Modeling Lapse Rates using Economic Variables.

A Case Study for a Life Insurance Company operating in Kenya

Project done in partial fulfillment of an MSc Actuarial Science Degree at the School of Mathematics

UNIVERSITY OF NAIROBI

John Ocheche

Registration number: I56/70679/2007

Supervisor: Dr PGO Weke

Statement of originality

By signing below I confirm that this is my original work and that in carrying out the work I have not engaged in any manner of plagiarism. I also confirm that the work has not been presented at any other University of learning institution.

John Ocheche

I56/70679/2007

Sign : ______________________

Date : 25/8/09

Supervisor

Dr PGO Weke

Sign : ______________________

Date : 26/08/2009

Executive Summary

Lapse rates are an important consideration in the pricing and implementation of Life Insurance Policies. *Lapses* are defined the premature withdrawal of policyholders from policies that they have previously taken up. The occurrence of a lapse is typified by the cessation of premium payments by some policyholder and in some cases a benefit payment may be made. It is therefore important to estimate lapse probabilities in the pricing of products since these have a direct impact on the duration of premium payments. Additionally it is important for the policy provider to have some idea the sort of lapse rates to expect where policies have been taken up in order to carry out reasonable Asset Liability Management. This project reviews past research that has been carried out on the estimation of future lapse rates and the factors that affect them. An attempt is also made to estimate lapses based on economic variables for a Life Insurance Company operating in Kenya. Amongst the key factors in the estimation of lapses is the type of insurance policy!

The model whose construction is attempted draws largely from Changki Kim (2005) 'Modeling lapses using economic variables' where the monthly lapse rate is constructed as a response variable in a generalized linear model. The predictor variables used are Market rates, Unemployment rates, Economic growth rates, and Financial Crises.

Acknowledgements

There are several people without whom this project may not have been possible.

The first is Dr Weke who gave me an ear throughout the project and never failed to provide necessary guidance. Then there is Dr Achia. I am grateful for the holes that he not only poked in my project but also helped me to fix.

My classmates; George, Caro, Mumo, and Golda. This has been an arduous journey for all of us and I acknowledge their support.

I also acknowledge the support of my sponsors Baba and Patrick; Abed, my boss, for allowing me to miss work on several occasions to complete work on the project.

I know that without Mama's prayers all this would not have been possible so a special acknowledgement goes to her, and to my sisters for repeatedly nagging me with the question 'have you finished your project?', and more recently 'have you presented yet?'

Last and most importantly I thank the Lord for providing me with all the support and strength to be able to complete this project successfully.

CONTENTS

1. Introduction

1.1. Background

Life Insurance is a contractual arrangement between the insured, referred to as policyholder, and the insurance provider. The insurance provider promises to pay the insured's dependants a certain sum of money, called the sum assured, on the occasion of his [insured] death. In exchange the insured is obligated to pay sums of money, usually at regular intervals to the insurance provider throughout or during some term in his lifetime. The premiums paid are rather small in amount when compared to the sum assured. This is especially so in the case where the death benefit is payable on condition that the death occurs within a pre-specified term. The term might come to a conclusion death not having occurred in which case the death benefit would not be paid! The premiums are calculated by equating their mean present value to the mean present value of the benefits. Mean present values are used because it is not known for certain when the premium payments will stop. The premium payments may stop because of the policyholders' death or due to his lapse.

Lapse is therefore the premature withdrawal from the policy by the insured party and as such is an important consideration in the pricing process of a life insurance policy. Death needs no definition. Mean present value calculations for both the benefits and premiums require that the insured's expected length of future lifetime be known. This in necessary in order to establish the duration that the premiums will be paid and roughly the point at which a death benefit may be paid. Several mortality tables are used in practice to determine this future lifetime. However, the insured may also no longer pay premiums if they lapse! Lapses are therefore just as important as the mortality rates in the process of calculating the mean present values of benefits and subsequently the required premium level for a given sum assured.

The process of policy development also requires that cash-flow projections are modeled. Reasonable profit margins and critical masses are investigated using these models and indeed the opinion that a new policy should be sold to the market is arrived at through this process. Lapse rates form an integral part of this investigation.

Premiums received by a Life Insurance provider are invested in order that they receive a return before they are paid as benefits. It is important to invest the premiums in assets whose duration is more or less similar to that of the liabilities – a process in this case referred to as Asset Liability Management (ALM). That way cash is available to pay liabilities when they become due. Some life policies include surrender values that are paid on the insured's withdrawal. Appropriate lapse rates should therefore contribute to the ALM process in giving some idea when some liabilities will become payable. Where it is expected that lapses may be quite high in early years, investment of the premiums should be in assets whose nature is short term.

The problem with lapses especially when compared with mortality is that they depend on range of factors that are not easy to forecast. (Scott 1999) states that unlike mortality rates which are largely predictable by actuaries, lapse rates depend on a range of economic and commercial factors that are less easy to forecast accurately. The Kenyan life insurance industry, which is the field where we intend to carry out this project, currently uses no scientific method to arrive lapse rates. At best it is the experience of similar products that is used in determining lapse rates. This is arguably a suitable method. New products of a similar kind to previous ones should roughly show similar lapse rates. However studies carried out elsewhere have developed models that depend on several factors. These 'factors' serve as predictor variables in a model where the lapse rate is the dependent variable. This project therefore aims to develop a lapse rate model that is suitable for the Kenyan life insurance market where the lapse rates are arrived at based on such factors.

The discussion has so far focused on life insurance policies. In practice most life insurance providers (companies) sell other types of products as well. These may include but not be limited to:

- Education plans; these are savings plans where a benefit, that includes the savings and interest earned, is payable at the end of a pre-specified term to cover the school fees of the insured's dependents.

- Endowment plans; these are policies that pay a benefit in the event of the policyholders death should it occur within a pre-specified term or at the end of that term if the policyholder is then alive.

- Annuity; this is a regular payment made by the insurance provider to a policyholder for either a pre-specified term or throughout his future lifetime. The policyholder pays a premium – normally one lump sum amount – at the beginning of the term in order to get this benefit.

