KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY

MODELING LAPSE RATES WITH ECONOMIC VARIABLES: THE VECTOR AUTOREGRESSIVE MODEL APPROACH

By

ARYEH ISAAC

A THESIS SUBMITTED TO THE DEPARTMENT OF MATHEMATICS, KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF MSc. ACTUARIAL SCIENCE.

April, 2016

Declaration

I hereby declare that this submission is my own work towards the award of the MSc degree and that, to the best of my 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.

Isaac Aryeh (PG1475613)	....	....
Student	Signature	Date

Certified by:

Dr. Omari Sasu	....	....
Supervisor	Signature	Date

Certified by:

Prof. S. K Amponsah	....	....
Head of Department	Signature	Date

Dedication

This thesis is dedicated to my parents: Alexander Adu Aryeh and Felicia Kwartemaa Aryeh for their support towards my education.

Abstract

Life insurance contracts are accompanied by risks. This study focuses on the risk involving termination of life policies by the policyholders. This study seeks to identify the main determinants of lapse rates in the Ghanaian life insurance industry. The data available span from 2006 to 2014 and were recorded on monthly basis. Literature research on predictive modeling of lapse rates in the insurance industry led to the choice for predicting with a VAR Model. A total of five (5) variables were analyzed on their relationship with the average lapse rate. Of these variables inflation, stock market return and interest rate proved valuable for modeling. The vector auto-regression (VAR) model was used to capture the evolution and the interdependencies between the variables. All the variables in the VAR were treated symmetrically by including for each variable an equation explaining its evolution based on its own lags and the lags of all the other variables in the model. The main limitation of this research was the small amount of data and their level of detail. Since the results of this research are promising it is recommended to extend the research to other parts of the country. To increase the statistical strength and accuracy of the inferences that can be made it is also recommended to examine the lapse rates at a policyholder level.

Acknowledgements

My humble gratitude goes to the Almighty God for his guidance and motivation towards the successful completion of this work.

I wish to extend my heartfelt gratitude and thanks to my able supervisor, Dr. A.Y Omari-Sasu for his patience and immense contribution which invaluably improved this thesis.

I am also very grateful to Priscilla Hansen, Michael Sakyi and Charles Avornyo for their support towards the successful completion of this work.

My profound thanks go to my brothers Eben, Ben, Emma and Alex Aryeh for their encouragement and support towards the attainment of this noble dream.

However, I am completely answerable to any limitation that may be detected in this work.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background of the Study

Life insurance is both a large industry and the most valuable method for individuals to financially protect their loved ones upon death. Mojekwu (2011) pointed out that insurance companies are vital segments of the financial system in African countries and it is necessary to ensure their stability in order to maintain public confidence in the financial structure. Most African countries were using the United Kingdom assured lives and annuitant tables such as the A24-29, a (55), A49-52 rates which are rated up (Mojekwu, 2011). Life insurance companies, deal with long-term as well as short-term plans. Consequently, it is necessary to safeguard the interest of the policyholders and promote national interest.

Life insurance is essentially a contract that guarantees a sum of money to designated beneficiaries upon the insured's death or perhaps to the insured if he or she lives beyond a certain age. The cast of characters involved with a life insurance policy other than the insurer include: The owner of the policy, the person whose life is being insured and the beneficiaries who are paid under the policy. The owner of the policy has the power to name or change the beneficiary, the right to assign the policy, cash it in it for its surrender value, or use it as collateral in obtaining a loan, and also the obligation to pay the premiums. The same person may occupy all three positions by naming their estate as beneficiary, or each of the three positions may be held by a separate person.

Just as insurance is broken down into several categories, life insurance also contains various subcategories. Some common forms of life insurance policies are Whole Life, Term Life, Endowment Life Insurance, Industrial Life Insurance, Mutual Life Insurance, Group Insurance, Annuities, Universal Life Insurance etc. Each of these policies is designed to address a particular set of circumstances that people may encounter (The Messenger, 2015). With the pace of product innovation the life insurance and annuity industry has seen over the past two decades and the corresponding increase in the complexity of financial products, the list of actuarial assumptions that can materially impact product profitability has grown and includes a group of assumptions that are collectively referred to as "policyholder behavior assumptions"(or "customer behavior assumptions"). In an actuarial context, the term policyholder behavior refers to any assumption for which experience is directly driven by the decisions policyholders make regarding the exercise of benefits and guarantees within their contracts (Campbell,*et al*, 2014). Policyholders may exercise their right to terminate a contract; this event is called a lapse. One of the problems with policies that lapse (at an early stage) is that not enough premium payments have been made to cover the policy expenses. (Michorius, 2011).

Lapsation and policy terminations may be initiated by consumers at any time by failing to pay premium billings on term and whole life policies.

The non-renewal of term and whole life policies may also occur on renewal anniversaries for policies with guaranteed renewable riders (Mauer & Holden, 2007).

From the company's perspective, life insurance products, and especially whole life products, typically entail large underwriting and upfront origination costs, heavily driven by sales costs and commissions. (Carson & Dumm. 2000). This cost structure provides insurers a strong motive toward lower lapse rates. Too high a lapse rate may impair the ability of the insurer to recoup these costs, given the projected benefits payouts required under these lines of insurance. Companies that are in a stronger market position will be able to price more aggressively than companies in a weaker market position.

According to Salami (1996) existing customers claimed that they allow the policies to lapse due to poor premium pricing and poor relationship marketing on the side of the life policy operators. Furthermore, they complained that in Nigeria, premium charged in life assurance do not take cognisance of inflation rate. In most cases, the value of the sum assured which is payable would have been eroded by inflation in the economy. However, this non- incorporation of inflation factor into the returns on life-policy impacts negatively on the interest of potential life-policy holders thereby affecting the trends and patterns of the mode of exit. On the other hand, the life insurance operators argued that inflation is taken care of by issuing policies on the profit basis Mojekwu (2011).

It is against this background that the study seeks to examine the main determinants of lapse rate in Ghana with the aim of modelingthe lapse rates.

1.2 Problem Statement

According to the Geneva Association (2012), high lapse rates remain a key issue for life insurance industry. In its second half-year financial stability report in 2011, the European Insurance and Occupational Pensions Authority (EIOPA) expressed concerns regarding potential risks with "higher than expected lapse rates in life insurance"(the Geneva Association, 2012)

According to Sandhu and Bice (2014), although most insurers have some form of retention programs in place, the upward trending nature of lapse rate suggests that these programs have had limited effect. The surrendering of a private life insurance contract is subject to individual choice. Although many insurance contracts include the option to surrender, changing or cancelling them still has a cost, as with all contracts (Geneva association, 2012). For example, in the case of a life insurance contract, there is usually a surrender charge for which the insurer deducts the pre-financed costs from the surrender value. In addition, customers generally buy life insurance policies for long-term purposes. In contrast to other financial products, customers do not generally expect liquidity from life insurance policies at any given time but rather at predetermined times. They have disincentives to surrender-if there is an option to surrender at all. Given that surrendering a life policy can be a costly process, why would a policyholder still want to do so? What drives insurance policyholders to consider surrendering their policies despite the long-term purpose of these products? What elements influence their decision to surrender? How do insurers manage the risks around surrenders, including liquidity risk? These concerns as highlighted by the Geneva association, (2012) seek to be the focus of the present study.

Literature on life insurance lapsation is very scarce especially in the Ghanaian context. In view of this the study seeks to investigate into the main determinants of lapse rates in Ghana with the aim of modeling the lapse rates.

