KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY,KUMASI

A SURVIVAL ANALYSIS OF THE SURRENDER OF LIFE INSURANCE POLICIES OF LIFE INSURANCE COMPANIES IN GHANA

By

KWEKU DAVID

OSAFO AGYARE BENJAMIN

A THESIS SUBMITTED TO THE DEPARTMENT OF MATHEMATICS, KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF BACHELOR OF SCIENCE (BSc.HONS) IN ACTUARIAL SCIENCE

May 15, 2017

Declaration

I hereby declare that this submission is our own work towards the award of the BSc. Actuarial Science degree and that, to the best of our knowledge, it contains no material previously published by another person nor material which had been accepted for the award of any other degree of the university, except where due acknowledgment had been made in the text.

Kweku David	....	....
Student	Signature	Date

Osafo Agyare Benjamin	....	....
Student	Signature	Date

Certified by:

Dr. G. A. Okyere	....	....
Supervisor	Signature	Date

Certified by:

Dr. R. K. Avuglah	....	....
Head of Department	Signature	Date

Dedication

This study is dedicated to our dear parents who have supported our education from the basics through this level amid all difficulties. May The Almighty God bless them and replenish all their sacrifices.

Abstract

In recent years, life insurance companies in Ghana have had to grapple with the challenge of policy surrenders, as the rate keeps increasing to unprecedented levels by the day, in spite of the increasing education and awareness to boost the insurance penetration rate, which is currently below two per cent (2%). The factors having significant effects on surrender of life policies, however, are company specific. This study was therefore conducted to estimate the prospect of a life policy holder surrendering, and to identify the factors having significant effects on this surrender in the life insurance companies.

With regards to these study objectives, secondary data was obtained from the SIC Life from the period of 1st January, 1998 to 31st December 2016, constituting a data size of 73642 policyholders. The Kaplan-Meier estimates were computed for these data points to identify what time period experienced most surrender, and was evident that, most life policyholders tend to surrender within their third year of policy subscription, with a surrender probability of ...., signifying the highest among all other policy years. With age, gender, income level, customer's dissatisfaction with services provided, need for emergency fund, interest rate arbitrage and subscription onto new policies as the covariates under consideration, a Cox-Proportional Hazard(Cox-ph) was fitted. Results indicated that, age,..... do no significantly affect the likelihood of a policyholder surrendering his/her life policy at 5% significance level. However, at 5% level of significance, gender, income level, customer's dissatisfaction with services provided, need for emergency fund, interest rate arbitrage and subscription onto new policies proved to have significant effect on policy surrender.

Acknowledgements

Our profound gratitude firstly goes to the Almighty God who in His loving kindness and abundant grace made it possible for us to bring this project to a successful completion. We would also like to express our heartfelt appreciation to our supervisor, Dr. G A Okyere for his priceless support, pieces of advice, suggestions and recommendations through all the stages of this project. We also express much gratitude to Mr Seydo Leander Ataki(Head, Actuarial Department) of SIC life for tirelessly helping compile the data for this study. We are indeed most grateful to him. The entire staff of Mathematics department can in no way be left out. We thank them for their continuous inspiration and contribution to the success of this project. Lastly, we would like to thank our families, friends, the teaching assistants of Mathematics department and all our course mates at KNUST for their support, advice and prayers which has led to the successful completion of this programme.

Contents

List Of Abbreviations

NIC National Insurance Commission

IAIS International Association of Insurance Supervisors

EFH Emergency Fund Hypothesis

IRH Interest Rate Hypothesis

PRH Policy Replacement Hypothesis

SIC State Insurance Company

List of Tables

List of Figures

Chapter 1

INTRODUCTION

1.1 Introduction

This chapter consists of background of study, research problem statement, objectives of the study, methodology, thesis justification and organization.

1.2 Background Of Study

The earliest known instance of insurance dates back to the Babylonian period circa 2250 BC, when most of the commerce and business deals used to happen through sea routes. During this period the Babylonian developed a type of loan insurance for marine businesses. Example of which can be found in the Code of Hammurabi. Upon receipt of a loan to fund his shipment, a merchant will typically pay the lender an additional premium in exchange for the lender's guarantee to cancel the loan should the shipment be stolen or lost at sea. In effect the lender assumed the perils of the transit at a premium of rate of interest. The maritime loan therefore cannot be considered a stand - alone insurance contract, although the practice proved effectively enough spreading through Europe, the Mediterranean, to the Adriatic and the Low Countries. Insurance originally evolved as a commercial instrument for risk, and it was not until after 1666, as a result of the Great Fire of London that insurance for households, aptly named "Fire of London" emerged (Buckham et al., 2010). Insurance hence started from this notion and now it is far more than marine insurance all over the world. There are many more types of insurance which have entered the market over this period of time.

Insurance as defined by the Financial Consumer Agency of Canada, FCAC (2011) is a way of reducing your potential financial loss or hardship which can help cover the cost of unexpected events such as theft, illness or property damage. Insurance can also provide your loved ones with a financial payment upon your death. In insurance, an entity known as the insurer assumes the risk of certain unforeseen events in the future of the policyholder (insured). For the insurer to assume the full responsibility of the risk, the policyholder makes contractual payments of premium calculated by the insurer at specific times to the insurer. Various types of insurances policies are available in the market, among which are auto policies, life insurance policies, disability insurance policies and so on.

1.2.1 Types of Insurance Policies

Health Insurance:

Health insurance is the insurance against the risk of incurring medical expenses among individuals. For instance in Ghana, the National Health Insurance Scheme was introduced in the year 2003 to provide financial risk protection against the cost of quality basic health care for all residents in Ghana.

Auto Insurance Policy:

Auto insurance (also known as car insurance, motor insurance or vehicle insurance) is insurance for cars, trucks, motorcycles, and other road vehicles. Its primary use is to provide financial protection against physical damage and/or bodily injury resulting from traffic collisions and against liability that could also arise there from. The specific terms of vehicle insurance vary with legal regulations in each region. To a lesser degree vehicle insurance may additionally offer financial protection against theft of the vehicle and possibly damage to the vehicle, sustained from things other than traffic collisions, such as keying and damage sustained by colliding with stationary objects. Comprehensive and Third Party Property Insurance is sold separately to cover property damage additionally, and can include fire, theft, collision, and other property damage (Wikipedia, 2017).

Life Insurance Policy:

A life insurance policy is a contract with an insurance company where in exchange for premium payments, the insurance company provides a lump-sum payment, known as a death benefit, to beneficiaries upon the insured's death. Typically, life insurance is chosen based on the needs and goals of the insured.

Term life Insurance:

Term life insurance is designed to provide financial protection for a specific period of time, such as 10 or 20 years. With traditional term insurance, the premium payment amount stays the same for the coverage period. After that period, policies may offer continued coverage, usually at a substantially higher premium payment rate. Term life insurance is generally less expensive than permanent life insurance.

Whole life Insurance:

Whole life insurance is a type of permanent life insurance designed to provide lifetime coverage. Because of the lifetime coverage period, whole life usually has higher premium payments than term life. Policy premium payments are typically fixed, and, unlike term, whole life has a cash value, which functions as a savings component and may accumulate tax-deferred over time.

In Ghana and the world at large, life insurance companies usually face some complications such as the problem increasing surrender activities, inflation, delays in claim payments, fraudulent claims, price undercutting and among others. However, one which has raised much concern is surrender of life insurance policies since in Ghana most of the insurance policies have surrender option.

Surrendering of life insurance policies is when a policyholder stops payment of premiums and curtails the term of the insurance policy before maturity. Surrender option (an American-style put) thus allows the policyholder to sell back the contract to the original seller (the issuer) and receives a compensation value (surrender value), (Mac-Issaka, 2015). In recent years, life insurance companies in Ghana have had to grapple with the challenge of policy surrenders, as the rate keeps increasing to unprecedented levels by the day, in spite of the increasing education and awareness to boost the insurance penetration rate, which is currently below two per cent (2%). Staggering data available indicate that for the period of January to December, 2013, partial and full surrenders/cancellations were in the region of seventy three per cent (73%) (Zogbenu, 2016).

Surrendering activities have many implications on the operations of life insurance companies. Few of these are as follows:

- Insurance companies function through the concept of risk pooling and risk sharing. This means that the losses of the few are spread over the group; Average loss is substituted for actual loss. Whether this system functions successfully is dependent on the concept of "the law of large numbers", which states that the greater the number of exposures, the more likely the actual results will approach the expected results. The law of large numbers permits an insurer to estimate future losses with some accuracy. Lapsing of policies makes it difficult for insurance companies to construct accurate estimates. Lapsed policies have a detrimental effect on risk pooling and sharing so, if policies start lapsing then business is sure to deteriorate.

- Insurance companies tend to make profit out of premium received after some years a policy has completed. This is due to the fact that the first year of a policy is very expensive for the insurance company, compared with the premium taken in the first year since premium rates are determined at a uniform rate. It is a known fact that the insurer has to spend very heavily in the initial years in order to procure a policy in other words, "new business strain". The high initial cost incurred when issuing a policy is expected to be recovered through premium installments, paid over a number of years (generally, from three to five years). If the policies lapse soon after they are issued, the insurer has no way to recover those expenses. Such lapses lead to a rise in first-year premium to expense ratios, which leads to a fall in profitability for the insurer.

1.3 Background Of Study Area

The SIC Life Company originally existed as the Life Division of the reputable multi-line insurer - the State Insurance Company of Ghana Limited (SIC). SIC has been in operation since 1962 when it was registered as a Public Corporation. In 1995 however, it was converted into a public limited liability company as part of the Government of Ghana's initiative to divest part of its shareholding in all State Corporations.

In compliance with the Insurance Law 2006, Act 724, the Life Division of the reputable State Insurance Company of Ghana Limited (SIC) became SIC LIFE COMPANY LIMITED (SIC Life) in 2007. Currently, SIC LIFE is the largest and most reputable insurance company in Ghana controlling the largest share of the insurance market for both Life and Non-Life Insurance business lines.

The Company is the leader in the Life Insurance industry in terms of key performance indicators such as capital base, shareholders' fund, total assets, profitability and market share. With its solid financial base, coupled with highly motivated and experienced human resource, SIC Life is well positioned to maintain its dominance of the Life Insurance Market through prudent management and sound technical practices.

SIC Life enjoys a favourable image within the Life Insurance industry both locally and internationally. SIC Life has a wide network of Area and Branch Offices throughout the country. It also has a large Agency Force whose focus is to respond professionally to the varying needs of the insuring public.

1.4 Problem Statement

Insurance companies over the years have developed certain expertise which they place at the disposal of their clients which has come to be known as maximum efficiency. Insurance policies are hence designed by insurance companies to help policyholders hedge against certain unforeseen future occurrences by transferring the risk form the policyholder to the insurance company. For the insurer to assume the full risk (responsibility to pay the insured a sum of money or the equivalence of the lost property), the policyholder is made to pay premiums calculated by the insurer at some dedicated times usually monthly since most individuals receive income with that frequency. In Ghana, premiums for life insurance policies are calculated based on some factors such as age of the insured, sex, term of the policy, health condition of the insured and among few other determinants.