Previous studies have shown that lapse rates vary for the different types of policies shown above. This project also seeks to develop models that are appropriate for the different policy types.

1.2. Statement of the problem

There currently exists no proper lapse model in the Kenyan Life Insurance market that has been developed based on factors that contribute to lapse. The models used are based on a simple analysis of previous experience. Lapses in the first year for a new life insurance product are regarded as equal to the lapse rate over the first year for a previous life insurance product. The analysis does not view the lapse rate as a dependent variable whose determination is based on several contributing factors, but rather as a rate dependent only on the type of product in question. In other words the underlying factors that may contribute to lapse are not viewed individually, neither are they varied or even tested for significance. Herein lies the problem of the current models used; that lapse rates despite being an important component in the pricing and profitability of life insurance policies are not modeled as best they can.

Previous studies have been carried out abroad to develop lapse rates – several of these are discussed in the Literature Review section. In these cases the lapse rates are viewed as variables dependent on a number of predictor variables. It is thought that the predictor variables used in some of these cases may be different from those that affect the Kenyan market – important variables may have been left out by the models and those that are unimportant included! Importing the models directly for use in Kenya might therefore present another problem. This study also aims to find out whether that should be a genuine concern.

1.3. Objectives

To review lapse rate models developed for use in life insurance by previous scholars and practitioners.

To develop lapse rate models for the following policy types that are suitable for the Kenyan Life Insurance market by modifying previously developed models to include factors that are relevant to Kenya:

- Protection
- Education
- Annuity
- Endowment

To check the suitability of the developed models on a set of data obtained from a local insurer.

Related to the above establish lapse rates that are calculated based on relevant predictor variables as opposed to the current situation where lapse rates for a new policy during a given period are arrived at by simply taking the proportion of individuals who left during that period for an existing similar policy.

1.4. Significance of the study

The required conclusion of this project will present proper lapse rate models for different types of life insurance policies in Kenya. The conclusion should provide a significant contribution to the profitability of life insurance companies in Kenya by allowing them to price their products (set premiums) based on more accurate lapse rates and to carry out the ALM process better by matching more closely the term of the assets to that of liabilities. Policyholders', on the other hand, should expect to buy more reasonably priced products.

2. Literature Review

This section comprises notes on some papers that have been written by scholars on lapse rates.

2.1. *Lapse Rates (Richardson and Hartwell 1951)*

"It is definitely apparent that the agent writing the business is a major factor in persistence, possibly even more powerful than the economic period.

It has been suggested, for example, that lapsation arises from overselling, failure to meet needs, making a 'poor' sale, failing to follow up, failing to create a prestige relationship, inadequate training or poor selection of the agent, or failure to create confidence in the company or agent."

Anonymous

Richardson and Hartwell began their report by presenting a brief history on the study of lapse rates;

In the earliest discussions lapses were considered to be dependent on the selling agents 'selection of policyholders; and other prevailing economic conditions. Indeed the dependence on economic conditions gained dominating significance in the depression after 1929!

About 1925 an attempt to discover measurable factors which are associated with lapse ensued; by the late thirties a large number of studies had reported the effect on lapse rates of income, occupation, sex, age, previous insurance, premium frequency, plan, and several other variables. There was also historical evidence that economic disaster affected most violently the termination at durations subsequent to the second year. The problem, however was that all analyses carried out failed to recognise the interaction of various factors, a deficiency that was remedied by a study published in 1949 by the Agency Management Association (an organization of life insurance selling agents).

The authors' main concern was, however, with the causes of high lapse rates and the characteristics of a business which had either high or low persistence.

Their paper presented the results of a large number of studies that were made by the Mutual Life Office of New York in order to determine the effect on the lapse rates in the early policy years, of various characteristics of the insured, policy, agent and sale, and certain economic factors. The studies were restricted to lapse rates in the early policy years since those appeared to be more affected by the different characteristics than those at later durations. Moreover from the standpoint of both the policyholder and the company, a lapse at an early duration is generally considered to be much more serious than a termination at a later duration. Most policies terminated at later durations have normally fulfilled a real economic need!

The various characteristics were segregated as follows for study and the following observations made on each one.

Characteristics of the policyholder	
Income	Considered the most important factors affecting lapse rate.
Occupation	Had a substantial effect on persistence. Lowest rate were amongst students and the highest salesclerks.
Age	Lapse rate decreased with increases in age. However there was a close correlation between age an income which 'substantially accounted for the superficial evidence that age and persistence were related'.

Characteristics of the policy	
Frequency of premium payment	Lapse rates were higher for monthly that for quarterly premium policies.
Plan of insurance	Pronounced persistence differences existed by plan. Predominantly endowment policies had good persistence and the limited payment life policies above average rate; only life at age 85 fell below the average rate.
Amount of policy	'When amount of policy was related to *income* groups, it became clear that in no case was the amount of policy a criterion of the lapse rate.'
Amount of premium	There was evidence that persistence increases as the amount of premium increases. However when the premium was analysed against income groups the trend was substantially lessened.

Other factors	
By gender	Persistence was significantly better for women than for men. The over-all lapse rate for women was 15.8% compared with 18.6% for men.
By dependency or marital status	Little differences existed between various marital status classifications except for dependent single women who had a very low lapse rate. This group was probably in the category of students and it has already been noted that persistence was highest among students.
By type of sale – 'needs based' or 'package sale'	The needs sale policies on average required higher premium and due to this exhibited higher lapse rates than package sales.
Policies with and without policy loans	The lapse rates on policies with loans varied from two and a half to nearly ten times that on policies without loans, as might be expected, since in very many cases a policy loan is a first step toward surrender.
Lapse rates by geographical area	More mature sections of the country had lower lapse rates than those that were expanding fast economically. 'Rapid economic expansion comes frequently from an optimistic attitude, and where there is optimism there must be some excessive optimism; in that situation people are easier to sell to and tend to buy things they cannot really afford.'

The material presented in the papers informed the reader on the facts about characteristics with high or low lapse rates. However it did tell how to tackle the problem of reducing lapse rates. The most promising fields for future investigation, it concluded, lie in interviews surveys, in direct contact with the policyholder, and in research directed at discovering factors underlying the wide differences between individual agents.