1.3 Research Objectives

The main research objective was to identify the main drivers of lapse rate in Ghanaian life insurance.

The specific objectives are to:

- Identify all the significant variables that contribute to lapse rate

- Identify amongst the variables, the most significant in determining the lapse rate

- To determine which model is most fit for forecasting lapse rates
- Ascertain whether the current lapse model used by life insurers can be improved

1.4 Significance of the Study

The risks arising from lapse are of high economic importance. As such, lapsation is of interest not only to academics, but is also highly relevant for the industry, regulators, and policymakers, especially in regard to designing an appropriate regulatory environment. The study will further broaden the knowledge base of life insurance companies on the drivers of lapse rate. It will also add up to the literature which will serve as a foundation for researchers and students studying in this area.

1.5 Scope of the Study

The study focused on the main drivers of lapse rate in Ghanaian Life Insurance Company with the aim of modeling the lapse rate.

1.6 Organisation of the Study

The study comprised five chapters. Chapter one forms the introductory aspect of the study which consisted of the background, the statement of the problem, purpose, significance and organization of the study. Chapter two comprises the literature review which was made up of the definition, other sub topics, concept and theories underlining the study's variable, and the empirical frame work of the work. Chapter three deals with the methodological part of the study consisting of the research design used, the targeted/accessible population, the sample and sampling techniques, instrument, data analysis and data collection procedure. Chapter four presents the analysis of the data collected. Finally, the chapter five deals with the summary, findings, recommendations and conclusion and other areas for further research.

Chapter 2

Literature Review

2.1 Introduction

This section reviews models underpinning the study. Below are some of the models reviewed in relation to the study.

2.2 Lapse models

2.2.1 Predictive Modeling

Predictive modeling can be thought of as the application of certain algorithms and statistical techniques to a data set to better understand the behavior of a target variable based on the co-relationships of several explanatory variables Michorius (2014). It involves analyzing data in order to understand risk, which is what actuaries have been doing for many years. The difference is the use of more advanced mathematics, algorithms and larger quantities of data, which makes the analysis computationally tractable. Rather than relying on a simple understanding of basic risk elements, predictive modeling enables the user to consider many confounding factors simultaneously by mining across a set of scenarios. This analysis will allow for making more informed decisions and limit the amount of subjective judgment required.

According to Michorius (2014), Predictive Modeling is now used by a wide range of practitioners in both life and non-life insurance and is used to answer an ever expanding set of questions.

Furthermore dwelling on the predictive model as reviewed in Michorius (2014), it is observed that predictive models are favored relative to traditional models because they capture more risks and can account for inter-variable correlation (Briere-Giroux *et al*, 2010). Whereas some variables can be accurately predicted with a predictive model, others are more complex and cannot be predicted with high accuracy. Depending on the amount of underlying drivers and the assumed relationship between drivers and variables there are various models which can be opted for. The most basic type of predictive models with explanatory variables are the one-factor models. These models suggest a significant relationship between a single variable and lapse rates. A common predictor is the reference market rate, which is the interest rate provided by a competitor (Briere-Giroux *et al*, 2010). This choice is justified by suggesting that the policy holder (constantly) compares similar products and chooses its policy based on its relative costs. Examples of such functions which are currently used by insurance companies are the

- Arctangent model: $r = a + b \times \arctan(m \times \Delta^{-n})$
- Parabolicmodel:$r = a + b \times \sin(\Delta) \times \Delta^2$
- Exponential model: $r = a + b \times e^{(m \times \frac{CR}{MR})}$

  In which

  r is the monthly lapse rate

  a, b, m, n are coefficients

  Δ is the reference market rate minus crediting rate minus surrender charges

The predictive model is relevant to the study as the study hopes to identify the model that is best fit to forecast lapse rate.

2.2.2 CART - Models

CART models produce classification or regression tree, dependent on whether the dependent variables is categorical or numeric. They are non-parametric forecasting methods. The main procedure of the CART-model is the step by step division of lapse data into smaller groups, based on binary rules. At each step the algorithm select the variables or combinations of variables which provide the greatest purity of data in order to form homogeneous data sets. The algorithm stops with dividing the data set as soon as an (arbitrarily) chosen criterion has been reached. Examples of such criterion have been reached. Examples of such criteria: equality of observations of the explanatory variables in a given class, a minimum number of observations per node or a specific potential of increase in data purity (Loisel & Maume-Deschamps, 2010). The advantage of this method are that the results are easily interpretable, because of the tree structure and that it is non-parametric and nonlinear (StatSoft, n.d.). Another advantage is the fact that CART-models can include continuous as well as categorical predictor variables which is extremely useful with variables such as gender, division and product type (Michorius, 2011).

The main disadvantages are the complexity of the CART algorithms and instability of the model. A small change in data may lead to a huge change in the outcome. Guszcza (2005) states that the model does a poor job t modeling linear structure, but can be used as a pre-test to analyze which variables or combination of variables might be explanatory. The CART-model is just one type of tree model and by far not the most complex. Because of the non-linearity and its subdivision of data, the trees can provide pretty accurate results (Michorius, 2011).

2.2.3 Stochastic Lapse Modeling

According to the Geneva Association, (2012), when calculating reserves or capital requirements, it may be preferable to use the most robust modeling available. Following this mindset, some may desire to model lapse rates stochastically. However, this would require selection of a distribution function for lapses (or the underlying factors affecting lapse rates).

2.2.4 Deterministic Lapse Modeling

The latest proposed PBR framework, presented at the Society of Actuaries Life 2007 Spring Meeting, is for lapse rates to be set deterministically. While this may be a reasonable accommodation given the problems with stochastic models, there will still be questions to consider when setting lapse assumptions.

2.2.5 Generalized Linear Models

Another model which is mentioned in literature for the modeling lapse rates is the generalized linear model. This model combines a number of linear variables into one regression model and uses a link-function, a function which transforms the data distribution of the linear variables, to predict an outcome variable, which is expected to have a distribution from the exponential family of distributions. GLMs are favored for they can provide accurate results when applied to lapse data, similar to the CART-models, while remaining interpretable (Briere-Giroux *et al*, 2010).

A GLM is suited to model many different types of functions, due to its link-function, and can predict (with) continuous as well a binary variables. This characteristics of great use in the insurance industry since the lapse variables can be either zero or one on a policy level and continuous between zero and one on an aggregated level.

2.2.6 Time Series

A time series is a sequential set of data points, measured typically over successive times. It is mathematically defined as a set of vectors $X_t, t = 1, 2, 3, \cdots$ where t represents the time elapsed (Cochrane, 1997; Hipel & McLeod, 1994; Ralchavoen, Lursinsop & Sanguanbhoki, 2003).

Components of a Time Series

A time series in general is supposed to be affected by four main components, which can be separated from the observed data. These components are: Trend, Cyclical, Seasonal and Irregular components. A brief description of these four components is given here.

One simple method of describing a series is that of classical decomposition. The notion is that the series can be decomposed into four elements:

(i) Trend (T_t) - long term movements in the mean;

(ii) Seasonal effects (I_t) - cyclical fluctuations related to the calendar;

(iii) Cycles (C_t) - other cyclical fluctuations (such as a business cycles);

(iv) Residuals (X_t) - other random or systematic fluctuations.

The idea is to create separate models for these four elements and then combine them, either additively

$$X_t = T_t + I_t + C_t + E_t$$

or multiplicatively

$$X_t = T_t \times I_t \times C_t \times E_t$$

Models for Time Series

1. **Time Series Data**

   A time series is a set of statistics, usually collected at regular intervals. Time series data occur naturally in many application areas.