Despite the importance of policy surrenders, most insurance companies have not yet tracked or organized their lapse data in a manner that allows them to accurately predict surrendering dynamics and their main causes. Even if they have, their models have not been too effective, since surrender activity in life insurance companies in Ghana is on the rise with available data indicating that for the period of January to December, 2013, partial and full surrenders/cancellations were in the region of seventy three per cent (73%) (Zogbenu, 2016).

1.5 Objectives Of Study

1. To use the Kaplan-Meier estimator to estimate the probability of a life insurance policy holder surrendering.

2. To determine the factors having significant effect on surrender of life insurance policies using the Cox proportional hazard model.

1.6 Justification of Study

The study focuses on one of the major problems in the life insurance sector, surrender of life policies. The study thus considers the Kaplan-Meier estimator and the Cox Proportional hazard model to evaluate surrender probabilities and the significant factors affecting surrender of policies respectively. Hence the introduction of survival concepts.

1.7 Thesis Organization

This thesis is organized into five main chapters. Chapter 1 presents introduction of the thesis. This consists of background of study, research problem statement, objectives of the study, methodology, thesis justification and organization. Chapter 2 is literature review, which briefly looks at works done by other researchers on the topic. Chapter 3 is formulation of the mathematical model. Chapter 4 deals with analysis of data collected, formulation of model instances, algorithms, computational procedures, results and discussion. Chapter 5 looks at summary, conclusions and recommendation of the results.

Chapter 2

LITERATURE REVIEW

2.1 Introduction

This chapter consist of the some definitions and meaning of insurance, a brief history of insurance in Ghana, the National Insurance Commission(NIC) and the composition of the life and non-life insurance. The chapter also considers the general overview of the life insurance sector of Ghana and its challenges. Review of existing literature on the drivers of surrender of life insurance policies is also done in this chapter.

2.2 Definition And Meaning Of Insurance

Scholars have endlessly contended about what insurance is, and whether its essential nature is transfer, pooling, some combination of the two, or something else altogether (Denenberg, 1963). This has led to many definitions as to what insurance is.

According to Brooker et al. (2012), insurance, in law and economics, is a form of risk management primarily used to hedge against the risk of a contingent loss. Insurance hence may be defined as the equitable transfer of the risk of a loss, from one entity to another, in exchange for a premium. An insurer typically is a company selling the insurance; and an insured typically is the person or entity buying the insurance.

Müller (1981) also defined insurance as a device for the reduction of uncertainty of one party, called the insured, through the transfer of particular risks to another party, called the insurer, who offers a restoration, at least in part, of economic losses suffered by the insured.

Again, insurance as explained by the Financial Consumer Agency of Canada, FCAC(2011) is a way of reducing your potential financial loss or hardship which can help cover the cost of unexpected events such as theft, illness or property damage. Insurance can also provide your loved ones with a financial payment upon your death. In insurance, an entity known as the insurer assumes the risk of certain unforeseen events in the future of the policyholder (insured).

Dorfman (2008) in financial perspective viewed insurance as a financial arrangement that redistributes the costs of unexpected of unexpected losses. It involves the transfer of potential losses to an insurance pool which combines all the potential losses and then transfer the cost of the predicted losses back to those exposed. Thus insurance involves the transfer of loss exposures to an insurance pool, and the redistribution of losses among the members of the pool. An insurance system redistributes the cost of losses by collecting a premium payment from every participant(insured) in the system. In exchange for the premium payment, the insurer promises to pay the insured's claim in the event of a covered loss.

This characterization of insurance stresses a major aspect which can be found implicitly or explicitly in most the definitions, namely the reduction of risk through some transfer mechanism (Müller, 1981). Thus though different definitions have been given by various scholars, they all seem to address the same issue in several ways.

2.3 Brief History Of Insurance In Ghana.

The 19th century saw the British merchants trading with the Colonies of Great Britain, where goods being shipped into the British colonies were to be carried by ships owned by British citizens as stated by the law. By implication, the goods being carried by the ships were insured by insurance companies in the United Kingdom. As such the agents of the insurance companies in the UK came to represent their companies in Ghana where the goods were sent. Thus insurance transactions were done through the foreign trading companies in Ghana acting as chief agents of insurance companies in the United Kingdom and other foreign countries. Ghanaian insurance industry at the time comprised mainly: insurers; sellers of insurance, insured; purchasers of insurance and intermediaries; agents of insurance companies who acted between the insurers and the insured and were also legible to accept proposal, sign and issue insurance cover on behalf of the insurance companies in the United Kingdom under the Act of British Parliament. There were no insurance broking, claim adjusting and reinsurance firms at the time. Because all insurance transactions were done in the United Kingdom and elsewhere, there were no local insurance legislations in Ghana. For instance, premium rates for various life insurance policies in Ghana were based on those of United Kingdom because there were no local statistics and tables on premium rates in Ghana at the time.

Towards Ghana's independence in 1957, local insurance companies began to emerge. The first among them was Gold Coast Insurance Company which was formed in 1955. General Insurance Company and Cooperative Insurance Society also followed in the years 1957 and 1958 respectively. Later Government of Ghana purchased Gold Coast Insurance Company and took over Cooperative Insurance Society and merged them to form the State Insurance Corporation (SIC) which was incorporated in February 1962. This was a strategic development to enable SIC compete favorably with foreign insurance companies. Most of the insurance businesses of Government owned organizations were handled by SIC; as a result, it grew rapidly. This resulted in remarkable changes in Ghanaian insurance industry between the periods of 1962 and 1970. A lot of rules and regulations were introduced and Acts were passed into laws in the insurance industry. The new insurance Act created the office of the commissioner of Insurance to regulate the operations of the insurance in Ghana.Bancassurance was also later introduced into the insurance market in Ghana. Under this agreement, the bank as a corporate agent will use its branch network to sell the insurance products of an insurance company. Example of such cases is the agreement between Enterprise Life and Standard Chartered Bank which was approved in 2007. The Number of agents has also increased. Seven hundred new agents were licensed in 2008, bringing the total number of agents officially registered with the NIC to 1,200 in 2009, although it is estimated that there are about 4,000 agents selling insurance products in the insurance markets in Ghana. The era also saw innovative products sprung up. Products such as family income protection plan, education endowment policy and funeral insurance policy. Currently, the industry is experiencing influx of companies from foreign countries and this has brought keen completion to the industry (Afriyie, 2006).

2.4 The National Insurance Commission

National Insurance Commission (NIC) was established under Insurance Law 1989 (PNDC Law 227), but now operates under Insurance Act, 2006 (Act 724). The object of the Commission, as detailed in Act 724 is to ensure effective administration, supervision, regulation and control the business of Insurance in Ghana. NIC is mandated to perform a wide spectrum of functions including licensing of entities, setting of standards and facilitating the setting of codes for practitioners. The Commission is also mandated to approve rates of insurance premiums and commissions, provide a bureau for the resolution of complaints and arbitrate insurance claims when disputes arise. Other responsibilities include the provision of recommendation to the sector Minister for policy formulation, supervision of practitioners, enforcement of compliance and public education. The development of strong relationships with regulators from other countries and international bodies such as the International Association of Insurance Supervisors (IAIS) and ensuring the conformity of practitioners to internationally accepted standards are also key mandates of the Commission. The enactment of Act 724 was a major milestone towards a robust insurance regulatory environment as it empowers and grants adequate powers to the commission. Together there are several initiatives in the past decade; the new Law provides a strong regulatory framework for Ghana's insurance industry (NIC, 2017)

2.5 Life and Non-life Insurance Businesses

Various attempts were made to classify insurance into different categories by various individuals and institutions. Some classified insurance under the categories of life, fire, marine and miscellaneous. Others classified it under the sub-titles such as insurance of person, insurance of property, insurance of interest and insurance of liability. In a broad term, insurance can be classified into Life insurance and Non-life (General) insurance. Life insurance is a contract in which the insurer, in consideration of certain premium, either in a lump sum or by other periodical payments, agrees to pay to the assured, or to the person for whose benefit the happening of a specified event contingent on the human life or at the expiry of certain period. The monthly premiums for a life insurance are generally based upon the age, health, and occupation information of the applicant, in addition to the total benefits to be paid to him for his policy. For life insurance, the risk ensured against is death. The life insurance company pays the sum assured to the insured in the event of death.

As stated by Pal, Bolda and Garg (2007), Life insurance enjoys maximum scope because life is the most important property of the society or an individual. Each and every person requires insurance, thus Life insurance provides protection to the family at the premature death or gives adequate amount at the old age when earning capacities are reduced. The insurance is not only a protection but a sort of investment because a certain sum is returnable to the insured at death or the expiry of a period. General insurance on the other hand is basically policy that protects the insured against losses and damages other than those covered by Life insurance. In Ghana, insurance is classified under the two broad names; Life and General (Non-life). Various policies are classified under these two broad titles. For instance liability and engineering policies are classified under General insurance. Life insurance policy in Ghana is no longer insurance against death but also includes endowment, funeral and other policies. Most companies have adopted products innovation as a strategy to gain competitive advantage and this consequently has increased insurance products especially Life insurance in the industry(Afriyie, 2006).

2.6 Overview Of The Life Insurance Sector.

The insurance sector is growing rapidly as a result of reforms, improved regulations and a better understanding of the relevant product on the part of Ghana's consumers, and the market has been a star performer. This is seen as total premiums rose to 23% in 2013 over 2012, and the 5-year average growth for the sector was 32%, according to the insurance commissioner as quoted in the local press. Business has not only expanded along with the economy, but also out performed it (Group, 2014). The Ghanaian life insurance market is consolidated with the top five companies accounting for about 80% share in 2012 (up from 78% in 2011). There were 19 life insurers as at May 2014. In 2012, the life insurance gross written premium (GWP) stood at GHS355.8 million, witnessing a growth of 31.7%. Ghana's top three life insurers' performance in 2012 is summarized below:

- SIC Life Insurance collected GHS100.3 million representing a premium growth of 39%. SIC is listed on the Ghana Stock Exchange and owned by government (40%) and others including institutions and individuals (60%)

- Enterprise Life Assurance reported total premium collection of GHS89.1 million, representing a growth of 45%.

- Glico Life Insurance reported a gross premium of GHS36 million witnessing 4.65% growth.

The life insurance GWP witnessed a growth of 31.7%. There is still room for growth in this sector especially with the development of micro insurance and innovative distribution channels such as banc assurance.

The portfolio mix (proportion of life and non-life premium to the total premium) has shifted from 32/68% in 2007 to 42/58% in 2012 The portfolio mix is expected to witness a further structural shift in favour of life insurance segment The future growth of life insurance segment is expected to be driven by a large untapped market and an established regulatory framework, which is making Ghana a lucrative market for international players (KPMG, 2014).

2.7 Challenges Facing the Life Insurance Sector.

Despite the increase in performance and the penetration of life insurance business in the country, the sector is facing a couple of challenges over the years. Among these are:

- The problem of inflation: one of the problems facing the insurance industry is the abnormal increases in prices which has taken place in the last few years. Although this has been a world-wide phenomenon, the higher prices has had adverse effect on expense loading in premium rates which has led to policyholders having low confidence in the life policies.