2.2. *Expected Lapse Tables (Brzezinski 1975)*

The study constructed a set of Select and Ultimate lapse tables. Twelve tables were constructed in total; Select for the first 15 policy years by number and amount of pension insurance, high early cash-value-insurance, permanent and term insurance; Ultimate for sixteenth and later policy years each by number and amount of pension insurance, and all other insurance excluding pension insurance.

Mention was also made the lapse tables most widely used in the Insurance Industry at the time. These were the previously published Linton and Moorhead tables. The Linton tables, long accepted as standards, were published by M.A. Linton in 1924 in the *Record of the American Institute of Acuaries (RAIA,* XXIII, 283*)*. The Moorhead tables were published by E.J. Moorhead in 1960 in the *Transactions of the Society of Actuaries (TSA,* XII, 545*)*. These tables, or variations of them at any rate, based on Company Experience (or Company Experience Tables) had been commonly used by actuaries for asset share calculations, model-office projections, gross premium calculations, agent's compensation calculations, and various other purposes.

The tables constructed in Brzezinski (1975) were intended to replace the use of Linton and Moorhead lapse tables in the future analysis of lapse trends in the inter-company experience of the Life Insurance Marketing and Research Association (LIMRA) long term lapse study, and to provide companies with up-to-date basis for lapse comparisons of various kinds. Linton and Moorhead had been used to calculate expected lapses in LIMRA's first long-term lapse study but it had been found that neither the Linton nor the Moorhead were appropriate since they did not represent the substantial variation in lapse rates associated with an insured's age at issue, characteristics of the insurance being purchased, and the insured's attained age. As a result it was decided to attempt the construction of a new set of expected lapse rates. The tables were based upon a selected sub-sample of the combined lapse experience of contributors of LIMRA's long term lapse study observed between 1971 and 1972 policy anniversaries - although several years of data was have to be combined to make it possible to develop mortality tables, the much higher level of lapse rates makes it possible to develop tables with a much smaller volume of exposure.

2.3. *Statistical Analysis of Life Assurance Lapses (Haberman and Renshaw 1986)*

Data covering lapse or withdrawal experience for seven Scottish Life Insurers in the calendar year 1976 were investigated with particular reference to the various policy characteristics. Lapse was used to denote the removing of a policy from the live file due to premature termination of the contract with or without payment of a surrender value.

A Report of the Faculty of Actuaries Withdrawals Research Group published in 1978 presented the data for 1976 in a factual way without attempting to set up any theoretical models. The authors of the Report identified nine characteristics with which the withdrawal rate may be expected to vary. The use of theoretical models for such a data set, however, would have the advantage of providing a structure to the date so as to improve the estimation of underlying parameter values. Further, statistical theory would enable different models to be compared and contrasted so that conclusions about the data may be reached. For this reason further study on the data was conducted.

The raw data were edited and the way in which policy lapses, the response, varied with the following covariates was investigated; Age at Entry, Duration of Policy, Office, Type of Policy. A linear model was formed where the response was equated to a systematic component and an error component. Attempts were made to fit a variety of model structures using independent normal homoscedastic errors to the following variables:

- annual lapse rate
- lapse frequency
- log of odds of lapsing

One of the problems met in the investigation was caused by the very nature of lapses. The withdrawal of a policy, unlike a death claim is voluntary. Hence groups of policies with given characteristics may vary both in their overall propensity to withdraw and in the timing of these withdrawals. This element of volition means that cross-sectional investigations are deficient in attempting to describe such phenomena. Hence a cohort approach is more natural and satisfactory. The situation is similar to that in demography where, although a cross-sectional approach may be adequate for mortality investigations, it is unsatisfactory in attempting to describe phenomena like fertility, first marriage, remarriage and divorce. Indeed in these cases such an approach can often lead to fallacious conclusions.

The pertinent details and results are contained in the paper but it was concluded that modeling large complex data sets may be viewed as a balancing act between modeling complexity and the need to encapsulate the salient underlying features present in the data. The simpler the model, the simpler the interpretation of the underlying data generating mechanism! Modeling does not necessarily have a unique solution, but a model may be deemed appropriate only if it achieves this goal! One way of assessing this is through a thorough graphical analysis of model residuals which, ideally, should be 'pattern free'. Additionally, what might be termed 'fine tuning' might then be attempted and its effects formally assessed. The development of Generalised Linear Modeling facilitates such modeling objectives.

2.4. *Lapse Statistics (Records of the Society of Actuaries 1983)*

This record included discussions held at a workshop of the Society of Actuaries entitled *Individual Life Insurance and Annuity Product Development Section – Selected Topics.*

Statistics were presented to support the suggestion that whereas the most important cost of pricing consideration in a term insurance policy would be mortality, the ultimate level of mortality and its effect on cost would be greatly affected by persistency.

Lapse rates were considered under two types of term policies distinguished by the nature of their rate scale. Category 1 Term Policies allowed for the variation of rates on policy renewal whereas Category 2 Term Policies had new issue rates and renewal rates that were equivalent.

Statistics presented on Category 1 were as follows:

- First year lapse rates range from low 20%'s to high 30%'s.
- Second and later years range from high 20%'s to low 40%'s
- Unlike traditional situations, lapse rates grow worse by duration and, typically, are higher for older ages than for younger ages. The reason is probably linked to the fact that increases in the term rate from one year to the next is far greater at the older ages and there appears to be a strong connection between the increase in the term rates and the lapse rate.
- Lapse rates worsen as policy size increases reflecting the selection by the mode sophisticated buyer. There is as much as a 10% point difference in lapse rates for small versus large policies.

Statistics presented on Category 2 were as follows:

- First year lapse rates are in the low 20%
- Lapse rates after the first year appear to be level in the range 24% to 30%.
- Lapses also increase as the policy size increases, reflecting the poorer persistency inherent in the large policy.