   - economics - e.g., monthly data for unemployment, hospital admissions, etc.
   - finance - e.g., daily exchange rate, a share price, etc.
   - environmental - e.g., daily rainfall, air quality readings.

- medicine - e.g., ECG brain wave activity every 2-8 secs.

The methods of time series analysis pre-date those for general stochastic processes and Markov Chains. The aims of time series analysis are to describe and summarise time series data, fit low-dimensional models, and make forecasts. We write our real-valued series of observations as $X_{-2}, X_{-1}, X_0, X_1, X_2 \cdots$ a doubly infinite sequence of real-valued random variables indexed by Z.

2. **Vector Auto-Regressions**

A way to summarize the dynamics of macroeconomic data is to make use of vector auto regressions. VAR models have become increasingly popular in recent decades. They are estimated to provide empirical evidence on the response of macroeconomic variables to various exogenous impulses in order to discriminate between alternative theoretical models of the economy. This simple framework provides a systematic way to capture rich dynamics in multiple time series, and the statistical toolkit that came with VARs was easy to use and to interpret (Luetkepohl, 2011). As Sims (1980) and others argued in a series of influential early papers, VARs held out the promise of providing a coherent and credible approach to data description, forecasting, structural inference and policy analysis.

According to Luetkepohl (2011), Sims (1980) advocated vector auto regression (VAR) models as alternatives to the use of multivariate simultaneous equations model extensively for macro econometrics analysis. At that time longer, and more frequently observed macroeconomic time series called models which described the dynamic structures of the variables. VAR models lend themselves for this purpose. They typically treat all variables as a priori endogenous. Thereby they account for Sim's critique that the exogeneity assumptions for some of the variables in simultaneous equations models are ad hoc and often not backed by fully developed theories. Restrictions, including exogeneity of some of the variables, may be imposed on the VAR models based on statistical procedures (Luetkepohl, 2011).

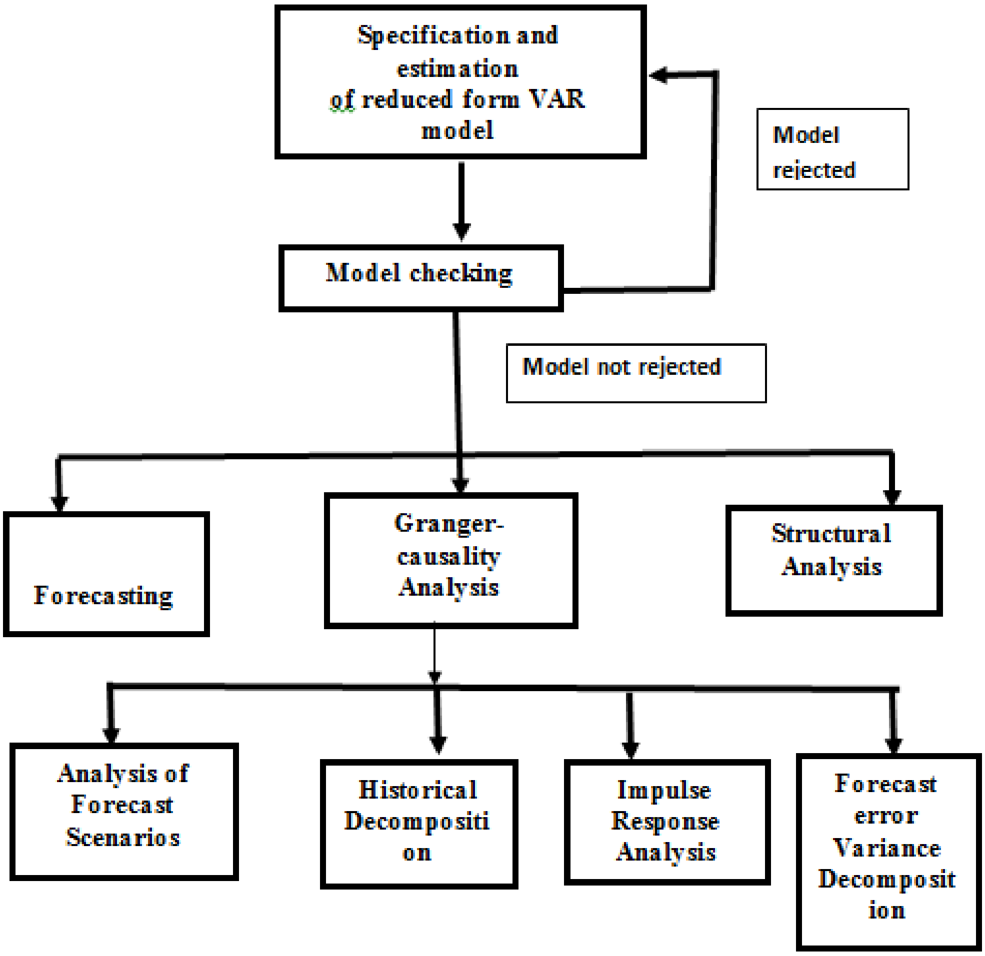
VAR models are natural tools for forecasting. Their set up is such that current values of a set of variables are partly explained by past values of the variables involved. They can also be used for economic analysis, however, because they describe the joint generation mechanism of the variables involved. Structural VAR analysis attempts to investigate structural economic hypotheses with the help of VAR models. Impulse response analyses, forecast, scenarios are the tools which have been proposed for disentangling the relations between the variables in a VAR model (Luetkepohl, 2011).

Traditionally, VAR models are designed for stationary variables without trends. Trending behavior can be captured by including deterministic polynomial terms. In the 1980s the discovery of the importance of stochastic trends in economic variables and the development of the concept of cointegration by Granger (1981), Engle and Granger (1987), Johansen (1995) and others have shown that stochastic trends can also be captured by VAR models. If there are trends in some of the variables it may be desirable to separate the long run relations from the short run dynamics of the generation process of set variables. Vector error correction models offer a convenient framework for separating long-run and short-run components of the data generation process (Luetkepohl, 2011).

The vector auto-regression (VAR) model is one of the most successful, flexible, and easy to use models for the analysis of multivariate time series. It is a natural extension of the univariate autoregressive model to dynamic multivariate time series. The VAR model has proven to be especially useful for describing the dynamic behavior of economic and financial time series and for forecasting. It often provides superior forecasts to those from univariate time series models and elaborate theory-based simultaneous equations models. Forecasts from VAR models are quite flexible because they can be made conditional on the potential future paths of specified variables in the model (Luetkepohl, 2011).

In addition to data description and forecasting, the VAR model is also used for structural inference and policy analysis. In structural analysis, certain assumptions about the causal structure of the data under investigation are imposed, and the resulting causal impacts of unexpected shocks or innovations to specified variables on the variables in the model are summarized. These causal impacts are usually summarized with impulse response functions and forecast error variance decompositions (Luetkepohl, 2011).