- The next problem which has raised much concern is the problem of surrendering of life insurance policies. Surrendering of life insurance policies is when a policyholder stops payment of premiums and curtails the term of the insurance policy before maturity. Surrender option (an American-style put) thus allows the policyholder to sell back the contract to the original seller (the issuer) and receives a compensation value (surrender value),(Mac-Issaka, 2015). In recent years, life insurance companies in Ghana have had to grapple with the challenge of policy surrenders, as the rate keeps increasing to unprecedented levels by the day, in spite of the increasing education and awareness to boost the insurance penetration rate, which is currently below two per cent (2%). Staggering data available indicate that for the period of January to December, 2013, partial and full surrenders/cancellations were in the region of seventy three per cent (73%) (Zogbenu, 2016).

Surrendering activities have many implications on the operations of life insurance companies. Few of them are as follows:

- Insurance companies function through the concept of risk pooling and risk sharing. This means that the losses of the few are spread over the group; Average loss is substituted for actual loss. Whether this system functions successfully is dependent on the concept of "the law of large numbers", which states that the greater the number of exposures, the more likely the actual results will approach the expected results. The law of large numbers permits an insurer to estimate future losses with some accuracy. Lapsing of policies makes it difficult for insurance companies to construct accurate estimates. Lapsed policies have a detrimental effect on risk pooling and sharing so, if policies start lapsing then business is sure to deteriorate.

- Insurance companies tend to make profit out of premium received after some years a policy has completed. This is due to the fact that the first year of a policy is very expensive for the insurance company, compared with the premium taken in the first year since premium rates are determined at a uniform rate. It is a known fact that the insurer has to spend very heavily in the initial years in order to procure a policy in other words, "new business strain". The high initial cost incurred when issuing a policy is expected to be recovered through premium installments, paid over a number of years (generally, from three to five years). If the policies lapse soon after they are issued, the insurer has no way to recover those expenses. Such lapses lead to a rise in first-year premium to expense ratios, which leads to a fall in profitability for the insurer.

2.8 Review Of Existing Literature

With surrendering of life policies becoming a limiting factor to the development of life insurance sector, a proper understanding of lapse drivers and the underlying dynamics is important for insurance managers and regulators as lapse influences an insurer's liquidity and profitability through acquisition cost, adverse selection, and cash surrender values Eling and Kiesenbauer (2012).

Our study contributes to existing literature through the investigation of the factors having significant effect on surrender of life insurance policies. Eling and Kiesenbauer (2014) puts it that existing empirical literature on lapse can be distinguished based on the explanatory variables considered. The first set of literature uses environmental characteristics including macroeconomic indicators and company data. These literature materials initially studied only the impact of interest rates and unemployment on lapse; referred to as Interest Rate Hypothesis (IRH) and Emergency Fund Hypothesis (EFH) respectively.

Outreville (1990) in his work, analyzed the effects of macroeconomic variables and life insurance industry trends on early lapsation. Using U.S. and Canadian data from the period 1955 –79, Outreville found that unemployment has a significantly positive effect on early lapsation while personal income has a significantly negative effect. He finds no significant relationship between interest rates and ordinary life insurance policies that lapse within 13 months of the policy issue date. Therefore, Outreville finds support for the EFH but not for the IRH.

Dar and Dodds (1989) explored the relationships amongst interest rates, unemployment, and net flow of funds into endowment life insurance policies and surrender activity in the United Kingdom.19 Using data from 1952 through 1985, they found net cash flows into endowment policies were directly related to the difference between rates on those policies and rates on alternative investments. While they also reported a direct relation between surrender activity and unemployment, no such relationship was identified between interest rates and surrender activity. Based on the findings from various models and tests, Dar and Dodds concluded that interest rate considerations drive life insurance savings decisions while "emergency" cash needs drive surrender activity, which supports the EFH.

Outreville's work was extended by Kim (2005a,b), Cox and Lin (2006), and Kiesenbauer (2012) considering additional economic indicators (such as gross domestic product and capital markets development) and company characteristics (including company size and legal form).

Kuo et al. (2003) tested the relative importance of the EFH and the IRH on life insurance lapse behavior using aggregate data from the ACLI for the period 1951 through 1998. The authors first report evidence was consistent with support for both the IRH and the EFH. However, while the results support both of these hypotheses, the authors also found that interest rates have a greater economic impact on lapse behavior.

Utilizing monthly cash flow data that allowed unbundling of a universal life policy's cash flows into premiums, loans, and surrenders, Hoyt (1994) examined the factors that drive surrender activity. He found that unemployment was the macroeconomic variable with the most explanatory power for universal life surrenders, lending support to the EFH (Russell et al., 2013).

Other literature on surrender of life policies considered other factors as the drivers of policy lapsation. Outreville (1990) used the ratio of new business to existing business as a proxy for replacements and found that this measure was directly related to lapsation. Outreville's replacement theory is referred to as the Policy Replacement Hypothesis (PRH). The PRH contends that consumers surrender policies for the purpose of replacing the original policy with one that has a better price or more favorable terms.

Eling and Kiesenbauer (2014) in their contribution to existing literature on the drivers of policy surrender considered remaining policy duration, distribution channel and supplementary cover, which were deemed significant lapse drivers. In their findings, they concluded that the considered product and policyholder characteristics have a statistically significant impact on the lapse rate development, but the magnitude of the effects varies. The largest variations are observed for calendar year, contract age, remaining policy duration, and premium payment (single vs. regular). They saw that the direction of impact was consistent with the existing literature (Renshaw and Haberman, 1986; Kagraoka, 2005; Cerchiara et al., 2009; Milhaud et al.,2010), except for product type which has only a limited effect on lapse rates.

Russell et al. (2013) also utilized state level data from 1995 through 2009 to test each of hypotheses (EFH, IRH and PRH). They modeled life insurance surrender activity as a function of the need for cash as the result of household liquidity constraints or interest rate/investment arbitrage opportunities, and life insurance market dynamics. The analysis considered variables such as unemployment (to test EFH), interest rate (to test IRH), real per capita income (to test EFH), policy replacement activities (to test PRH), age and among others. There was evidence of a positive correlation between short term interest rates and surrender activity, consistent with the notion of the Interest Rate Hypothesis that individuals may surrender policies for the purpose of exploiting higher market rates. Their findings also provided some support for EFH and PRH. Therefore according to Russell et al. (2013), insureds tend to surrender policies when market interest rates increase, when real per capita income decreases, and when policy replacement activity increases.

Chapter 3

METHODOLOGY

3.1 Introduction

This chapter consist of the methodology used to implement the survival analysis methods used in the valuation of surrender of life insurance policies using the Cox PH model. In order to understand the Cox PH model of examining and modeling, it is necessary to have a good understanding of some methods used in examining the time to event data. The branch of statistics which deals with analysis of time to the occurrence of one or more events such as death in biological organisms and patient's response to a therapy is Survival analysis. Survival analysis attempts to answer questions such as: what proportion of a population will survive past a certain time?,Of those that survive, at what rate will they die or fail?, Can multiple causes of death or failure be taken into account?, How do a particular characteristics increase or decrease the probability of survival?, etc. Survival analysis is widely employed in various field; in engineering, this concept is called *reliability theory* or *reliability analysis*, and in economics or sociology, it is referred to as *duration analysis* or *duration modeling.*

3.2 Concept Of Survival

Survival time data measure the time to a certain event, such as failure, death, response, relapse, the development of a given disease, parole, or divorce. These times are subject to random variations, and like any random variables, form a distribution. The distribution of survival times is usually described or characterized by three functions: (1) the survivorship function, (2) the probability density function, and (3) the hazard function. These three functions are mathematically equivalent?if one of them is given, the other two can be derived.In practice, the three functions can be used to illustrate different aspects of the data. A basic problem in survival data analysis is to estimate from the sampled data one or more of these three functions and to draw inferences about the survival pattern in the population. In Section 3.4 we define the three functions and discussed the equivalence relationship among the three functions.

3.3 Censoring

In estimating the life time distribution, in a perfect world we would be able to set-up an experiment such that at time 0 we would be able to observe a large number of identical new born lives. We would then be able to observe these individuals throughout the years and note when they die. At any time,t, we would then be able to determine that an estimate of the probability of survival to this age is the number of lives still alive divided by the number of lives at the start of the experiment. Unfortunately we have problems with this simple approach.

As humans now live for over 100 years, using this method to study human mortality will lead to the experiment taking over 100 years to complete. This also leads to the additional problem that this approach gives a retrospective measure of mortality and hence the mortality rates at the younger ages will no longer be applicable if mortality rates have changed in the intervening years (which we would expect).

Moreover in practice, it is likely to be impossible to observe all lives from birth until death. Many lives will be lost to the investigation before death and to exclude such lives from investigation may bias the result. This problem is known as **censoring**.

3.3.1 Types Of Censoring

As we have seen, data can be censored for a number of reasons and the effect it will have on our results will vary so we need to classify the types of censoring present in our data.

Right-censoring

Censoring mechanism cuts short observations in progress e.g. ending the investigation on a fixed date. (If you think of a time line then the line is abruptly cut off and we lose information to the right of the date that the investigation ends.)

Right censoring also occurs when:

- life insurance policyholders surrender their policies.
- active lives of a pension scheme retire.
- endowment assurance policies mature.

Left-censoring

Censoring mechanism prevents knowledge of when entry into the particular state occurred e.g. in medical studies, lives will be subject to regular medical examinations. Discovery of a condition only tells us that onset occurred in the period since the previous examination date so when measuring the number of days someone will survive with a certain disease we may not know when the disease actually starts e.g. we don?t know the start of the left side of the time line.

Left censoring occurs, for example:

- when estimating functions of exact age and you don't know the exact date of birth.
- when estimating functions of exact policy duration and you don't know the exact date of policy entry

- when estimating functions of the duration since onset of sickness and you don't know the exact date of becoming sick.

Interval-censoring

Censoring mechanism only allows us to say that a particular event occurred within some known interval of time e.g. in mortality investigations, we may only know the calendar year of death (rather than the exact date of death). Again, if we think of time lines we can only know the right end of the line occurs somewhere within a calendar year.

Further examples of interval censoring include the following situations:

- when you only know the calendar year of withdrawal.
- when estimating functions of exact age and you only know that deaths were aged x nearest birthday at the date of death

Random censoring

As stated above people may unexpectedly leave an investigation through an exit that is not the one being studied e.g. if we are using policyholders then a policyholder who cancels their life assurance policy will no longer be part of the investigation i.e. they exited the investigation without dying. We can introduce a bit of notation now and state that the time at which observation of life i is censored is a random variable,C^i.

Then, the observation will be censored if $C^i < T^i$, where T^i is the random variable denoting the future lifetime of life i . Of course, exits other than death may also be of interest (e.g. sickness in pension schemes) so we may find that we need to include these exits as well in a multiple decrement model.

Examples of random censoring include:

- life insurance withdrawals.
- members of a company pension scheme may leave voluntarily when they move to another employer.

Informative and non-informative censoring

Censoring is non-informative if it gives no information about the lifetimes Ti.

This just means that the mortality of the lives that remain in the at-risk group is the same as the mortality of the lives that have been censored. In the case of random censoring, the independence of each pair T_i ,C_i is sufficient to ensure that the censoring is non-informative. Informative censoring is more difficult to analyse, essentially because the resulting likelihoods cannot usually be factorised.