Additional statements were that in the early years of a policy, lapses result in a loss since acquisition cost has not been recovered. On the other hand, in later policy years, increased lapse rates generally result in anti-selection and poorer than expected mortality experience since an inordinate proportion of the lapses will be select and those whose health has deteriorated will tend to persist. The way to mitigate the resulting potential loss is to design products that have a level rate for a number of years with a reduced first year commission that in total pays more commission to the agent but provides the company with revenues to defray the cost of high lapses.

2.5. *Mortality Rates as a function of Lapse Rates (Mortality and Lapse Discussions SOA 1998)*

This paper presents a panelists' discussion on the results of two Society of Actuaries Research studies. The first study entitled *Mortality Rates as a function of Lapse Rates (MRLR)* explored the traditional anti-selection hypothesis, examined the difference in mortality between companies and related them to the differences in overall lapse rates. The other study *Analysing the relationship between mortality and lapse rates* assessed the suitability of models involving random mortality rates in analysing insured life mortality and developed models for the relationship between mortality and lapse rates.

MRLR covered mortality experience in 1991 and 1992 (13 US Companies with a total exposure of 1.293 trillion and 1.5 billion deaths) by amount on standard, ordinary policies issued on a smoker and nonsmoker basis. To compare mortality with lapse experience, companies were segregated into high, medium and low lapse categories and then the mortality for each of the categories studied in turn. The experience, however, was not segregated by plan.

As expected there were substantial differences in the groups of companies for all durations. The lowest lapse rate was 11.3%, the medium 13.9% and the high 18.2% for an aggregate lapse rate over all durations of 14.8%. A surprising result, on the other hand, was the convergence of mortality rates at high durations in addition to an increasing trend thought to be related to lapse rates.

The objective of the second study was to develop a model for analysing the relationship between mortality and lapses on an individual life basis. The study began be defining selective lapsation as the good risks that have a greater tendency to lapse their life insurance policies than the poorer risks. Good risks may feel that they do not need life insurance or may see more attractive options elsewhere in terms of replacing their life insurance policies. This results in a poorer group of persistent policyholders. If this occurs over time then the group of persistent policyholders gets worse and worse and experience on mortality worsens at later durations. In a case where the product is traditional and there is experience on mortality and lapses and nothing different is expected then there is not anything special that needs to be done. However, if heavy losses are expected then the pricing actuary needs to think about the impact that the heavy lapses will have on mortality experience.

Selective lapsation occurs because although an insurer attempts to classify risks, it's impossible to end up with a completely homogeneous group of insureds. There will always be a good risk/bad risk mix within a group and the possibility of selective lapsation. In addition there is also the possibility of deterioration of some insured lives and every effort on the part of the insured to keep his or her policy in force.

A random value, frailty, was defined in the study to mean the length of time the individual is expected to live. The distribution of this variable was studied and a mortality model generated that allowed for heterogeneity within a group. The model was then extended to allow for deterioration in the health of an individual so that it could be set in a life insurance context and allow for both mortality and lapse. An extensive discussion of the model is included in the actual paper.

3. Methodology

3.1. *(Modeling Surrender and Lapse Rates)*

The Methodology used in this project is largely based on Changki kim (2005) on Modeling Surrender and Lapse Rates with Economic Variables. It is therefore first important to understand Changki Kim (2005)

Changki summarized a few existing models used at the time by some insurance companies and went on the develop one of his own. His study was carried out for policies in Korea which were first categorized into

- Protection: includes Whole of Life and Term Assurance Policies which make payments on the death of a policyholder at any point in his life or during a pre-specified term respectively.

- Endowment: Pure Endowments that provide a payment if the policyholder is alive at the end of a certain term. No payment is made if the policyholder dies during the term.

- Annuity: a series of payments made at regular intervals as long as the policyholder is alive.

- Education: a savings scheme where the policyholder makes regular contributions over a specified term and when the policy matures the maturity value is used to pay school fees. The savings earn an interest over the term.

Some of the models currently used by insurance companies included (only two out of five models are mentioned; the first because Changki compared his results with those provided by that model; and the second just as a further example to show what the models look like):

A. Arctangent model

B. Parabolic model

The Arctangent model is stated as follows:

$$q_s = a + b * \arctan(m\Delta - n)$$

where q_s is the monthly surrender rate

Δ is reference market rate – crediting rate – surrender charges

a, b, m, n are constants

The Parabolic model is:

$$q_s = a + b * sign(\Delta) * \Delta^2$$

where q_s is the monthly surrender rate

Δ is reference market rate – crediting rate – surrender charges

a, b are constants

$sign(\ )$ is +1 if $(\ )$ is positive, and -1 if $(\ )$ is negative

In both models, the 'reference market rate' refers to the prevailing interest rate on debt instruments that are available in the capital markets. These include Treasury Bills and Treasury Bonds.

The 'crediting rate' is the interest used in the calculation of premium rates for the products. For all the products the policyholder is required to pay a calculated premium in order to get a benefit. It is the benefits payable to the policyholder and the event on which they are paid that have been described above.

Surrender charges are a deduction made on the expected benefit when it is paid on surrender.

The difference between Lapse and Surrender is that Lapses occur when the policyholder fails to pay premiums for a certain period of time. On the other hand a Surrender is an active move made by the policyholder to withdraw from using a product in which event a calculated surrender value becomes payable to the policyholder. Lapses in a sense occur automatically when it is noticed from the policy providers' books that premiums have not been paid for a given policyholder over a certain period of time. In this project, however, the terms have been used synonymously.

It is clear from the formulae used above that the main determinant variable is interest and more accurately the difference between the reference and crediting interest rates. The crediting rates are regarded as constant in the case of the mentioned types of policy. However there may be periods where a bonus is provided in addition to the sum assured on the policy and again for the mentioned products the bonuses would be provided at the discretion of the policy provider.

Since the credited rates are constant, it follows that the lapses really just depend on the reference market rates. A decrease in the reference rates leads to a decrease in surrender rates whereas an increase in reference rates causes an increase in surrender rates.