Figure 2.1: VAR analysis; Source: EUI Working Paper, 2011



2.3 Empirical Review

This section of the literature review, reviews studies related to the study's research objectives. The following are related studies review in relation to the present study;

Russel, *et al* (2013) analyze life insurance policy surrender activity to determine whether surrender is a function of certain macroeconomic variables and, therefore, highly correlated across policies. Their findings supported the Emergency Fund Hypothesis and the Interest Rate Hypothesis. In addition, they provided evidence that surrenders are significantly related to policy replacement activity, as in Outreville (1990), which they referred to as the Policy Replacement Hypothesis. The significant relationship between policy surrender and macroeconomic factors strongly supports insurer efforts to understand and actively manage disintermediation risk via insurance contract features and investment policy (Russell *et al*, 2013)

Eling and Kochanski (2013) cited in Bauer *et al* (2014) conducted a survey research on life insurance lapsation and review more than 50 theoretical and empirical contributions. On the empirical side, literature has formed a number of hypotheses as to what factors drive lapsation. More precisely, according to the interest rate hypothesis (IRH), policyholders lapse in response to changes in interest rates; the related and more recent policy replacement hypothesis (PRH) presumes that policies are lapsed with the intention to purchase another insurance contract as a replacement; and the so-called emergency fund hypothesis (EFH) contemplates that policyholders predominantly lapse to meet unexpected funding requirements.

Cited in Bauer *et al* (2014) using Cox proportional hazards regressions based on German data, found that those with higher wealth and income are less likely to lapse their policy. Furthermore, different occupation groups show different lapse profiles and (recent) unemployment appears to be a key driver for lapsing. They also found that age does appear to be significant once one control for wealth and policy years. They conclude that there is ample support for the EFH whereas they "rule out"the value-based hypotheses-although this may be due to the specific time period (2005-2011) in which value-driven lapsation may not have been opportune (due to the low/decreasing interest rate environment).

Furthermore, Fang *et al* (2012) in their study, attempt to disentangle the drivers for lapsing a policy. They employed a semi-structural discrete choice model-calibrated to life insurance holdings from the Health and Retirement Study (HRS) data - they conclude that a large portion of policy lapses are driven by idiosyncratic shocks that are largely unrelated to health, income, and bequest motives - especially when individuals are relatively young. However, as the policyholders age, the shocks are more systematic and, initially, are predominantly related to income and health. Over the life-cycle, the bequest motive factor becomes increasingly significant (Bauer *et al*, 2014).

Kim (2005) "using a data collected from a Korean company from the period 1951-1998, employed logit and complementary log-log-model and found that lapse depends on additional exogenous factors beyond interest rates and unemployment rates."

Kiesenbauer (2011) "using regression model found the following as factors that influence lapse; buyer confidence, GDP, company age, distribution, legal form, company size, participation rate factors beyond interest rates, unemployment and company characteristics."Furthermore, Eling and Kiesenbauer (2011) "using the poison model, binomial model and negative binomial model found that product characteristics such as product type or contract age and policy holders characteristics such as age and gender are important lapse drivers."

Kim (2005) "in the quest to model lapse rates used the logit model and found that it closely fit the experience of the data even under extreme financial conditions. Cox and Lin (2006) however, found that Tobit model is better that logit model. They further postulated that the poison model and the negative binomial regression model are more appropriate to model lapse."

The trend of literature reveals that although a number of studies have been done in this area, literature in modeling lapse rates in Africa and especially in Ghana is very scarce. In view of this the present study focused on data form a Ghanaian life insurance company. Furthermore, majority of the studies model lapse rate using regression models and generalized linear models, however, the present study seeks to employ the time series model, specifically, the vector auto-regression model as the main model for modeling the lapse rates. The rational here is to find out which economic variable is most significant in determining lapse rates. This is done by modeling all six (6) variables together using the vector auto-regression approach. This approach was chosen since it has all the elements to answer the research questions. For example, the causality test which is used to test which variable is more significant in determining lapse rates and the impulse response function used to investigate the relationship between the variables in the VAR model.

Chapter 3

Methodology

3.1 Introduction

This chapter discusses the statistical and mathematical methods used in the analysis. After a brief description of the research design, dataset and data analysis, the basics of forecasting is introduced and the foundation for the use of this statistical technique is laid by looking at the definition of a time series and the approaches to time series forecasting. Some basic concepts mostly used in modeling time series data are introduced.

The second part of the chapter looks at the modeling approach employed to analyze the data set, the Vector Auto-regression. Thus, the autoregressive unit root tests and stationarity tests have been treated as a prelude to vector autoregression. The steps involved and the appropriate test statistics used in the analysis are also introduced.

3.2 Research Design

The research design is a quantitative design and time series and deductive in nature. The techniques used in this section are mainly graphical and descriptive statistics. The procedure will enable the researcher gain an insight into the data set, extract important variables and their distributions and detect other anomalies.

3.3 Sources of Data

Life insurers treat lapse data highly confidential not only in the Ghanaian market. Lapse information are therefore publicly available only to a limited extent. The study used data from a life insurance company in Ghana which is referred throughout this study as life insurer Y. The insurer is one of the leading life insurers with expertise in life insurance service delivery to a cross section of Corporate, Small and Medium Enterprises and individuals among others. Since it cost too much time to get the needed authority to obtain the right data, only aggregated lapse rates were available for analysis instead of policy specific data. The data available belonged to the years 2006 to 2014 and were recorded on monthly basis. The data were not modified before analysis. The information on variables such as inflation, unemployment, GDP, stock market return were taken from the Bank of Ghana and Ghana Statistical Service websites within the same period. The interest rates were extracted directly from life insurer Y's database.

3.4 Data Analysis

The extracted data were used as input for R, which was used for the formation of summary statistics and figures. R is a well-known statistical software program which was developed by the R Development Core Team and is known for its predictive analytics, the function which was used in this study.

3.4.1 Stochastic Process:

A stochastic process is a model for a time-dependent random phenomenon. So, just as a single random variable describes a static random phenomenon, a stochastic process is a collection of random variables X_t, one for each time t in some set J. The process is denoted $X_t : t\epsilon J$. The set of values that the random variables X_t are capable of taking is called the state space of the process, S.

3.4.2 Stationary Time Series:

A stochastic process may be classified as either Strict or Weakly stationary. It is said to be strictly stationary, if the joint distributions of $X_{t_1}, X_{t_2}, ..., X_{t_n}$ and $X_{k+t_1} X_{k+t_2}, ...,$ X_{k+t_n} are identical for all $t_1,$ $t_2...,t_n$ and $k + t_1, k + t_2, ...k + t_n$ in J and all integers n. This means that the statistical properties of the process remain unchanged as time elapses. Weak stationarity requires that the mean of the process $m(t) = E(X_t)$ is constant and that the covariance of the process $Cov(X_s, X_t) = E[(X_s - m(s))(X_t - m(t)]$ depends only on the time difference $t - s$. The time difference $t - s$ is referred to as the lag. If a process is strictly stationary then it will also be weakly stationary. A weakly stationary process is not necessarily strictly stationary. Weak stationarity considers only the first two moments of the joint distribution of the set of random variables X_t.