Examples of informative censoring include:

- Withdrawal of life insurance policies, because these are likely to be in better average health than those who do not withdraw. So the mortality rates of the lives that remain in the at-risk group are likely to be higher than the mortality rates of the lives that surrendered their policies.
- Ill-health retirements from pension schemes, because these are likely to be in worse than average health than the continuing members. So the mortality rates of those who remain in the pension scheme are likely to be lower than the mortality rates of the lives that left through ill-health retirement.

An example of non-informative censoring is:

- The end of the investigation period (because it affects all lives equally, regardless of their propensity to die at that point).

Type I censoring

The censoring time, C_i , for each life i is known in advance e.g. in the study of members of an occupational pension scheme, the only other decrement may be retirement at a fixed normal retirement age (NRA).

Type II censoring

The investigation is continued until a specified number of deaths have occurred.

This can simplify the analysis as the number of deaths is non-random.

3.4 Survival Analysis

To study, we must introduce some notation and concepts for describing the distribution of "time to event" for a population of individuals. Let T denotes the positive random variable representing time to event of interest.

Cumulative Distribution function is $F(t) = Pr(T \leq t)$ with probability density function $f(t) = F'(t)$.

Survival function is

$$S(t) = P(T > t) = 1 - F(t)$$

Note: we use $S(t) = \bar{F}(t)$ throughout.

Hazard function

$$h(t) = \lim_{\delta t \to 0} \left(\frac{Pr(t \leq T < t + \delta t | T \geq t)}{\delta t} \right)$$

If T is discrete and positive integer-valued then $h(t) = Pr(T = t | T \geq t) = Pr(T = t)/S(t-1)$.

Cumulative hazard function

$$H(t) = \int_0^t h(s)ds$$

We have the following **relations** between these functions:

(i)

$$
\begin{aligned}
h(t) &= \lim_{\delta t \to 0} \left(\frac{S(t) - S(t+\delta t)}{\delta t S(t)} \right) \\
&= -\frac{S'(t)}{S(t)} \\
&= -\frac{d}{dt}(logS)
\end{aligned}
$$

(ii)

$$S(t) = \exp(-H(t)), \text{ since } S(0) = 1$$

(iii)

$$f(t) = h(t)S(t)$$

3.5 Estimating Survival Distribution

From section 2, we saw that due to the existence of the problem of censoring in data, the survival distribution cannot be solely estimated using the simple approach, where the lives of a large number of new born would be observed throughout the years and note when they die, in order to estimate the probability of survival to an age at any time,t. To deal with this problem of estimating the survival distribution, censored observations are considered when estimating the survival distribution. Two methods are used in estimating the survival distribution:

Under a non-parametric approach, we make no prior assumptions about the shape or form of the distribution.

Under a parametric approach, we assume that the distribution belongs to a certain family (eg normal,exponential,Weibull,etc.) and use the data to estimate the appropriate parameters (eg mean and variance).

3.5.1 Likelihood and Censoring

If the censoring mechanism is independent of the event process, then we have an easy way of dealing with it.

Suppose that T is the time to event and that C is the time to the censoring event.

Assume that all subjects may have an event or be censored, say for subject i one of a pair of observations $(\tilde{t_i}, \tilde{c_i})$may be observed. Then since we observe the minimum time we would have the following expression for the likelihood (using independence)

$$L = \prod_{\tilde{t_i} < \tilde{c_i}} f(\tilde{t_i}) S_C(\tilde{t_i}) \prod_{\tilde{c_i} < \tilde{t_i}} S(\tilde{c_i}) f_C(\tilde{c_i})$$

Now define the following random variable:

$$\delta = \begin{cases} 1 & if\ T < C \\ 0 & if\ T > C \end{cases}$$

For each subject we observe $t_i = \min(\tilde{t_i}, \tilde{c_i})$and δ_i, observations from a continuous random variable and a binary random variable. In terms of these L becomes

$$L = \prod_i h(t_i)^{\delta_i} S(t_i) \prod_i h_C(t_i)^{1-\delta_i} S_C(t_i)$$

3.5.2 Non-parametric Estimators

Under a non-parametric approach, we make no prior assumptions about the shape or form of the distribution. Thus it is not assumed that the data follows a particular known distribution. The two most widely used non-parametric methods used in estimating the survival function are Kaplan-Meier(product-limit) and the Nelson-Aalen methods of estimate.

If there are observations $x_1, ..., x_n$ from a random sample then we define the empirical distribution function

$$\hat{F}(x) = \frac{1}{n}\#\{x_i : x_i \leq x\}$$

This is appropriate if no censoring occurs. However if censoring occurs this has to be taken into account. We measure the pair (X, δ) where $X = \min(T, C)$ and δ is as before

$$\delta = \begin{cases} 1 & if\ T < C \\ 0 & if\ T > C \end{cases}$$

Suppose that the observations are (x_i, δ_i) for $i = 1, 2, ..., n$.

$$\begin{aligned} L &= \prod_i f(x_i)^{\delta_i} S(x_i)^{1-\delta_i} \\ &= \prod_i f(x_i)^{\delta_i} (1 - F(x_i))^{1-\delta_i} \end{aligned}$$

What follows is a heuristic argument allowing us to find an estimator for $S(t)$,the survival function, which in the likelihood sense is the best that we can do. Suppose that there are failure times $(0 <) < t_1 < ... < t_i < ...$ Let $s_{i1}, s_{i2}, ..., s_{ic_i}$ be be the censoring times within the interval $[t_i, t_{i+1})$ and suppose that there are d_i failures at time t_i (allowing for tied failure times). Then the likelihood function becomes

$$\begin{aligned} L &= \prod_{fail} f(t_i)^{d_i} \prod_i \left(\prod_{k=1}^{c_i} (1 - F(s_{ik})) \right) \\ &= \prod_{fail} (F(t_i - F(t_i-))^{d_i}) \prod_i \left(\prod_{k=1}^{c_i} (1 - F(s_{ik})) \right) \end{aligned}$$

where we write $f(t_i) = F(t_i) - F(t_i-)$; the difference in the cdf at time t_i and the cdf immediately before it. Since $F(t_i)$ is an increasing function, and assuming that it takes fixed values at the failure time points, we make $F(t_i-)$ and $F(s_{ik})$ as small as possible in order to maximise the likelihood. That means we take

$F(t_i-) = F(t_{i-1})$and $F(s_{ik}) = F(t_i)$. This maximises L by considering the cdf $F(t)$ to be a step function and therefore to come from a discrete distribution, with failure times as the actual failure times which occur. Then

$$L = \prod_{fail}(F(t_i) - F(t_i))^{d_i} \prod_{i}(1 - F(t_i))^{c_i}$$

So we have showed that amongst all cdf's with fixed values $F(t_i)$ at the failure times t_i; then the discrete cdf has the maximum likelihood, amongst those with d_i failures at t_i and c_i censorings in the interval $[t_i, t_{i+1})$. Let us consider the **discrete case** and let

$$\Pr(\text{fail at } t_i|\text{survived to } t_i-) = h_i$$

Then

$$S(t_i) = 1 - F(t_i) = \prod_{j=1}^{i}(1 - h_j)$$

$$f(t_i) = h_i \prod_{j=1}^{i-1}(1 - h_j)$$

Finally, we have

$$L = \prod_{j=1}^{i} h_i^{d_i}(1 - h_j)^{n_i - d_i}$$

where n_i is the number at risk at time t_i. This is usually referred to as the number in the risk set.

Note

$$n_{i+1} + c_i + d_i = n_i$$

Kaplan-Meier Estimator

The Kaplan-Meier estimate is a mechanical process which allows for an easy estimation of the survival function from an investigation where censoring occurs(Cofie et al,2012). The Kaplan-Meier or product limit estimator is the limit of the life-table estimator when intervals are taken so small that only at most one distinct observation occurs within an interval. **Kaplan** and **Meier** demonstrated in a paper in JASA (1958) that this estimator is "maximum likelihood estimate". This estimator for $S(t)$ uses the mle estimators for h_i. Taking logs

$$l = \sum_i d_i \; log \; h_i + \sum_i (n_i - d_i) \; log(1 - h_i)$$

Differentiate with respect to h_i

$$\begin{aligned} \frac{\partial l}{\partial h_i} = \frac{d_i}{h_{h_i}} - \frac{n_i - d_i}{1 - h_i} = 0 \\ \Rightarrow \hat{h}_i = \frac{d_i}{n_i} \end{aligned}$$

So the Kaplan Meier estimator is

$$\hat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i} \right)$$

where

$$\begin{aligned} n_i &= \#\{ \; in \; risk \; set \; at \; t_i \} \\ d_i &= \#\{ \; events \; at \; t_i \} \end{aligned}$$

Note that $c_i = \#\{\text{censored in } [t_i, t_{i+1})\}$. If there are no censored observations before the first failure time then $n_0 = n_1 = \#\{\text{ in study}\}$. Generally we assume $t_0 = 0$

Nelson-Aalen Estimator

The next non-parametric estimator for $\hat{S}(t)$ to be talked about is the Nelson-Aalen estimator. The Nelson-Aalen estimator for the cumulative hazard function is

$$\hat{H}(t) = \sum_{t_i \leq t} \frac{d_i}{n_i} = \left(\sum_{t_i \leq t} \hat{h}_i \right)$$

This is natural for a discrete estimator, as we have simply summed the estimates of the hazards at each time, instead of integrating, to get the cummulative hazard.

This correspondingly gives an estimator of S, which is of the form

$$\begin{aligned}\tilde{S}(t) &= \exp\Big(-\hat{H}(t)\Big) \\ &= \exp\Bigg(-\sum_{t_i \le t} \frac{d_i}{n_i}\Bigg)\end{aligned}$$

The Greenwood's Formula

Variance Of Kaplan-Meier Estimator

Since Kaplan-Meier estimates are often used to compare the lifetime distributions of two or more populations ? for example, in comparing medical treatments ? their statistical properties are important. Greenwood's formulae for approximating the variance of the Kaplan-Meier estimates is given by:

$$Var\Big(\hat{S}(t)\Big) = \Big(\hat{S}(t)\Big)^2 \sum_{t_j \le t} \frac{d_j}{n_j(n_j - d_j)}$$

Variance Of Nelson Aalen Estimator

Corresponding to Greenwood's formula for the variance of the Kaplan-Meier estimator, there is a formula for the variance of the Nelson-Aalen estimator.The variance of the Nelson Aalen estimator for the cummulative hazard is given by:

$$Var\Big(\hat{H}(t)\Big) \approx \sum_{t_j \le t} \frac{d_j\Big(n_j - d_j\Big)}{n_j^3}$$

and

$$\begin{aligned}Var\Big(\hat{S}(t)\Big) &= Var\Big(\exp\big(-\hat{H}(t)\big)\Big) \\ &\approx \Big(e^{-H}\Big)^2 \sum_{t_j \le t} \frac{d_j\Big(n_j - d_j\Big)}{n_j^3} \\ &\approx \Big(\hat{S}(t)\Big)^2 \sum_{t_j \le t} \frac{d_j\Big(n_j - d_j\Big)}{n_j^3}\end{aligned}$$