Changki argues that whereas interest rates have a significant impact on lapses, there are other economic variables that affect lapse. These include the following variables:

A. Policy Age Since Issue
B. Economy Growth Rate
C. Unemployment Rate
D. Financial Crises (the data used in the investigation was for a 3 year period that also included the South East Asia Financial Crisis of December 1997 to December 1998)
E. Seasonal effects

In addition he also argues that the lapses caused by changes in the reference rates do not necessarily happen immediately. An immediate increase in reference rates may be the reason that some policies lapse, not immediately, but after a 3 or 4 or even 5 month period. The reason is that whereas some policyholders make an immediate decision to surrender their policies, there are others who may prefer to wait a while to see whether the increase is a lasting one. Changki's model therefore factors in the difference between reference and crediting rates but with different lag periods. The meaning of the term model now changes from the singular to the plural. Changki developed 2 Generalised Linear Models for the monthly lapse rates. These were (explained further on the next page)

- Logit model
- Complementary Log-Log model

The Logit model is of the form:

$$\ln\left(\frac{q_s}{1-q_s}\right) = \beta_0 + \beta_1 V_1 + \ldots + \beta_n V_n$$

where q_s is the monthly surrender rate

V_j is the j-th predictor variable. There are n predictor variables in total.

β_j is the coefficient of the j-th predictor variable.

The Complementary Log-Log (CLL) model is of the following form:

$$\log\left(-\log\left(1-q_s\right)\right) = \beta_0 + \beta_1 V_1 + \ldots + \beta_n V_n$$

again q_s is the monthly surrender rate

V_j is the j-th predictor variable. There are n predictor variables in total.

β_j is the coefficient of the j-th predictor variable.

Changki carried out Logistic Regression to obtain, for each of the fields mentioned above for Korean lapse data, values for the coefficients in the above Logit and CLL models. He further carried out tests of significance on each of the explanatory variables. His results showed that each variable had a *p*-value less than 1% concluding that each value was significant in explaining the response.

Changki then performed the following steps to refine the model:

a. Selected a few significant explanatory variables.

Few significant variables were selected using the Backward Elimination Method. Four criteria were used in this process though we only state two of the criteria since all four resulted in the same reduced model. The two are:

- Akaike Information Criteria (AIC)
- Bayesian Information Criteria (BIC)

b. Set up reduced models with the explanatory variables.

The variables used that remained after the Backward Elimination process above were used to construct new Logit and CLL models. In other words new parameters were obtained for those variables. This would result in a more superior model than that first since it is the most significant variables that would now be included.

Changki then compared the lapse predictions of the new models with those resulting from the 'arctangent model' predictions.

For Protection and Education policies, the overall estimated errors of the Logit and CLL models were smaller than those of the arctangent model.

For the Endowment and Annuity plans, no conclusion could be made on whether the Logit and CLL models were better than the arctangent.

c. Transformed the policy age.

The policy age was transformed because the surrender rates increase during the early years and decrease as time passes. There is therefore a real possibility that the fit may be decreased if the real policy ages are used without transformation.

The transformation formula was $-\frac{1}{policyage}$. In the case of the Protection and Endowment plans, the Logit and CLL models remained better after transformation than the arctangent model. The Logit and CLL model for the Annuities was now also better than the arctangent!

3.2. (*Determination of Surrender and Lapse rates*)

The actual methodology carried out in this project derives from Changki (2005) which has been described above.

The steps carried out are as follows:

Segregate data according to the policy type categories and so that we can develop lapse rates for each policy type. The lapse rates are expected to be different across policy types. Policy types are; Protection, Education, Endowment, Annuity.

Assign the following predictors for each lapse month under investigation

- Economy Growth Rate
- Inflation Rate (with lag 0 to 5 months)
- Difference between Reference and Market Rates (with lag 0 to 5 months)
- Unemployment Rate
- Seasonal effects

On R (statistical software) fit a Generalised Linear Model (GLM) of the following form to the data:

$$\ln\left(\frac{q_s}{1-q_s}\right) = \beta_0 + \beta_1 V_1 + \ldots + \beta_n V_n$$

where q_s is the month of lapse

V_j is the j-th predictor variable. There are 9 predictor variables in total.

β_j is the coefficient of the j-th predictor variable.

We fit a Generalised Linear Model because of its property of constraining the 'odds' response variable between 0 and 1. The response variable in the equation is the natural logarithm of the ratio of lapse in a given month to non-lapse in the same month.

Carry out Logistic Regression analysis to establish which of the estimated variables are significant.

Perform Backward Elimination using the Akaike Information Criteria to reduce the model to a parsimonious one. A parsimonious model is one which is neither over-fit nor under-fit ie the model is a best fit for a given set of data.

Set up a reduced model with the variables that have resulted from above. In this step we perform step '2.' again in order to establish a new Generalised Linear Model that incorporates only the variables that have already produced the best fit as per the Akaike Information Criterion.

Transform policy age at lapse using $-\frac{1}{policyage}$ and test using Akaike Information Criteria to see if the model has improved.