Autoregressive unit root tests

Augmented Dickey-Fuller Test

Test Regression

$$y_t = \beta' D_t + \Phi y_{t-1} + \sum_{j=1}^{p} \mathit{\Psi}_j \Delta_{t-j} + u_t$$

$$\Delta y_t = \beta' D_t + \pi y_{t-1} + \sum_{j=1}^{p} \mathit{\Psi}_j \Delta y_{t-j} + u_t with \pi = \mathit{\Phi} - 1$$

Test Statistic

$$ADF_t : t_{\mathit{\Phi}=1} = \frac{\ddot{\mathit{\Phi}} - 1}{SE(\mathit{\Phi})}$$

$$ADF_t : t_{\pi=0} = \frac{\hat{\pi}}{SE(\pi)}$$

Stationarity Tests

KPSS

$$y_t = \beta' D_t + u_t + u_t, u_t \sim I(0)$$

$$u_t = u_{t-1} + \epsilon_t, \epsilon_t \sim WN * (0, \sigma^2)$$

Hypothesis

$$H_0 : \sigma_\epsilon^2 = 0 \quad \textbf{and} \quad H_1 : \sigma_\epsilon^2 > 0$$

Test Statistic

$$LM = \frac{T^{-2} \sum_{t=1}^{T} S_t^2}{\hat{\lambda}^2}$$

3.4.3 Vector Auto-regressive Models

Vector auto-regression (VAR) is a model used to capture the evolution and the interdependencies between multiple time series, generalizing the univariate AR models. All the variables in a VAR are treated symmetrically by including for each variable an equation explaining its evolution based on its own lags and the lags of all the other variables in the model (Pfaff & Taunus, 2007). According to Sims (1980), in multivariate time series, one of the most flexible and tractable framework for analyzing economic time series is the vector auto-regression (VAR) model. A VAR model describes the evolution of a set of k variables (called endogenous variables) over the same sample period ($t = 1, ..., T$) as a linear function of only their past values. The variables are collected in k × 1 vector y_t, which has the $i^t h$ element, $i_(i, t)$, the observation at time 't'of the $i^t h$ variable. A $p^t h$ order VAR, denoted VAR(p) is

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + ... + A_p y_{t-p} + e_v$$

where the l-periods back observation y_{t-1} is called the l^{th} lag of y, c is a k × 1 vector of constants (intercepts), A_i is a time-invariant k × k matrix and e_t is a k $\times$ 1 vector of error terms satisfying

$E(e_t) = 0$ - every error term has a mean zero

$E(e_t, \acute{e_t}) = \Omega$ - the contemporaneous covariance matrix of error terms is $\omega(ak \times k$ positive - semi-definite matrix)

$E(e_t, e\acute{_{t-k}}) = 0$ for any non-zero k - there is no correlation across time; in particular, no serial correlation in individual error terms.

A p^{th} order VAR is also called a VAR with p lags. The process of choosing the maximum lag p in the VAR model requires special attention because inference is dependent on correctness of the selected lag order.

3.4.4 Model Selection Criteria (Optimal Lag Order Selection)

The standard model selection criteria which are used in this context choose the VAR order which minimizes them over a set of possible orders $m = 0, ..., p_{max}$. The general form of a set of such criteria is

$$C(m) = \log\det\left(\sum_m\right) + cTp(m)$$

Where $\hat{\sum}_m = T^{-1}\sum_{t=1}^{T}\hat{u}_t$ is the residual covariance matrix estimator for a model of order m, $\rho(m)$ is a function of the order m which penalizes large VAR orders and cT is a sequence which may depend on the sample size and identifies the specific criterion.

The term $\log\det\hat{\sum}m$ is a non-increasing function of the order m while $\rho(m)$ increases with m. The lag order is chosen which optimally balances these two forces.

Examples of criteria of this type are Akaike's information criterion (Akaike 1973, 1974),

$$AIC(m) = \log\det(\hat{\sum}_m) + \frac{2}{T}mK^2$$

where

$$cT = \frac{2}{T}$$

the Hannan-Quinn criterion (Hannan and Quinn 1979, Quinn 1980)

$$HQ(m) = \log\det(\hat{\sum}_m) + \frac{2\log\log T}{T}mK^2$$

where

$$cT = \frac{2\log\log T}{T}$$

and the Scharwz(or Rissanen) criterion (Rissanen 1978; Schwarz 1978).

$$SC(m) = \log det(\hat{\sum}_m) + \frac{\log T}{T}mK^2$$

where

$$cT = \frac{\log T}{T}$$

In each case $\rho(m) = mK^2$ is the number of VAR parameters in a model with order m. Denoting by $\hat{p}(AIC), \hat{p}(HQ)$ and $\hat{p}(SC)$ the orders selected by AIC, HQ and SC, respectively, the following relations hold for samples of fixed size $T \geq 16$:

$$\hat{p}(AIC) \leq \hat{p}(HQ)\hat{p}(AIC)$$

Thus, AIC always suggests the largest order, SC chooses the smallest order and HQ is in between (Lütkepohl 2005, Chapters 4 and 8). Of course, this does not preclude the possibility that all three criteria agree in their choice of VAR order. The HQ and SC criteria are both consistent, that is, the order estimated with these criteria converges in probability or almost surely to the true VAR order p under quite general conditions, if p max exceeds the true order. On the other hand, the AIC criterion tends to overestimate the order asymptotically. These results hold for both I(0) and I(1) processes (Paulsen 1984).

3.4.5 Estimation of VAR Models: Conditional MLE

$$f(Y_T, Y_{T-1} \cdots Y_1 | Y_0, Y_{-1} \cdots Y_{-p+1}; v) = \prod_{t=1}^{T} f(Y_t, |Y_{t-1}, Y_{t-2} \cdots Y_{t-p+1}; v) \quad (3.1)$$

$$(Y_t, Y_{t-1}, Y_{t-2} \cdots) \to N(c + \Phi_1 Y_{t-1} + \cdots \Phi_p Y_{t-p}, \Omega) \quad (3.2)$$

$$\prod{}' \equiv [c\Phi_1\Phi_2 \cdots \Phi_p] \qquad n \times (np+1) \quad (3.3)$$

$$X_t \equiv [1Y_{t-1}\Phi Y_{t-2} \cdots Y_{t-p}]' \qquad (np+1) \times 1 \quad (3.4)$$

$$Y_t = \prod{}' X_t + a_t \quad (3.5)$$

$$\ell(v) = \sum_{t=1}^{T} \log f(Y_t | past; v) \quad (3.6)$$

$$= -\frac{Tn}{2}\log(2\pi) + \frac{T}{2}\log|\Omega^{-1}| - \frac{1}{2}\sum_{t=1}^{T}\left[(Y_t - |\prod{}' X_t)'\Omega^{-1}(Y_t - |\prod{}' X_t)\right] \quad (3.7)$$

3.4.6 Stability:

One important characteristic of a VAR(p)-process is its stability. This means that it generates stationary time series with time invariant means, variances and covariance structure, given sufficient starting values. One can check this by evaluating the characteristic polynomial:

$$\det(I_k - A_1 z - \cdots - A_p z^p) \neq 0 \quad for|z| \leq 1 \quad (3.8)$$

If the solution of the above equation has a root for $z = 1$, then either some or all variables in the VAR(p)-process are integrated of order one, i.e., I(1). In practice, the stability of an empirical VAR(p)-process can be analyzed by considering the companion form and calculating the eigenvalues of the coefficient matrix:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} A^1_{11} & A^1_{12} \\ A^1_{21} & A^1_{22} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} A^2_{11} & A^2_{12} \\ A^2_{21} & A^2_{22} \end{bmatrix} \begin{bmatrix} y_{1t-2} \\ y_{2t-2} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \tag{3.9}$$

Or

$$y_{1t} + c - 1 + A^1_{11} y_{1t-1} + A^1_{12} y_{2t-1} + A^2_{11} y_{1t-2} + A^2_{12} y_{2t-2} + e_{1t} \tag{3.10}$$

$$y_{2t} + c - 2 + A^1_{21} y_{1t-1} + A^1_{22} y_{2t-1} + A^2_{21} y_{1t-2} + A^2_{22} y_{2t-2} + e_{2t} \tag{3.11}$$

whereby y_{it} is the value of the i^{th} variable at time t with the dimensions of the stacked vectors ε_t and v_t; (KP $\times$ 1) and the dimension of the matrix A is $(K_p \times K_p)$. If the moduli of the eigenvalues of A are less than one, then the VAR(p)-process is stable (Pfaff and Taunus, 2007, p.6).