An Illustration Of How to Use The Kaplan-Meier And Nelson-Aalen Estimators To Estimate The Survival Distributions

The population of elderly people in a prison is observed during the period 1 January 1994 to 31 December 1996. The duration of residence (measured to the nearest number of months) is recorded for those who die during the period, for those who are released from the prison during the period and for those who are still in residence on 31 December 1996. The recorded data measured in months are:

6* 6 6 6 7 9* 10*

10 11* 13 16 17* 20 23*

where * indicates those who were released from the prison during the period or who were still in residence on 31 December 1996(a censored data).
The estimate of the survival distribution using the Kaplan-Meier($\hat{S}_{KM}(t)$) and the Nelson-Aalen($\hat{S}_{NA}$) is shown in the table below.

j	t_j	n_j	c_j	d_j	$\frac{d_j}{n_j}$	$\hat{S}_{KM} = \prod_{j:t_j \le t}(1 - \frac{d_j}{n_j})$	$\hat{\Lambda}_{NA}(t) \sum_{j:t_j \le t}(\frac{d_j}{n_j})$	$\hat{S}_t = \exp[-\hat{\Lambda}_{NA}(t)]$
0	0	14	0	0	0	1.0000	0.0000	1.0000
1	6	14	1	3	$\frac{3}{14}$	0.7857	0.2143	0.8071
2	7	10	1	1	$\frac{1}{10}$	0.7071	0.3143	0.7303
3	10	8	2	1	$\frac{1}{8}$	0.6188	0.4393	0.6445
4	13	5	0	1	$\frac{1}{5}$	0.4950	0.6393	0.5277
5	16	4	1	1	$\frac{1}{4}$	0.3712	0.8893	0.4109
6	20	2	1	1	$\frac{1}{2}$	0.1856	1.3893	0.2492

Table 3.1: Using The Kaplan-Meier And Nelson-Aalen Estimators To Estimate Survival Distributions

The survival estimate of the Kaplan-Meier $\hat{S}_{KM}(t)$ is given by:

$$\hat{S}_{KM}(t) = \begin{cases} 1.0000 & 0 \leq t < 6 \\ 0.7857 & 6 \leq t < 7 \\ 0.7071 & 7 \leq t < 10 \\ 0.6188 & 10 \leq t < 13 \\ 0.4950 & 13 \leq t < 16 \\ 0.3712 & 16 \leq t < 20 \\ 0.1856 & 20 \leq t < 23 \end{cases}$$

The survival estimate using the Nelson-Aalen estimator() $\hat{\Lambda}_{NA}(t)$ is given by:

$$\hat{S}(t) = \begin{cases} 1.0000 & 0 \leq t < 6 \\ 0.8071 & 6 \leq t < 7 \\ 0.7303 & 7 \leq t < 10 \\ 0.6445 & 10 \leq t < 13 \\ 0.5277 & 13 \leq t < 16 \\ 0.4109 & 16 \leq t < 20 \\ 0.2492 & 20 \leq t < 23 \end{cases}$$

3.5.3 Parametric Estimators

A parametric approach is based on the assumption that the lifetime distribution belongs to a family of known parametric distributions. The most common forms are:

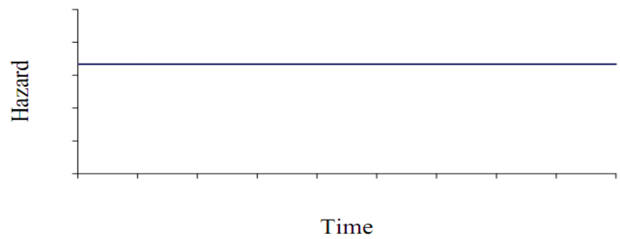
Exponential distribution

- used to model a constant hazard

- of a constant hazard rate, λ , means that the future lifetime of a life of age x , denoted by T_x, has an exponential distribution with parameter $\lambda \Rightarrow f_{T_x}(x) = \lambda e^{-\lambda x} (for x > 0)$

- this model could be used to reflect the hazard for an individual who remains in good health, where the level of hazard would reflect the risk of death from unnatural causes (e.g. accident) i.e. most useful at young ages

Figure 3.1: exponential-constant hazard

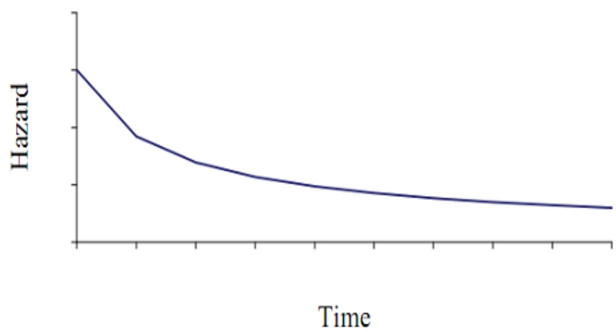


Weibull

- used to model a monotonically decreasing (or increasing) hazard

- this model could be used to reflect the hazard for patients recovering from major surgery (e.g. heart surgery), where the level of hazard is expected to fall as the duration since surgery increases i.e. if you survive the critical few hours after surgery then the hazard rate starts to fall

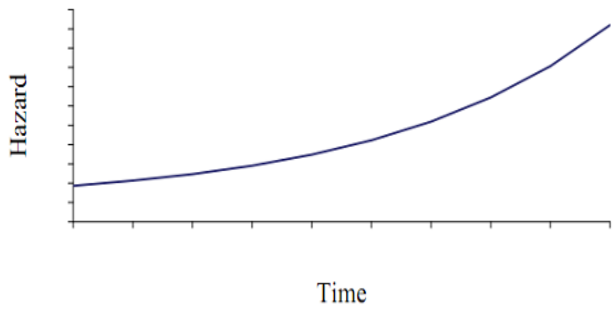
Figure 3.2: Weibull-monotonically decreasing hazard)



Gompertz-Makeham

- used to model an exponential hazard
- over longer time periods, this model can used to reflect human mortality, where we expect the level of the hazard to increase as age increases

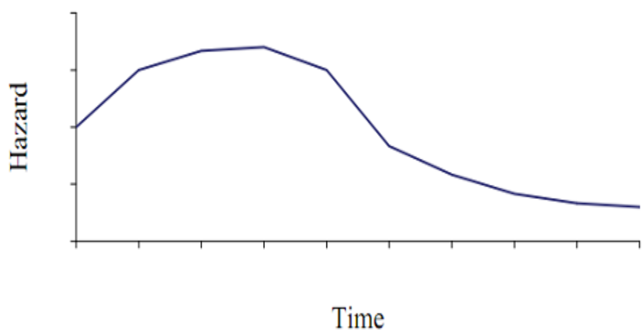
Figure 3.3: Gompertz-Makeham(exponential hazard)



Log-logistic

- used to model a humped hazard
- this model could be used to reflect the hazard for patients with a disease most likely to cause death in the early stages (e.g. tuberculosis), where the level of the hazard increases as the initial condition becomes more severe but then decreases once patients have survived the period of highest risk.

Figure 3.4: Log-logistic hazard("humped" hazard)



The table below illustrates the survival, hazard rate, density and expected life to failure of some parametric distributions:

Distribution	$S(t)$	$\lambda(t)$	$Density f(t)$	$E(T)$
Exponential	$e^{-\lambda t}$	$\lambda(>0)$	$\lambda e^{-\lambda t}$	$\frac{1}{\lambda}$
Weibull	$e^{-\lambda t^{\alpha}}$	$\alpha\lambda t^{\alpha-1}(\alpha,\lambda>0)$	$\alpha\lambda t^{\alpha-1}e^{-\lambda t^{\alpha}}$	$\frac{\Gamma(1+\frac{1}{\alpha})}{\lambda^{\frac{1}{\alpha}}}$
Gamma	1- $I(\lambda t,\beta)$	$\frac{f(t)}{S(t)}$	$\frac{\lambda^{\beta}t^{\beta-1}e^{-\lambda t}}{\Gamma\beta}$	$\frac{\beta}{\lambda}$

Table 3.2: survival, hazard rate, density and expected life to failure of some parametric distributions

3.6 Modeling Survival with Parametric Regression Models

Parametric models can be used with a single homogeneous population (as for the non-parametric approach). Alternatively, such models can be fitted to smaller homogeneous sub-groups, and confidence intervals for the fitted parameters will give an objective test of the differences in the lifetime distribution between the groups. However, fully parametric models are difficult to apply without knowledge of the precise form of the hazard function (which will often be the object of the investigation). As a result, a semi-parametric approach is more common. The two popular types of parametric regression models are the Accelerated Life Models(ALT) and the Proportional Hazard Models(PH).

3.6.1 Covariates

Again, a strictly non-parametric approach to survival analysis is limited when considering the effect of covariates on survival, where we define a covariate as any quantity recorded in respect of each life under observation, which is likely to affect the future lifetime distribution (e.g. age, sex, health). As stated, we could use these covariates to divide the total population into smaller homogeneous groups and then compare the Kaplan-Meier estimates for each sub-group. The problem is that as we introduce more covariates the size of the subgroups will reduce and hence less credible estimates will be obtained.

A solution is to use a more direct approach and construct a regression model (of which the most common form is the Cox regression model) in which the effects of the different covariates on survival are modeled directly.

Types of Covariates

Covariates are split into three main groups:

- a direct quantitative measure (e.g. age, height, weight)
- an indicator (e.g. 0 for male and 1 for female, or 0 for non-smoker and 1 for smoker)
- a quantitative interpretation of a qualitative measure (e.g. severity of symptoms ranging from 0 to 5, with 0 representing no symptoms and 5 representing extremely severe symptoms).

3.6.2 Accelerated Life Models(ALT)

Suppose there are (several) groups, labelled by index i. The accelerated life model has a survival curve for each group defined by

$$S_i(t) = S_0(\rho_i t)$$

where $S_0(t)$ is some baseline survival curve and ρ_i is a constant specific to group i

If we plot S_i against $logt, i = 1, 2, ..., k$, then we expect to see a horizontal shift as

$$S_i(t) = S_0\left(e^{(log\rho_i + logt)}\right)$$

3.6.3 Proportional Hazard Models

The Cox regression model is an example of a semi-parametric approach, which does not require the precise form of the hazard function in advance.

For each life in the observation, assume that we have p covariates of interest.

Thus the $(1 \times p)$ vector, $\underline{z}_i = (X_{i,1}, X_{i,2}, ..., X_{i,p})$, represents the values of the covariates in respect of life i at the start of the observation period.

According to the Cox model, the hazard function for life i at time t, denoted by $\lambda(t, \underline{z}_i)$, is given by:

$$\lambda(t, \underline{z}_i) = \lambda_0(t) \times exp(\underline{\beta}.\underline{z}_i^T) = \lambda_0(t) \times exp\left(\sum_{j=1}^{p} \beta_j \times X_{i,j} \right)$$

where $\underline{\beta} = (\beta_1, \beta_2, ..., \beta_p)$ is a $(1 \times p)$ vector of regression parameters and $\lambda_0(t)$ is the *baseline hazard* at time t .