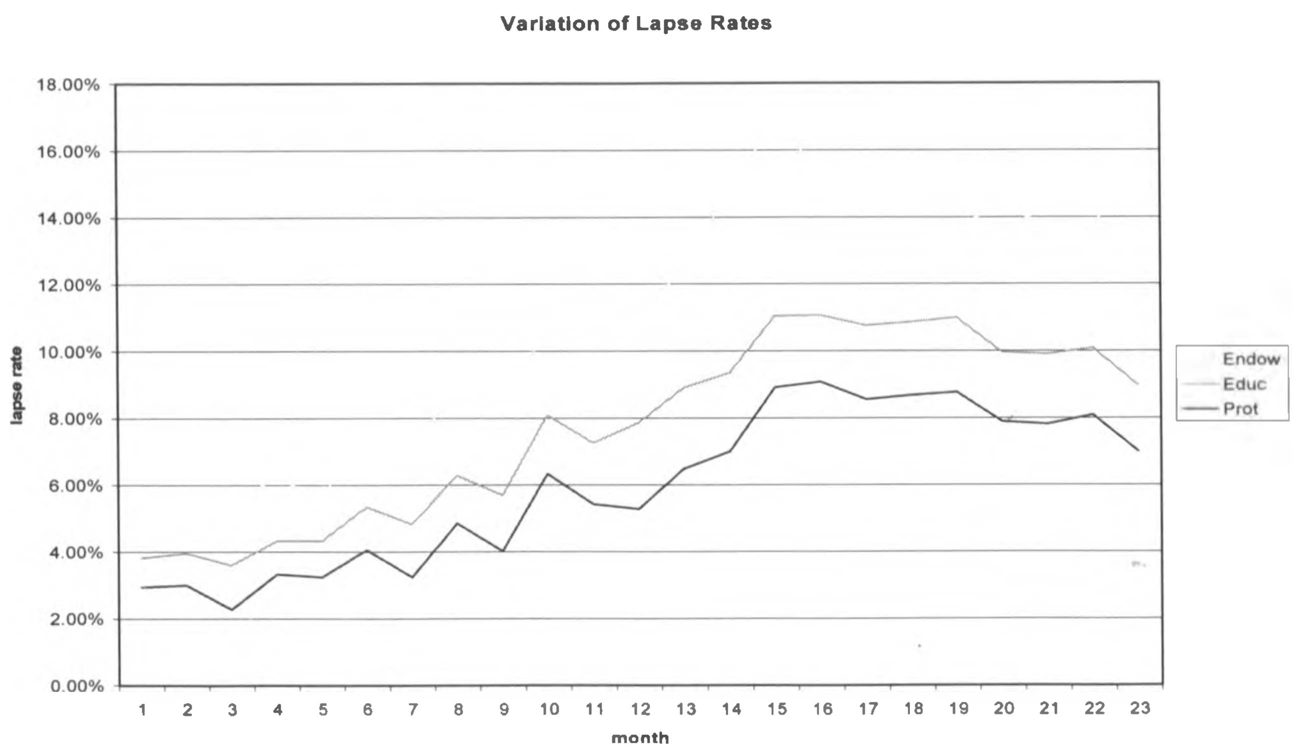
4. Data, Data Analysis and Results

4.1. (Data)

The following table summarises the data that was used in the project. The actual data used is contained in the appendix.

Year	2005	2006
Number of Lapses	305	281
New Policies in Year	1168	1464
Total Policies End of Year	863	2046

The graph that follows gives an impression of the lapse variation across the study period.



Other Economic variables used are as follows:

Economic Growth Rates denoted by the increase in Gross Domestic Product were obtained from the Kenya National Bureau of Statistics (KNBS).

Inflation Rates and Unemployment Rates were obtained from the (KNBS).
The Reference rates used were the 91 Day Treasury bill rates over the period of lapses.

An assumption of 4% per annum was made on the credited rates. Credited Rates are assumed to be constant for the data used in the study and it is the difference between reference market rates and credited rates that is important.

4.2. (Data Analysis)

After segregating the data according to policy types, an attempt was made to fit a Generalised Linear Model (GLM) using the statistical software R.

The following summary table was generated for the GLM fit for the education policies.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-7.9617	194.8417	-0.041	0.967
Gdp	88.2014	2586.8563	0.034	0.973
Inflzero	16.5826	1307.3499	0.013	0.990
Inflone	-21.0914	2028.0891	-0.010	0.992
Infltwo	4.1229	1974.2799	0.002	0.998
Inflthr	18.2315	2012.0619	0.009	0.993
Inflfour	13.9054	2390.2044	0.006	0.995
Inflfive	-44.0966	2126.9539	-0.021	0.983
Diffzero	6.7615	1115.0955	0.006	0.995
Diffone	-10.8662	1043.4935	-0.010	0.992
Difftwo	0.3157	623.0255	0.001	1.000
Diffthr	9.0859	565.0574	0.016	0.987
Difffour	-12.8040	589.8052	-0.022	0.983
Difffive	6.0967	451.6567	0.013	0.989
Unemp	NA	NA	NA	NA

From the table it can be seen that none of the coefficients are significant in explaining the response. The same result was obtained for the endowment and protection policies (there we no annuity policies). We draw a conclusion that is contrary to what was expected. The economic variables used do not affect the lapse rates is the given months. We cannot therefore proceed to develop a model. In the next section we attempt to explain why the response seems not to be affected by the predictors.

5. Conclusion

This project has reviewed past research on lapse rates that has been previously conducted. Further the project has attempted to develop a generalised linear model for the lapse rates of an insurance company where the predictor variables are economic variables. However it has been found that the economic variables are not significant for the lapse rates in the data used.

There are three possible reasons that may cause the variables to be insignificant:

Either the lapse rates for the insurance company whose data was used are actually not dependent on the economic variables used or,

The period of data that the lapse data represents is too short to allow for any relationship between lapses and economic variables to be elucidated (the period of lapse data is two years) or,

The data that has been used belongs to an insurance company which targets low income earners who may remain largely unaffected by economic fluctuations since their incomes are small.

In conclusion it is recommended that more substantial lapse data (from several insurance companies) should be collected over several years analysed to establish how the lapse rates are affected by economic fluctuations.

References and Appendix

ALBERT F. AND BRAGG G.1998 Mortality Rates as a function of Lapse Rates. *Session 31PD Mortality and Lapse Rates.1-6.*

BRITTON, D.W 1983 Lapse Rates in Product Development.
Record of Society of Actuaries.3-8.

BRZEZINZKI, JOSEPH R. 1975 LIMRA 1971-72 Expected Lapse Tables. *Transactions of the Society of Actuaries* 27:267 – 312.

DOBSON J.A 2002 An Introduction to Generalised Linear Models. *Chapman and Hall/CRC*

FARAWAY J.J 2006 Extending the Linear Model with R *Chapman and Hall/CRC*

JONES B.L.1998 Analysing the relationship between mortality and Lapse Rates. *Session 31PD Mortality and Lapse Rates.6-22*

RENSHAW A.E AND HABERMAN S 1986 Statistical analysis of Life Assurance Lapses. *Journal of Insurance 459-497*

RICHARDSON, C.F.B, AND J.M. HARTWELL. 1951 Lapse Rates. *Transactions of the Society of Actuaries* 375-96.