3.5 Diagnostic testing(VAR)

3.5.1 Statistical Tests

Serial correlation: Portmanteau Test

The portmanteau test for residual autocorrelation checks the null hypothesis that all residual auto-covariances are zero, that is, $H_0 : E(u_t u'_{t-1}) = 0 (i = 1, 2, \cdots)$. It is tested against the alternative that at least one autocovariance and, hence, one autocorrelation is nonzero. The test statistic is based on the residual auto-covariances and has the form:

$$Q_h = T \sum_{j=1}^{h} tr(\hat{C_j}' \hat{C_0}^{-1} \hat{C_j} \hat{C_0}^{-1}) \tag{3.12}$$

where $\hat{C_j} = T^{-1} \sum_{j=1+1}^{h} \hat{u}_t u'_{t-j}$ and the $\hat{u}_t$'s are the estimated residuals (L$\ddot{u}$tkepohl (2008)).

ARCH Test

A multivariate ARCH model of order q for the residual vector u_t has the form

$$vech(\Sigma_{t|t-1}) = \beta_0 + \beta_1 vech(u_{t-j}u'_{t-1}) + ... + B_q vech(u_{t-q}u'_{t-q}) \tag{3.13}$$

where vech is the column stacking operator for symmetric matrices which stacks the columns from the main diagonal downwards and $\Sigma_{(t|t-1)}$ is the conditional covariance matrix of u_t given $u_{t-1}, u_{t-2}, \cdots$. For this model one may want to test the pair of hypotheses

$H_0 : B_1 = \cdots = B_q = 0$ versus $H_1 : B_1 \neq 0$ for at least one i $\in 1, \cdots, q$

If H_0 is true, there is no ARCH in the residuals. The relevant LM statistic can be obtained by using the auxiliary model

$$vech(\hat{u_t \hat{u}'_t}) = \beta_0 + B_1 vech(\hat{u}_{t-j}\hat{u}'_{t-1}) + \cdots + B_q vech(\hat{u}_{t-q}\hat{u}'_{t-q}) \tag{3.14}$$

and computing

$$LM_{tARCH}(q) = \frac{1}{2}TK(K+1)\left(1 - \frac{2}{K(K+1)}tr\hat{\Omega}\hat{\Omega}_0^{-1}\right) \tag{3.15}$$

where $\hat{\Omega}$ is the residual covariance matrix of the $\frac{1}{2}K(K+1)$ -dimensional regression model with $q > 0$ and Ω_0 is the corresponding matrix for the case $q = 0$ (L$\ddot{u}$tkepohl (2008)).

Normality Test (Residual Analysis):

Multivariate normality test displays the χ^2-statistics associated with the skewness and kurtosis of the residuals which may be used for tests of nonnormality. Most tests for normality are based either on comparing the empirical cumulative distribution with the theoretical normal cumulative distribution (Kolmogorov-Smirnov, Anderson-Darling, Chi-Square) or empirical quantiles with the theoretical normal quantiles (PPCC, Wilk-Shapiro). In contrast, the Jarque-Bera test is based on the sample skewness and sample kurtosis.

Testing the null hypothesis:

H_0: normal distribution, skewness is zero and excess kurtosis is zero; against the alternative hypothesis:

H_1: non-normal distribution.

That is to say, the null hypothesis is of normality, and rejection of the hypothesis (because of a significant p-value) leads to the conclusion that the distribution from which the data came is non-normal.

The test is specifically looking for skewness and kurtosis that is different from that of the normal (it squares the standardized deviations and sums them) and will tend to be significant when skewness and kurtosis deviating from the values at the normal are present.

The Jarque-Bera test statistic is defined for all five (5) variables as:

$$\frac{N}{6}\left(s^2 + \frac{(K-3)^2}{4}\right)$$

with S, K and N denoting the sample skewness, the sample kurtosis, and the sample size, respectively. It turns out that this test statistic can be compared with a χ^2 (chi-square) distribution with 2 degrees of freedom. The null hypothesis of normality is rejected if the calculated test statistic exceeds a critical value from the $\chi^2_{(2)}$ distribution (Jarque and Bera, 1987).

Causality Test

There are many ways of conducting causality test but in respect of this study, the Granger causality test is reviewed. It is used in time series analysis to know whether changes in a variable will have an impact on changes in other variables. VAR models also open up the possibility for analyzing the relation between the variables involved. Analyzing the causal relations is of particular interest. According to Granger (1969), causality in the time series context has become quite popular in applied work. He called a variable y_1 t causal for a variable y_2 t if the information in y_1 t is helpful for improving the forecasts of y_{2t}

According to Granger, causality can be further sub-divided into long-run and short-run causality. This requires the use of error correction models or VECMs, depending on the approach for determining causality.
Long-run causality is determined by the error correction term, whereby if it is significant, then it indicates evidence of long run causality from the explanatory variable to the dependent variable. Short-run causality is determined with a test on the joint significance of the lagged explanatory variables, using an F-test or Wald test.

$$(\frac{y_{1t}}{y_{2t}}) = \sum_{i=1}^{p} \begin{bmatrix} \alpha_{11,i} & \alpha_{12,i} \\ \alpha_{21,i} & \alpha_{22,i} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \mu_t \tag{3.16}$$

for a bivariate VAR(p) setting.

For stationary and integrated processes it turns out thaty_{1t} is not Granger-causal for a variable y_{2t} if and only if

$$\alpha_{21,i} = 0, i = 1, 2...p$$

that is, if it does not appear in the y_{2t} equation of the model.

3.5.2 Prediction / Forecasting

A forecast of y_{t+s} on the basis of $y_t, y_{t+1} \cdots$ is represented as

$$y_{t+s|t} = u + \mathbb{F}_{11}^{(s)}(y_t - u) + \mathbb{F}_{12}^{(s)}(y_{t-1} - u) + \mathbb{F}_{1p}^{(s)}(y_{t-p+1} - u) \tag{3.17}$$

3.5.3 Impulse Response Function

Impulse response analysis is a standard tool for investigating the relations between the variables in a VAR model. If the VAR(p) process y_t is I(0), it has a Wold moving average (MA) representation of the form

$$y_t = \sum_{j-0}^{\infty} \psi_j \mu_{t-j} \tag{3.18}$$

$\psi_k = I_k$ and the $j = \psi_j{}'s(1, 2, \cdots)$ are $(K \times K)$ coefficient matrices and deterministic terms are ignored to simplify the exposition. The marginal response $y_{n,t+j}$ to a unit change in, y_{mt}, holding constant all past values of y_t is given by $(n, m)^{th}$ elements of the matrices ψ, viewed as a function of j. More precisely, the elements of ψ_j represent responses to μ_t innovations, that is, to forecast errors. Therefore these quantities are sometimes called forecast error impulse responses (L$\ddot{u}$tkepohl 2005, Section 2.3.2). Since $\psi_j \longrightarrow 0$ as $j \longrightarrow \infty$ for stationary processes, the effect of an impulse vanishes over time. In other words, it is transitory. Any nonsingular matrix P with the property that $PP' = \sum_\mu$ can be used to define orthogonalized shocks as $\epsilon_t = P^{-1}\ u_t$. Clearly, these shocks have the property that $\epsilon_t \sim (0, I_k)$ and they are contemporaneously uncorrelated. The responses to such shocks are given by the coefficients of the MA representation

$$y_t = \sum_{j=1}^{\infty} PP^{-1} \mu_{t-j} = \sum_{j=1}^{\infty} \psi_j \epsilon_{t-j} \tag{3.19}$$

Forecast Error Variance Decomposition

$$G(L)_{z_t} = \mu_t$$

the VMA(vector moving average) representation of our VAR is:

$$z_t = \Gamma_0 \mu_t + \Gamma_1 \mu_{t-1} + ...$$

and the error in forecasting z_t in the future is, for each horizon s:

$$z_{t+s} - E(z_{t+s}) = \Gamma_0 \mu_{t+s} + \Gamma_1 \mu_{t+s-1} + ...\Gamma_{s-1} \mu_{t+1}$$

from which the variance of the forecasting error is:

$$var(z_{t+s} - E_t z_{t+s}) = \Gamma_0 \sum_{\mu} \Gamma_0^1 + \Gamma_1 \sum_{\mu} \Gamma_1^1 + ... + \Gamma_{s-1} \sum_{\mu} \Gamma_{s-1}^1$$

on the basis of this formula, we can compute the share of the total variance of the forecast error for each variable attributable to the variance of each structural shock.