The influence of each covariate in $\underline{z}_i = (X_{i,1}, X_{i,2}, ..., X_{i,p})$ on the hazard function for life i is included multiplicatively through the scalar quantity

$$\beta.\underline{z}_i^T = \sum_{j=1}^{p} \beta_j \times X_{i,j}.$$

In this simple formulation of the Cox regression model:

- the baseline hazard, $\lambda_0(t)$, is a function of time t , but is independent of the covariates $\underline{z}_i$
- the (so-called) constant of proportionality for life i , $exp(\underline{\beta}.\underline{z}_i^T)$, is a function of the covariates z_i, but is **independent of time** t

The *extended Cox regression model* allows us to analyse time-dependent covariates (i.e. covariates that are a function of time t).

Even though we are only using the simple Cox regression model it is still possible to analyse the effects of time-dependent covariates (e.g. age, height, weight) using the simple Cox model if:

- the value of these covariates does not change greatly over the period of the investigation
    - e.g. for a 6-month investigation, age can be considered as time-independent
- the main effect of these covariates depends on the value at a specific point in time
    - e.g. the weight on an individual at the onset of a disease

Note that, if the j^{th} covariate can only take positive values (e.g. age, height, weight), then:

- if the j^{th} regression parameter, β_j, is positive, there is a positive correlation between the covariate and the hazard rate.
- if the magnitude of the j^{th} regression parameter, β_j , is large, there is a strong correlation between the covariate and the hazard rate.

So if we have $\underline{\beta} = (0.7, -0.1, 0.02)$ then the first and third covariates have a positive correlation to the hazard rate i.e. when these values increase for a person their hazard rate also increases whereas the second covariate has a negative correlation. Also, the first covariate has a much stronger correlation than the other two covariates i.e. changes in the value of this covariate have more effect on the hazard rate.

Because the Cox model is constructed by changing the base hazard function we have a very simple formula when comparing the ratio of the hazard functions at time t of two different lives with covariate vectors $\underline{z}_1$ and $\underline{z}_2$:

$$\frac{\lambda(t, \underline{z}_1)}{\lambda(t, \underline{z}_2)} = \frac{exp(\underline{\beta}.\underline{z}_1^T)}{exp(\underline{\beta}.\underline{z}_2^T)}$$

This ratio does not depend on the time t, giving rise to the alternative name for the model, the *proportional hazards* model.

It should be noted that the Cox regression model is not the only model with proportional hazards. We can formulate an alternative model with a hazard function for life i at time t given by

$$\lambda(t, \underline{z}_1) = \lambda_0(t) \times g(\underline{z}_i)$$

where $g(\underline{z}_i)$ is any function of $\underline{z}$ only (and not of time t).

The benefits of using the Cox regression model is that it ensures that the hazard function is always positive for any value of t (as required in practice). Also, using the Cox model means that the logarithm of the hazard function is linear, which is useful both in theory and in practice.

The general shape of the hazard function for the population as a whole is determined by the baseline hazard function, $\lambda_0(t)$, while the exponential term, $exp(\underline{\beta}.\underline{z}_i^T)$, represents the differences between individuals.

The mortality of each individual is proportional to the baseline hazard function, $\lambda_0(t)$, and the constant of proportionality for each individual, $exp(\beta.\underline{z}_i^T) = \sum_{j=1}^{p} \beta_j \times X_{i,j}$, depends on the measurable values of the p covariates.

As a result, if we are mainly concerned with the effects of the different covariates on the lifetime function, we can ignore the baseline hazard and estimate the vector $\underline{\beta} = (\beta_1, \beta_2, ..., \beta_p)$ from the observed data.

This is known as a semi-parametric approach as the baseline hazard can be non-parametric whereas the vector is by definition parametric.

3.6.4 Estimating Regression Parameters

From the above, we can see the benefits of using a semi-parametric approach, so the next step is to look at how we can set about estimating the regression parameters $\underline{\beta} = (\beta_1, \beta_2, ..., \beta_p)$. This can be done by calculating the mortality rate for an individual life i at time t, $\lambda(t, \underline{z}_i)$, using the baseline hazard function (which we are assuming applies to all individuals), $\lambda_0(t)$, and the measurable values of the p covariates in respect of the individual life i, $\underline{z}_i$.

The partial likelihood

To estimate the vector $\underline{\beta} = (\beta_1, \beta_2, ..., \beta_p)$, we will maximise the following partial likelihood function.

Let $R(t_j)$ denote the set of lives alive and at risk just before the j^{th} observed lifetime at time t_j.
Assume that there is only one death at each observed lifetime t_j (i.e. $d_j = 1$ for $1 \leq j \leq k$).
Then, the probability that the j^{th} life observed to die at time t_j is one of the $R(t_j)$ lives at risk, conditional on only one death being observed at time t_j is:

$$\frac{\lambda(t, \underline{z}_1)}{\sum_{i \in R(t_j)} \lambda(t, \underline{z}_2)} = \frac{\lambda_0.exp(\underline{\beta}.\underline{z}_j^T)}{\sum_{i \in R(t_j)} \lambda_0.exp(\underline{\beta}.\underline{z}_j^T)} = \frac{exp(\underline{\beta}.\underline{z}_j^T)}{\sum_{i \in R(t_j)} exp(\underline{\beta}.\underline{z}_j^T)}$$

Thus, considering all observed lifetimes, t_j for $1 \leq j \leq k$, the total partial likelihood function is given by:

$$L(\beta) = \prod_{j=1}^{k} \frac{exp(\underline{\beta}.\underline{z}_j^T)}{\sum_{i \in R(t_j)} exp(\underline{\beta}.\underline{z}_j^T)}$$

The term *partial likelihood* is used because only those parts of the full likelihood involving the times at which deaths were observed are included and what was observed between the observed deaths is ignored. Unlike the Kaplan-Meier approach, the partial likelihood considers observed deaths only and not the times at which deaths were observed (or any censoring observed between the observed deaths). Thus, the above form of the partial likelihood gives the *comparative* risk of death at time t for a particular individual, given that a death occurs. Then, maximizing the total partial likelihood (based on all observed deaths) gives estimates of the regression parameters based on the order in which the observed deaths occurred. Hence, the model identifies the relative significance of the different factors that may be considered to affect mortality rates.

In effect what we are saying is that if we have 100 lives with 50 male and 50 female and the first death is a male then if we had no other information then we would assume that males die earlier but we would have little confidence in this result. However, if we observe all deaths and note that all 50 males died before a single female died then we would of course come to the conclusion that gender was a major factor in mortality (and we would in fact model the populations separately). What we are trying to do with the likelihood function is determine which covariates have a large effect on the hazard rate by observing which type of people die first compared to the number of people with those covariates in the population at the time of death. However, we are not including the actual timings in the calculation only the order of deaths (and we are assuming that only one death occurs at any particular time).

The likelihood $L(\underline{\beta})$ will usually have to be maximised numerically, normally using a statistical package.

In practice, the following problems may occur:

a) there may be more than one death at each observed lifetime, i.e. $d_j \neq 1$ for some values of j ; and

b) some observations may be censored at an observed lifetime.

It is usual to deal with (b) by including all of the lives censored at time t_j in the set $R(t_j)$.i.e. assuming that censoring occurs just after the deaths were observed (similar to the K-M method).

With (a), accurate calculation of the partial likelihood is difficult, as all possible combinations of d_j deaths from $R(t_j)$ lives at risk at time t_j should be included. In practice, the Breslow approximation is often used:

$$L(\underline{\beta}) \approx \prod_{j=1}^{k} \frac{exp(\underline{\beta}.\underline{s}_j^T)}{[\sum_{i\in R(t_j)} exp(\underline{\beta}.\underline{z}_i^T)]^{d_j}}$$

where $\underline{s}_j$ is the sum of the covariate vectors of each of the d_j lives observed to die at t_j.

The MLE of $\underline{\beta} = (\beta_1, \beta_2, ..., \beta_p)$, denoted by $\hat{\underline{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta_p})$, is found by solving $u(\hat{\underline{\beta}}) = 0,$ where:

$$u(\underline{\beta}) = \left(\frac{\partial logL(\underline{\beta})}{\partial \beta_1}, \frac{\partial logL(\underline{\beta})}{\partial \beta_2}, ..., \frac{\partial logL(\underline{\beta})}{\partial \beta_p} \right)$$

The vector of partial derivatives, $u(\beta)$, is known as the *efficient score function.*

It can be shown that $\hat{\underline{\beta}}$, the MLE of $\underline{\beta}$, is asymptotically (multivariate) Normal and unbiased. The asymptotic variance matrix, which allows the standard errors of the components of $\hat{\underline{\beta}}$ to be obtained,can be estimated by the inverse of the observed information matrix, denoted by $I(\underline{\beta})$. These standard errors are useful when evaluating the goodness-of-fit of a particular model.

The observed information matrix, $I(\hat{\underline{\beta}})$, is the $(p \times p)$ matrix of second partial derivatives:

$$I(\underline{\beta})_{i,j} = \frac{\partial^2 L(\underline{\beta}) log}{\partial \beta_i \partial \beta_j} \text{ for } 1 \leq i, j \leq p$$

evaluated at $\hat{\underline{\beta}}$.

Statistical Test Of The Covariates

In practice, there are likely to be several explanatory variables (or covariates) that may affect the hazard function. Part of the modelling process is the selection of those variables that have a significant effect. Therefore, criteria are needed for assessing the effects of covariates (either alone or in combination). In other words we don't want to look at hundreds of covariates as we will then be treating everyone individually and we will not be able to measure the main affects on mortality as we won't know which covariates are having most effect.

Two common criteria for selecting the most important covariates are the Wald Test and the likelihood ratio Test.

Hypothesis Tests:

For each covariate of interest, the null hypothesis is:

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

Wald Test

A Wald test of the above hypothesis is constructed as:

$$Z = \frac{\hat{\beta}_j}{se(\beta_j)} \; or \; \chi^2 = \frac{(\hat{\beta}_j}{se(\beta_j))^2}$$

This test for $\beta_j = 0$ assumes that all other terms in the model are held fixed.

If we have a factor A with a levels, then we would need to construct a χ^2 test with $(a-1)$ df, using a test statistic based on a quadratic form:

$$\chi^2_{(a-1)} = \hat{\beta}'_A Var\left(\hat{\beta}_A\right)^{-1} \hat{\beta}_A$$

where $\beta_A = (\beta_2, ..., \beta_a)'$are the $(a-1)$ coefficients corresponding to $Z_2, ..., Z_a$(or $Z_1, ..., Z_{a-1}$, depending on the reference group).

Likelihood Ratio Test

Suppose that we fit a model using p covariates.

Let L_p be the (maximised) log-likelihood in this case.

Suppose that we wish to consider the effect of adding an additional q covariates to the model then we will fit a model using $(p+q)$ covariates and get L_{p+q} as the log-likelihood in this case.

Then, under the null hypothesis that the additional q covariates have no effect in the presence of the original p covariates, the likelihood ratio statistic, given by: $-2 \times (L_p - L_{p+q})$ has an asymptotic χ^2 distribution with q degrees of freedom

3.7 Residual Analysis

In order to asses the goodness of fit of the Cox PH model to a given data, residual analysis must be performed on the fitted model. Various residual analysis are used in assessing the goodness of fit of the Cox PH model, among which are: Schoenfeld, Cox-Snell and Martingale residual analysis.In our study, the Cox-snell residual analysis will be employed to validate the final model.