SCOTT W.F 1999 Life Assurance Mathematics

employment

MonthID	Empl	Gdp	Inflzero	Inflone	Infltwo	Inflthr	Inflfour	Inflfive	Ref	Diffzero	Diffone	Difftwo	Diffthr	Difffour	Difffive	Unemp
1	0.012	0.055	0.110	0.111	0.112	0.113	0.114	0.115	0.083	0.043	0.040	0.011	-0.001	-0.013	-0.017	0.400
2	0.010	0.056	0.108	0.110	0.111	0.112	0.113	0.114	0.086	0.046	0.043	0.040	0.011	-0.001	-0.013	0.400
3	0.019	0.056	0.107	0.108	0.110	0.111	0.112	0.113	0.086	0.046	0.046	0.043	0.040	0.011	-0.001	0.400
4	0.019	0.057	0.106	0.107	0.108	0.110	0.111	0.112	0.087	0.047	0.046	0.046	0.043	0.040	0.011	0.400
5	0.012	0.058	0.105	0.106	0.107	0.108	0.110	0.111	0.087	0.047	0.047	0.046	0.046	0.043	0.040	0.400
6	0.026	0.058	0.104	0.105	0.106	0.107	0.108	0.110	0.085	0.045	0.047	0.047	0.046	0.046	0.043	0.400
7	0.022	0.059	0.103	0.104	0.105	0.106	0.107	0.108	0.086	0.046	0.045	0.047	0.047	0.046	0.046	0.400
8	0.026	0.059	0.107	0.103	0.104	0.105	0.106	0.107	0.087	0.047	0.046	0.045	0.047	0.047	0.046	0.400
9	0.028	0.060	0.110	0.107	0.103	0.104	0.105	0.106	0.086	0.046	0.047	0.046	0.045	0.047	0.047	0.400
10	0.025	0.060	0.114	0.110	0.107	0.103	0.104	0.105	0.082	0.042	0.046	0.047	0.046	0.045	0.047	0.400
11	0.030	0.060	0.117	0.114	0.110	0.107	0.103	0.104	0.078	0.038	0.042	0.046	0.047	0.046	0.045	0.400
12	0.034	0.061	0.121	0.117	0.114	0.110	0.107	0.103	0.081	0.041	0.038	0.042	0.046	0.047	0.046	0.400
13	0.035	0.061	0.124	0.121	0.117	0.114	0.110	0.107	0.082	0.042	0.041	0.038	0.042	0.046	0.047	0.400
14	0.054	0.061	0.128	0.124	0.121	0.117	0.114	0.110	0.080	0.040	0.042	0.041	0.038	0.042	0.046	0.400
15	0.058	0.062	0.131	0.128	0.124	0.121	0.117	0.114	0.076	0.036	0.040	0.042	0.041	0.038	0.042	0.400
16	0.044	0.062	0.135	0.131	0.128	0.124	0.121	0.117	0.070	0.030	0.036	0.040	0.042	0.041	0.038	0.400
17	0.040	0.062	0.138	0.135	0.131	0.128	0.124	0.121	0.070	0.030	0.030	0.036	0.040	0.042	0.041	0.400
18	0.035	0.063	0.142	0.138	0.135	0.131	0.128	0.124	0.066	0.026	0.030	0.030	0.036	0.040	0.042	0.400
19	0.033	0.063	0.145	0.142	0.138	0.135	0.131	0.128	0.059	0.019	0.026	0.030	0.030	0.036	0.040	0.400
20	0.037	0.064	0.141	0.145	0.142	0.138	0.135	0.131	0.060	0.020	0.019	0.026	0.030	0.030	0.036	0.400
21	0.043	0.064	0.137	0.141	0.145	0.142	0.138	0.135	0.065	0.025	0.020	0.019	0.026	0.030	0.030	0.400
22	0.042	0.065	0.133	0.137	0.141	0.145	0.142	0.138	0.068	0.028	0.025	0.020	0.019	0.026	0.030	0.400
23	0.036	0.066	0.129	0.133	0.137	0.141	0.145	0.142	0.064	0.024	0.028	0.025	0.020	0.019	0.026	0.400