Chapter 4

Analysis and Findings

4.1 Introduction

This chapter presents results and analysis of the study. The chapter consists of two main sections. The first section presents the trend analysis of the six (6) main variables of interest; lapse rates, unemployment rates, inflation rates, stock market return, GDP and interest rates. The second section deals with the presentation and analysis of summary statistics.

4.2 Model Selection Criteria

Table 4.1: Optimal Lag Order Selection

VAR Order	Log-likelihood	p(LR)	AIC	BIC	HQC
1	-1109.72478		24.768275	25.912029*	25.230090
2	-1064.03224	0.00000	24.559833	26.683949	25.417490
3	-1008.90476	0.00000	24.148489	27.252966	25.401988
4	-957.16531	0.00000	23.810007	27.894845	25.459347
5	-892.79606	0.00000	23.199915	28.265114	25.245098
6	-826.09306	0.00000	22.539636	28.585196	24.980660
7	-740.20462	0.00000	21.466766	28.492687	24.303632*
8	-690.79008	0.00000	21.178281*	29.184564	24.410989

This is a VAR system with lag order 8. OLS estimates, observations 2006:09-2014:05 (T=93) and log-likelihood = -690.79008. For the system as a whoole, the null hypothesis (the longest lag) is 7 and the alternative hypothesis (the longest lag) is 8. Likelihood ratio test implies, Chi-square (36) = 98.8291 [0.0000]. The asterisks below indicate the best (that is, minimized) values of the respective information criteria, AIC = Akaike criterion, BIC = Schwarz Bayesian criterion and HQC = Hannan-Quinn criterion.

From the statistics the Akaike information criterion was the best criteria used in choosing the optimal lag order with a value of 21.1783 at lag 8, which is lower than the best estimates of the other selection criteria's. A p-value of 0.0000 implies that the alternative hypothesis is true i.e. the longest lag is 8.

4.3 Estimation of VAR Models: Conditional MLE

Refer to Appendix A, Table 5.4

From the summary statistics, since the p-values of inflation, stock market rate and interest rate are below or slightly above the significant level of 5%, we reject the null hypothesis and conclude that the coefficients of inflation (4.92993), stock market return (-13.2187) and interest rates (-11.07880) are significant in determining lapse rates.

The coefficient of inflation implies that an increase in inflation will increase lapse rates. Conversely, the coefficients of stock market return and interest rates imply that a decrease stock market return and a decrease in interest rates will decrease lapse rates.

The F-test of zero restrictions also indicates that among all the variables, the interest rate (for all lags) is the most significant in determining lapse rates with a p-value of 0.0107.

VAR model for average lapse rate is significant since it has a p-value of 0.016167 compared to a significant level of 5%.

Table 4.2: Final Forecast Model

	Coefficient	Std. Error	t-ratio	p-value	
const	7.83532	1.66695	4.7004	0.00001	***
$INFL_1$	0.0483466	0.0261255	1.8506	0.06858	*
$INFL_2$	0.0236226	0.0289925	0.8148	0.41804	
$INFL_3$	0.0372398	0.0215486	1.7282	0.08850	*
$INFL_4$	0.0528083	0.0194398	2.7165	0.00836	***
$INFL_5$	-0.0136803	0.0309476	-0.4420	0.65986	
$INFL_6$	0.0208792	0.0295468	0.7066	0.48220	
$INFL_7$	0.0478825	0.0195736	2.4463	0.01703	**
$INFL_8$	-0.0136587	0.0214247	-0.6375	0.52593	
$STKMT_1$	0.447737	0.187444	2.3886	0.01969	**
$STKMT_2$	0.142544	0.152676	0.9336	0.35379	
$STKMT_3$	-0.379701	0.18755	-2.0245	0.04684	**
$STKMT_4$	0.111687	0.205196	0.5443	0.58802	
$STKMT_5$	-0.0734312	0.139231	-0.5274	0.59963	
$STKMT_6$	-0.17186	0.154491	-1.1124	0.26987	
$STKMT_7$	-0.198306	0.141107	-1.4054	0.16447	
$STKMT_8$	-0.12324	0.13649	-0.9029	0.36975	
INT_1	-0.0629868	0.0458146	-1.3748	0.17370	***
INT_2	0.00680454	0.0512813	0.1327	0.89483	
INT_3	-0.101415	0.0640049	-1.5845	0.11772	***
INT_4	-0.0720572	0.0811095	-0.8884	0.37746	***
INT_5	0.00715901	0.069383	0.1032	0.91812	**
INT_6	0.0207542	0.0713355	0.2909	0.77198	
INT_7	0.0742374	0.0707632	1.0491	0.29785	**
INT_8	-0.0145191	0.048139	-0.3016	0.76387	***

Mean dependent var	8.673763	S.D. dependent var	1.506772
Sum squared resid	28.34850	S.E. of regression	0.645670
R-squared	0.864279	Adjusted R-squared	0.816378
F(24, 68)	39.38509	P-value(F)	6.25e-31
ρ	-0.006597	Durbin-Watson	1.943604

All lags of INFL	F(8, 68) = 6.2156 [0.0000]
All lags of STKMT	F(8, 68) = 9.3979 [0.0000]
All lags of INT	F(8, 68) = 1.6509 [0.1268]
All vars, lag 8	F(3, 68) = 0.88677 [0.4525]

From the summary statistics, since the p-values of inflation, stock market rate and interest rate are below or slightly above the significant level of 5%, we reject the null hypothesis and conclude that the coefficients of inflation, stock market return and interest rates are significant in determining lapse rates.

The coefficient of inflation implies that an increase in inflation will increase lapse rates. Conversely, the coefficients of stock market return and interest rates imply that a decrease stock market return and a decrease in interest rates will decrease lapse rates.

The F-test of zero restrictions also indicates that among all the variables, the interest rate (for all lags) is the most significant in determining lapse rates with a p-value of 0.1268.

VAR model for average lapse rate is significant since it has a p-value of 0.016177 compared to a significant level of 5%.

