300	0.0062	0.0062	-79.2045	-10.1174
350	0.0053	0.0054	-91.4838	-10.4558
352	0.0047	0.0048	-103.797	-10.764
380	0.0042	0.00427	-116.122	-11.0434
390	0.0037	0.0039	-128.464	-11.3002
405	0.0031	0.0032	-153.169	-11.7573
415	0.0029	0.0030	-164.604	-11.9487
423	0.0027	0.0028	-176.034	-12.1293
435	0.0025	0.0027	-187.44	-12.2977
444	0.0024	0.0025	-198.249	-12.4561
453	0.0023	0.0024	-209.043	-12.603
500	0.0021	0.0022	-225.911	-12.807

Appendix V and VI illustrate the calculations which enter into the construction of the 95% confidence intervals for the median. For Appendix V, the entry in the fourth column is the middle term in (3.19), namely, $\frac{0.95-.05}{0.048734}$=9.233805. The entry in the fifth column is the middle term in (3.20), namely $\frac{[In\{-In[0.95]\}-In\{-In[0.5]\}][0.95In[0.95]}{0.048734}$=2.603397. To find the linear 95% confidence interval, we find all those values of $t_i$which have a value, in column four between -1.96 and 1.96. The same procedure applies to Appendix VI.