protection

MonthID	Prot	Gdp	Inflzero	Inflone	Infltwo	Inflthr	Inflfour	Inflfive	Ref	Diffzero	Diffone	Difftwo	Diffthr	Difffour	Difffive	Unemp
1	0.030	0.055	0.110	0.111	0.112	0.113	0.114	0.115	0.083	0.043	0.040	0.011	-0.001	-0.013	-0.017	0.400
2	0.030	0.056	0.108	0.110	0.111	0.112	0.113	0.114	0.086	0.046	0.043	0.040	0.011	-0.001	-0.013	0.400
3	0.023	0.056	0.107	0.108	0.110	0.111	0.112	0.113	0.086	0.046	0.046	0.043	0.040	0.011	-0.001	0.400
4	0.033	0.057	0.106	0.107	0.108	0.110	0.111	0.112	0.087	0.047	0.046	0.046	0.043	0.040	0.011	0.400
5	0.033	0.058	0.105	0.106	0.107	0.108	0.110	0.111	0.087	0.047	0.047	0.046	0.046	0.043	0.040	0.400
6	0.041	0.058	0.104	0.105	0.106	0.107	0.108	0.110	0.085	0.045	0.047	0.047	0.046	0.046	0.043	0.400
7	0.033	0.059	0.103	0.104	0.105	0.106	0.107	0.108	0.086	0.046	0.045	0.047	0.047	0.046	0.046	0.400
8	0.049	0.059	0.107	0.103	0.104	0.105	0.106	0.107	0.087	0.047	0.046	0.045	0.047	0.047	0.046	0.400
9	0.040	0.060	0.110	0.107	0.103	0.104	0.105	0.106	0.086	0.046	0.047	0.046	0.045	0.047	0.047	0.400
10	0.063	0.060	0.114	0.110	0.107	0.103	0.104	0.105	0.082	0.042	0.046	0.047	0.046	0.045	0.047	0.400
11	0.054	0.060	0.117	0.114	0.110	0.107	0.103	0.104	0.078	0.038	0.042	0.046	0.047	0.046	0.045	0.400
12	0.053	0.061	0.121	0.117	0.114	0.110	0.107	0.103	0.081	0.041	0.038	0.042	0.046	0.047	0.046	0.400
13	0.065	0.061	0.124	0.121	0.117	0.114	0.110	0.107	0.082	0.042	0.041	0.038	0.042	0.046	0.047	0.400
14	0.070	0.061	0.128	0.124	0.121	0.117	0.114	0.110	0.080	0.040	0.042	0.041	0.038	0.042	0.046	0.400
15	0.089	0.062	0.131	0.128	0.124	0.121	0.117	0.114	0.076	0.036	0.040	0.042	0.041	0.038	0.042	0.400
16	0.091	0.062	0.135	0.131	0.128	0.124	0.121	0.117	0.070	0.030	0.036	0.040	0.042	0.041	0.038	0.400
17	0.086	0.062	0.138	0.135	0.131	0.128	0.124	0.121	0.070	0.030	0.030	0.036	0.040	0.042	0.041	0.400
18	0.087	0.063	0.142	0.138	0.135	0.131	0.128	0.124	0.066	0.026	0.030	0.030	0.036	0.040	0.042	0.400
19	0.088	0.063	0.145	0.142	0.138	0.135	0.131	0.128	0.059	0.019	0.026	0.030	0.030	0.036	0.040	0.400
20	0.079	0.064	0.141	0.145	0.142	0.138	0.135	0.131	0.060	0.020	0.019	0.026	0.030	0.030	0.036	0.400
21	0.078	0.064	0.137	0.141	0.145	0.142	0.138	0.135	0.065	0.025	0.020	0.019	0.026	0.030	0.030	0.400
22	0.081	0.065	0.133	0.137	0.141	0.145	0.142	0.138	0.068	0.028	0.025	0.020	0.019	0.026	0.030	0.400
23	0.070	0.066	0.129	0.133	0.137	0.141	0.145	0.142	0.064	0.024	0.028	0.025	0.020	0.019	0.026	0.400

education

MonthID	Educ	Gdp	Inflzero	Inflone	Infltwo	Inflthr	Inflfour	Inflfive	Ref	Diffzero	Diffone	Difftwo	Diffthr	Difffour	Difffive	Unemp
1	0.009	0.055	0.110	0.111	0.112	0.113	0.114	0.115	0.083	0.043	0.040	0.011	-0.001	-0.013	-0.017	0.400
2	0.010	0.056	0.108	0.110	0.111	0.112	0.113	0.114	0.086	0.046	0.043	0.040	0.011	-0.001	-0.013	0.400
3	0.013	0.056	0.107	0.108	0.110	0.111	0.112	0.113	0.086	0.046	0.046	0.043	0.040	0.011	-0.001	0.400
4	0.010	0.057	0.106	0.107	0.108	0.110	0.111	0.112	0.087	0.047	0.046	0.046	0.043	0.040	0.011	0.400
5	0.011	0.058	0.105	0.106	0.107	0.108	0.110	0.111	0.087	0.047	0.047	0.046	0.046	0.043	0.040	0.400
6	0.013	0.058	0.104	0.105	0.106	0.107	0.108	0.110	0.085	0.045	0.047	0.047	0.046	0.046	0.043	0.400
7	0.016	0.059	0.103	0.104	0.105	0.106	0.107	0.108	0.086	0.046	0.045	0.047	0.047	0.046	0.046	0.400
8	0.014	0.059	0.107	0.103	0.104	0.105	0.106	0.107	0.087	0.047	0.046	0.045	0.047	0.047	0.046	0.400
9	0.017	0.060	0.110	0.107	0.103	0.104	0.105	0.106	0.086	0.046	0.047	0.046	0.045	0.047	0.047	0.400
10	0.016	0.060	0.114	0.110	0.107	0.103	0.104	0.105	0.082	0.042	0.046	0.047	0.046	0.045	0.047	0.400
11	0.016	0.060	0.117	0.114	0.110	0.107	0.103	0.104	0.078	0.038	0.042	0.046	0.047	0.046	0.045	0.400
12	0.026	0.061	0.121	0.117	0.114	0.110	0.107	0.103	0.081	0.041	0.038	0.042	0.046	0.047	0.046	0.400
13	0.024	0.061	0.124	0.121	0.117	0.114	0.110	0.107	0.082	0.042	0.041	0.038	0.042	0.046	0.047	0.400
14	0.024	0.061	0.128	0.124	0.121	0.117	0.114	0.110	0.080	0.040	0.042	0.041	0.038	0.042	0.046	0.400
15	0.021	0.062	0.131	0.128	0.124	0.121	0.117	0.114	0.076	0.036	0.040	0.042	0.041	0.038	0.042	0.400
16	0.020	0.062	0.135	0.131	0.128	0.124	0.121	0.117	0.070	0.030	0.036	0.040	0.042	0.041	0.038	0.400
17	0.022	0.062	0.138	0.135	0.131	0.128	0.124	0.121	0.070	0.030	0.030	0.036	0.040	0.042	0.041	0.400
18	0.022	0.063	0.142	0.138	0.135	0.131	0.128	0.124	0.066	0.026	0.030	0.030	0.036	0.040	0.042	0.400
19	0.022	0.063	0.145	0.142	0.138	0.135	0.131	0.128	0.059	0.019	0.026	0.030	0.030	0.036	0.040	0.400
20	0.021	0.064	0.141	0.145	0.142	0.138	0.135	0.131	0.060	0.020	0.019	0.026	0.030	0.030	0.036	0.400
21	0.021	0.064	0.137	0.141	0.145	0.142	0.138	0.135	0.065	0.025	0.020	0.019	0.026	0.030	0.030	0.400
22	0.020	0.065	0.133	0.137	0.141	0.145	0.142	0.138	0.068	0.028	0.025	0.020	0.019	0.026	0.030	0.400
23	0.020	0.066	0.129	0.133	0.137	0.141	0.145	0.142	0.064	0.024	0.028	0.025	0.020	0.019	0.026	0.400