3.7.1 Cox-Snell Residual

To examine the overall fit of the Cox PH model, the Cox-Snell residual analysis will be employed.

If X has the hazard function $\lambda\big(x|Z\big)$, then the cumulative hazard $\Lambda\big(x|Z\big)$ satisfies exponential(1). The Cox-Snell residual is defined(Cox and Snell,1986) as:

$$r_j = \hat{\Lambda}_0(T_j)\exp(\hat{\beta}' Z_j), j = 1, ...n$$

where, $r_j's$ are censored sample from a unit exponential distribution, given the assumed Cox model holds and $\hat{\beta}, \hat{\Lambda}_0$ close to the true values β, Λ_0.

Plot of $\hat{\Lambda}_r(r_j)$ versus r_j will produce a straight line with slope 1 if the Cox model provides a good fit of the data as in the figure below.

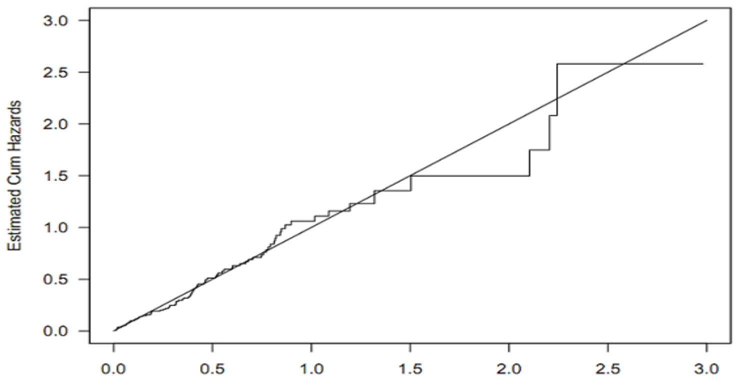


Figure 3.5: Cox-Snell Residual Analysis

Chapter 4

ANALYSIS

4.1 Introduction

In Chapter, we shall use survival analysis to model cases of surrender of life policies in SIC Life. The Kaplan-Meier estimator would be used to estimate surrender probabilities of life insurance policyholders. Again the Cox Proportional Hazard would be used to fit the data and validation be run using the Cox-Snell residual analysis.

4.2 Data Collection and Analysis

Data on surrender of life insurance policies was taken from SIC life from January, 1999 to December,2016. An excerpt of the data is given below:

4.2.1 Kaplan-Meier Estimates

```
> Kmsurvival<- survfit(Surv(time,event)~1,conf.int=.95)

> summary(Kmsurvival)
```

```
Call: survfit(formula = Surv(time, event) ~ 1, conf.int = 0.95)
```

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	73642	961	0.98695	0.000418	0.98613	0.98777
2	72419	3458	0.93982	0.000878	0.93811	0.94155
3	68269	15090	0.73209	0.001642	0.72888	0.73531
4	52107	10457	0.58517	0.001836	0.58158	0.58878
5	40497	7575	0.47571	0.001875	0.47205	0.47940
6	30383	6826	0.36884	0.001847	0.36524	0.37247
7	21171	4594	0.28880	0.001784	0.28533	0.29232
8	15767	3437	0.22585	0.001688	0.22256	0.22918
9	11629	2576	0.17582	0.001576	0.17276	0.17893
10	7774	2007	0.13043	0.001459	0.12760	0.13332
11	5490	1466	0.09560	0.001323	0.09304	0.09823
12	3868	1029	0.07017	0.001185	0.06788	0.07253
13	2694	811	0.04904	0.001035	0.04706	0.05111
14	1694	558	0.03289	0.000892	0.03119	0.03468
15	1052	396	0.02051	0.000742	0.01910	0.02202
16	557	190	0.01351	0.000639	0.01232	0.01483
17	298	103	0.00884	0.000560	0.00781	0.01001
18	145	18	0.00774	0.000547	0.00674	0.00889

The Kaplan-Meier estimates from the output above indicate that, the surrender probabilities increased substantially between times 1 and 4, with most event occurring at times 3 and 4 indicating that, most polices are surrendered within the first four years from the policy inception. These probabilities then decreased over time as the number of events decreased as the duration of the policy increased; showing that surrender of policies reduces relatively as the policy duration increases. This can be seen from the graph of the Kaplan-Meier plot below.

```
> plot(Kmsurvival,main="The Kaplan-Meier Estimate Of The Survival
```

```
Probabilities",xlab="Time",ylab="Survival probability")
```

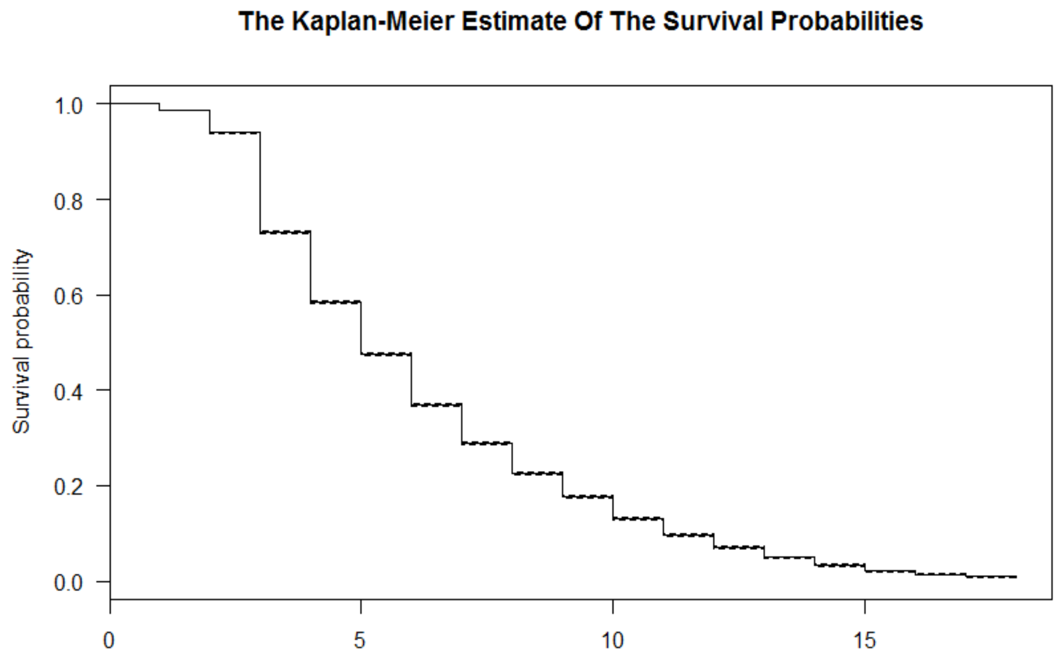


Figure 4.1: Kaplan-Meier Plot

4.2.2 The Cox Proportional Hazard Model

The Cox Proportional Hazard would be used to investigate the factors that are significant to the model. The factors considered in the model include: Age of the policyholder at policy inception, Gender of the policyholder, Income level of the policyholder at policy inception, Customer's dissatisfaction with services provided,Subscription unto of better policies,Interest rate arbitrage and customer's need for emergency fund, which could affect customer's prospect of surrendering his/her policies.

The Age factor was treated as a quantitative variable;

The Gender factor as a categorical covariate had two levels, male and female and were coded 0 and 1 respectively;

Income was treated as a quantitative variable;

Customers dissatisfaction with services provided was treated as a categorical variable with three levels as "no", "yes" and were coded 0 and 1 respectively;

Subscription unto of better policies was treated as a categorical variable with three levels as "no", "yes" and were coded 0 and 1 respectively;
Interest rate arbitrage was treated as a categorical variable with three levels as "no", "yes" and were coded 0 and 1 respectively; and
customer's need for emergency fund was treated as a categorical variable with three levels as "no", "yes" and were coded 0 and 1 respectively;
Thus, the mathematical model for the above factors is:
$\lambda(t|z) = \lambda_0 \exp(\beta_1 age + \beta_2 gender1 + \beta_3 Income + \beta_4 dissatisfaction1 + \beta_5 newpolicy1 + \beta_6 interestrate1 + \beta_7 emergencyfund1)$

the output of the above model is given below:

```
> CoxphModel<- coxph(Surv(time,event)~age + gender + Income +
dissatisfied.customer + new.policy + interest.arbitrage + emergency.fund,
method="breslow")
> summary(CoxphModel)
Call:
coxph(formula = Surv(time, event) ~ age + gender + Income+
 dissatisfied.customer+ new.policy + interest.arbitrage + emergency.fund,
 method = "breslow")

  n= 61552, number of events= 61552
   (12090 observations deleted due to missingness)

                            coef  exp(coef)   se(coef)       z Pr(>|z|)
age                   -1.208e-02  9.880e-01  4.864e-04 -24.834   <2e-16 ***
gender1                9.089e-02  1.095e+00  8.467e-03  10.735   <2e-16 ***
Income                -7.443e-07  1.000e+00  1.471e-06  -0.506   0.6129
dissatisfied.customer1  7.537e-02  1.078e+00  8.486e-03   8.882   <2e-16 ***
```

```
new.policy1            2.655e-02  1.027e+00  4.052e-02   0.655   0.5124
interest.arbitrage1    1.757e-02  1.018e+00  8.307e-03   2.115   0.0344 *
emergency.fund1        1.037e-01  1.109e+00  9.004e-03  11.514   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

                       exp(coef) exp(-coef) lower .95 upper .95
age                        0.988     1.0122    0.9871    0.9889
gender1                    1.095     0.9131    1.0771    1.1135
Income                     1.000     1.0000    1.0000    1.0000
dissatisfied.customer1     1.078     0.9274    1.0605    1.0964
new.policy1                1.027     0.9738    0.9485    1.1118
interest.arbitrage1        1.018     0.9826    1.0013    1.0344
emergency.fund1            1.109     0.9015    1.0898    1.1290

Concordance= 0.556  (se = 0.002 )
Rsquare= 0.014   (max possible= 1 )
Likelihood ratio test= 851.3  on 7 df,   p=0
Wald test            = 842.7  on 7 df,   p=0
Score (logrank) test = 843.5  on 7 df,   p=0
```

From the above it is observed that, a test of the significance of the aforementioned factors in the above model indicates that, the income and the subscription onto new policy factors had higher p-values of 0.6219 and 0.5124 respectively as compared to the significance level of 0.05, indicating that the two are insignificant factors in the model and hence we eliminate them and consider the remaining factors in our final model.

Final Model

The final model now becomes:

$\lambda(t|z) = \lambda_0 \exp(\beta_1 age + \beta_2 gender1 + \beta_3 dissatisfaction1 + \beta_5 interestrate1 + \beta_6 emergencyfund1)$

```
> final.model<-coxph(Surv(time,event)~age+gender+dissatisfied.customer+interest
> final.model
Call:
coxph(formula = Surv(time, event) ~ age + gender + dissatisfied.customer +
    interest.arbitrage + emergency.fund, method = "breslow")

                           coef exp(coef)  se(coef)      z      p
age                   -0.012084  0.987989  0.000486 -24.84 <2e-16
gender1                0.090841  1.095095  0.008466  10.73 <2e-16
dissatisfied.customer1 0.074999  1.077883  0.008468   8.86 <2e-16
interest.arbitrage1    0.017154  1.017302  0.008283   2.07  0.038
emergency.fund1        0.102740  1.108203  0.008884  11.57 <2e-16

Likelihood ratio test=851  on 5 df, p=0
n= 61552, number of events= 61552
   (12090 observations deleted due to missingness)
```

putting in the estimates of the β^s into the final model,

$\lambda(t|z) = \lambda_0 \exp(-0.012084 age + 0.090841 gender1 + 0.074999 dissatisfaction1 + 0.017154 interestrate1 + 0.102740 emergencyfund1)$

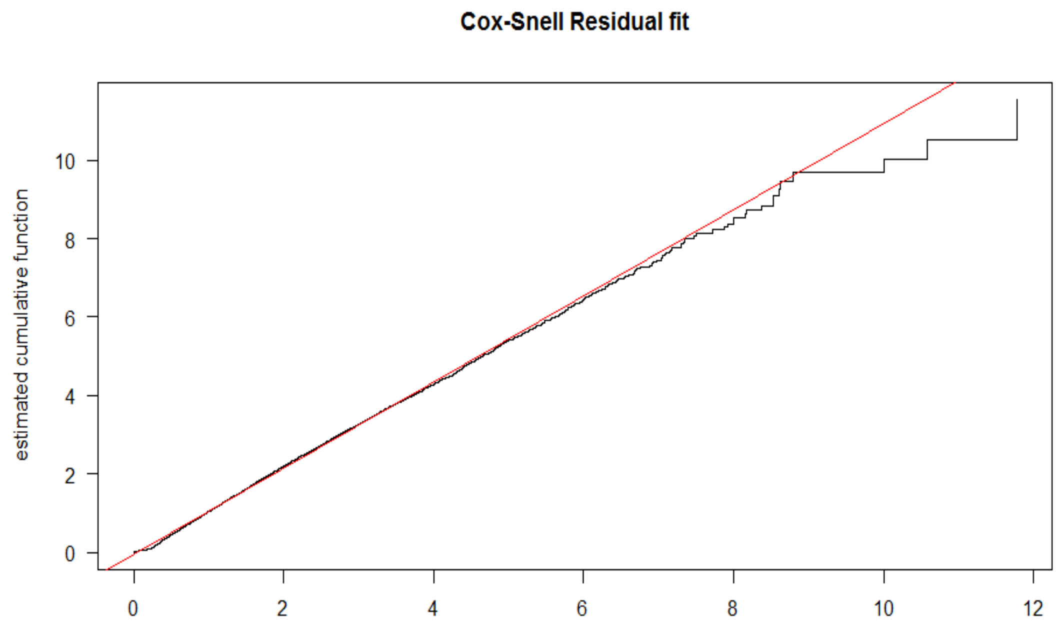


Figure 4.2: Cox-Snell Residual Plot

4.2.3 Residual Analysis

Using the Cox-Snell residual analysis, an investigation would be made to see if there is a straight line in the graph plot of the estimated cumulative function $(\cap\Lambda_r(r_j))$ against r_j.

```
> coxsnell<-event-resid(final.model,type="martingale")
> modval=survfit(Surv(coxsnell,event)~1)
> par(las=1,mfrow=c(1,1),mai=c(0.5,1.0,1,0.1),omi=c(1.0,0,0.5,0))
> fitres=survfit(coxph(Surv(coxsnell,event)~1,method="breslow"),type="aalen")
> plot(fitres$time,-log(fitres$surv),type="s",xlab="residuals",ylab="estimated
> modelline<-lm(-log(fitres$surv)~fitres$time)
> abline(modelline,col="red",lty=1)
```

From figure 4.2, it is seen that the plot produces a straight line, which indicates that the proportionality assumption is met and thus, the Cox-PH model provides a good fit to the data.

Chapter 5

Conclusion

5.1 Introduction

This chapter contains a brief summary of the findings obtained from the study in connection to the set objectives. Recommendations would also be made for the various interested life insurers and for future studies.

5.2 Summary Of Results

From the data obtained from SIC Life, it is observed from the Kaplan-Meier estimates that, persons who subscribed to life policies had higher surrender probabilities within their first 3 and 4 years of subscription, as the highest number of surrender events were recorded within these periods with probabilities of 0.20773 and 0.14692 respectively. The surrender probabilities then decreased marginally within periods 5 and 6 from 0.10946 to 0.10687. From period 6 to 7, there was a sharp decline in the probability to 0.08004. It then decreases steadily through to period 18.

The Cox-PH Model was used to determine among the factors considered, the ones which were significant to surrendering. It was observed that, among the factors: *age,gender,Income level, Customer's dissatisfaction,Subscription unto new policies,Interest rate arbitrage* and *customer's need for emergency fund*, only the income and subscription onto new policy factors were insignificant. Also the considering the coefficients of the covariates used in our final model, it is can be said that, the *gender, customer's dissatisfaction,interest rate arbitrage* and *customer's need for emergency fund* factors have a direct relationship with the surrender hazard rate.

The coefficient of the *gender* factor being 0.090841 implies that, as more males subscribe onto the policies, the surrender hazard rate tends to increase as compared to females.

The *customer's dissatisfaction* factor has a coefficient of 0.074999 which implies that as more policyholders become dissatisfied with the services provided by the insurer, the hazard rate for surrender also increases.

Also the interest arbitrage factor has a coefficient of 0.017154 implies that as the interest rate of other alternative investment become higher than insurance interest, more policyholders will surrender the policies for such investment avenues.

Moreover, the coefficient value of 0.102740 for the *customer's need for emergency fund* factor implies that, as more policyholders tend to need money to address some emergencies, the rate of surrendering increases.

The *age* factor has a coefficient value of -0.012084 indicating an inverse relationship with the hazard rate of surrendering. This implies that as the age variable increases in magnitude, the surrender hazard rate reduces, thus that young policyholders are more likely to surrender their policies as compared to the older policyholders.

5.3 Conclusions

- From the analysis, it is observed that the prospect of surrendering of life policies increases during first 3 years from the policy inception date. The highest probability of surrendering of policies occurs at time 3 (years) from the Kaplan-Meier estimates and has a probability value of 0.20773 being approximately 21%.

- From the Cox-PH model, it is found that with the factors considered, it can be inferred that surrender of life policies does not depend on the income level of the the policyholder and subscription onto new policies but on age, gender,policyholder's dissatisfaction with services provided by insurer, need for emergency fund, interest rate arbitrage and other factors that might have not been considered.

5.4 Recommendation

- It is recommended that, strategies should be developed towards enticing more females to subscribe onto the life policies.
- Insurers are recommended to design life policies that will suit the youth to increase/sustain their retention rate of their polices in the life companies.
- Management should implement measure that would help cater for client needs in respect of areas including response to feed back, grievances, among others.
- Insurers should implement policies that would allow its policyholders who have had their policy in force for some minimum number of years to qualify to apply for policy loans in times of emergency.

REFERENCES

Afriyie, D. (2006). Ghana insurance market: Past, present and the future. *University of Ghana, Accra*, pages 6–11.

Allison, P. D. (2010a). Definition of the event. *The Reviewer's Guide to Quantitative Methods in the Social Sciences*, page 413.

Allison, P. D. (2010b). *Survival analysis using SAS: a practical guide*. Sas Institute.

Altman, D. G. (1990). *Practical statistics for medical research*. CRC press.

Brooker, D. E., Stringham, D., Smith, C. A., Sharf, C. S., Petro, K. M., Mariano, R. A., Baruh, R. J., York, C. A., et al. (2012). System and method for managing a group insurance policy. from `https://www.google.com/patents/US8175952`. US Patent 8,175,952.

Buckham, D., Wahl, J., and Rose, S. (2010). *Executive's guide to solvency II*. John Wiley & Sons.

Cox, S. H. and Lin, Y. (2006). Annuity lapse rate modeling: Tobit or not tobit. *Society of Actuaries*, pages 1–12.

Dar, A. and Dodds, C. (1989). Interest rates, the emergency fund hypothesis and saving through endowment policies: some empirical evidence for the uk. *Journal of Risk and Insurance*, 56(3):415–433.

Denenberg, H. S. (1963). The legal definition of insurance: Insurance principles in practice. *The Journal of Insurance*, 30(3):319–343.

Dorfman, M. (2008). *Introduction to Risk Management and Insurance*. Pearson/Prentice Hall.

Eling, M. and Kiesenbauer, D. (2012). Does surplus participation reflect market discipline? an analysis of the german life insurance market. *Journal of Financial Services Research*, 42(3):159–185.

Eling, M. and Kiesenbauer, D. (2014). What policy features determine life insurance lapse? an analysis of the german market. *Journal of Risk and Insurance*, 81(2):241–269.

Group, O. (2014). *The Report: Ghana 2014:*. 4. Oxford Business Group.

Hoyt, R. E. (1994). Modeling of insurance cash flows for universal life policies. *Journal of Actuarial practice*, 2:197–220.

Jenkins, S. P. (2005). Survival analysis. *Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK.*

Kiesenbauer, D. (2012). Main determinants of lapse in the german life insurance industry. *North American Actuarial Journal*, 16(1):52–73.

Kim, C. (2005a). Modeling surrender and lapse rates with economic variables. *North American Actuarial Journal*, 9(4):56–70.

Kim, C. (2005b). Report to the policyholder behavior in the tail subgroups project. *Society of Actuaries, Schaumburg, IL.*

KPMG, G. (2014). Review of the ghanaian insurance industry. from `http://docplayer.net/storage/25/5888443/1491310789/XUnTCK_Y9HA0BsWU6Jzl8A/5888443.pdf`. [Online; accessed ?January ?20, ?2017].

Kuo, W., Tsai, C., and Chen, W.-K. (2003). An empirical study on the lapse rate: the cointegration approach. *Journal of Risk and Insurance*, 70(3):489–508.

Mac-Issaka, B. (2015). Valuation of surrender option using crank-nicolson and hopscotch methods. Master's thesis, KNUST.

Müller, W. (1981). Theoretical concepts of insurance production. *Geneva Papers on Risk and Insurance*, 6(21):63–83.

NIC (2017). About us. from http://nicgh.org/about-us/overview/. [Online; accessed 31-March-2017].

Outreville, J. F. (1990). Whole-life insurance lapse rates and the emergency fund hypothesis. *Insurance: Mathematics and Economics*, 9(4):249–255.

Rodrıguez, G. (2005). Non-parametric estimation in survival models.

Russell, D. T., Fier, S. G., Carson, J. M., and Dumm, R. E. (2013). An empirical analysis of life insurance policy surrender activity. *Journal of Insurance Issues*, 36(1):35–57.

Sandra-Odonnell (2015). Censoring. from http://documents.mx/download/link/censoringpdf. [Online; accessed 17-November-2016].

Wikipedia (2017). Vehicle insurance — wikipedia, the free encyclopedia. from https://en.wikipedia.org/w/index.php?title=Vehicle_insurance&oldid=761682012. [Online; accessed 26-January-2017].

Zogbenu, M. (2016). High surrender rates : Credit houses to blame? from http://www.graphic.com.gh/business/business-news/high-surrender-rates-credit-houses-to-blame.html. [Online; accessed 17-December-2016].