4.4 Stability

Table 4.3: Eigen analysis of the Correlation Matrix

Component	Eigenvalue	Proportion	Cumulative
1	0.8459	0.4229	0.42290
2	0.4294	0.2147	0.6376
3	0.3153	0.1577	0.7953
4	0.2748	0.1374	0.9327
5	0.0798	0.0399	0.9726
6	0.0548	0.0274	1.0000

Table 4.4: Eigenvectors (component loadings)

	PC1	PC2	PC3	PC4	PC5	PC6
AVLP	0.116	0.333	-0.932	0.002	-0.068	-0.052
INFL	0.592	0.001	0.027	0.082	0.035	0.801
UNEMPL	-0.556	0.186	0.022	0.214	-0.661	0.416
STKMT	0.566	-0.005	0.142	0.195	-0.671	-0.414
INT	-0.030	-0.702	-0.239	-0.587	-0.305	0.104
GDP	0.077	0.601	0.230	-0.752	-0.120	0.017

The determinant of covariance matrix = 0.11399945. Since the covariance matrix is symmetric, its eigenvalues are all real and positive and the eigenvectors that belong to distinct eigenvalues are orthogonal. As a consequence, the determinant of the covariance matrix is positive.

Also the VAR(8) process is stable since all the moduli of the eigenvalues of the matrix are less than one.

Diagnostic testing (VAR)

The serial correlation of the Portmanteau Test produced the following results: Portmanteau test: LB (23) = 1005.15, df = 540 [0.0000]

With a p-value of 0.0000 against 5% significant level we reject the null hypothesis and conclude that there exist residual autovariances

Chapter 5

Conclusion and Recommendations

5.1 Introduction

In this chapter, the results from the study are compared to the research objectives to assess if they have been met. Recommendations are also made for further research.

5.2 Summary of Results

From the statistics the Akaike information criterion was the best criteria used in choosing the optimal lag order with a value of 21.1783 at lag 8, which is lower than the best estimates of the other selection criteria's. A p-value of 0.0000 implies that the alternative hypothesis is true i.e. the longest lag is 8.

Also from the summary statistics, since the p-values of inflation, stock market rate and interest rate are below or slightly above the significant level of 5%, we reject the null hypothesis and conclude that the coefficients of inflation (4.92993), stock market return (-13.2187) and interest rates (-11.07880) are significant in determining lapse rates.

The coefficient of inflation implies that an increase in inflation will increase lapse rates. Conversely, the coefficients of stock market return and interest rates imply that a decrease stock market return and a decrease in interest rates will decrease lapse rates.

The F-test of zero restrictions also indicates that among all the variables, the interest rate (for all lags) is the most significant in determining lapse rates with a p-value of 0.0107. VAR model for average lapse rate is significant since it has a p-value of 0.016167 compared to a significant level of 5%.

Furthermore, since the covariance matrix is symmetric, its eigenvalues are all real and positive and the eigenvectors that belong to distinct eigenvalues are orthogonal. As a consequence, the determinant of the covariance matrix is positive.

Also the VAR(8) process is stable since all the moduli of the eigenvalues of the matrix are less than one.

Finally the portmanteau test indicates a p-value of 0.0000 against 5% significant level. Thus we reject the null hypothesis and conclude that there exist residual autovariances.

5.3 Conclusion

This thesis started with research questions which were to: identify are those variables which are seen as significant drivers of lapse rate, identify the variable which is most significant in determining lapse rates and to determine which model is most fit for forecasting lapse rates.

To answer the first and second questions, it had to be examined which variables could have exercised influence on the lapse rates using data on all contracts. To answer the second part of the question, examination had to be done to discover which variables were available as well as applicable to this data set. With the theory in scientific papers as a basis, complemented by company experience and logic, a list of ten (10) possible explanatory variables had been composed.

Literature review on predictive modeling of lapse rates in the insurance industry led to the choice for predicting with a VAR Model. Application of the vector auto-regression model provided formulae for forecasting future lapse rates.

A total of six (6) variables - including inflation, unemployment and lagged lapse rates - have been analyzed on their relationship with the average lapse rate. Of these variables inflation, stock market return and interest rate have proven to be valuable for modeling with interest rate been the most significant.

From the summary table the best model for forecasting lapse rates includes the variables inflation, stock market return and interest rate. Thus, existing forecast models which considers only interest rate as the only variable in determining lapse rates can be improved.

Although there is room for improvement of the model, it provides understandable results and is more accurate than the prediction of future lapse rates by assuming the mean of the lapse data to be a good predictor.

5.4 Recommendation

The main limitation of this research was the small amount of data and their level of detail. Since the results of this research are promising it is recommended to extend the research to other parts of the country.

To increase the statistical strength and accuracy of the inferences that can be made it is recommended to examine the lapse rates at a policyholder level. This will provide the option to include more variables in the modeling in order to explain individual behavior. It should also be verified whether the data are pure and not contaminated, since some data is expected to be contaminated due to policy mutations (Achmea, 2010b).

The promising results do stimulate monitoring of the model to check its reliability and accuracy. One of the ways to analyze the model is to compare the results with existing GLMs.

Appendix A

Table 5.1: Augmented Dickey-Fuller and Normality Test

ADF and Normality Test

Augmented Dickey-Fuller(AD) Test	Normality Test (Residual Analysis)
Augmented Dickey-Fuller (GLS) test for AVLP including 12 lags of (1-L)AVLP (max was 12) sample size 88 unit-root null hypothesis: a = 1 test with constant model: (1-L)y = b0 + (a-1)*y(-1) + ... + e 1st-order autocorrelation coeff. for e: 0.001 lagged differences: F(12, 75) = 4.402 [0.0000] estimated value of (a - 1): -0.490665 test statistic: tau = -1.65761 asymptotic p-value 0.09213	Test for normality of AVLP: Doornik-Hansen test = 188.385, with p-value 1.23801e-041 Shapiro-Wilk W = 0.716615, with p-value 1.14197e-012 Lilliefors test = 0.231623, with p-value ~= 0 Jarque-Bera test = 142.229, with p-value 1.30399e-031

Table 5.2: Equation 1: Var results for average lapse rate

	Coefficient	Std. Error	t-ratio	p-value
Const	625.313	272.407	2.2955	0.02653
$AVLP_1$	0.05429	0.16679	0.3255	0.74635
$AVLP_2$	0.210735	0.161245	1.3069	0.19803
$AVLP_3$	0.219594	0.226291	0.9704	0.33715
$AVLP_4$	0.101873	0.204473	0.4982	0.62081
$AVLP_5$	-0.216323	0.175328	-1.2338	0.22382
$AVLP_6$	-0.078229	0.20079	-0.3896	0.69871
$AVLP_7$	-0.15316	0.174597	-0.8772	0.38513
$AVLP_8$	-0.123047	0.148233	-0.8301	0.41097
$INFL_1$	4.92993	1.9059	2.5867	0.01307
$INFL_2$	1.75384	2.27955	0.7694	0.44578
$INFL_3$	-2.4663	2.07292	-1.1898	0.24052
$INFL_4$	-0.347546	2.64668	-0.1313	0.89613
$INFL_5$	-1.19719	2.41948	-0.4948	0.62319
$INFL_6$	0.86002	2.13221	0.4033	0.68865
$INFL_7$	1.00357	1.99908	0.5020	0.61816
$INFL_8$	0.00649368	1.89112	0.0034	0.99728
$UNEMPL_1$	11.0676	24.4665	0.4524	0.65323
$UNEMPL_2$	7.53267	19.3086	0.3901	0.69833
$UNEMPL_3$	-21.1592	17.8171	-1.1876	0.24137
$UNEMPL_4$	7.97484	18.8852	0.4223	0.67488
$UNEMPL_5$	-4.14535	18.7224	-0.2214	0.82580
$UNEMPL_6$	-27.9422	17.9813	-1.5540	0.12736
$UNEMPL_7$	-7.73842	17.5934	-0.4398	0.66220
$UNEMPL_8$	-21.275	20.9392	-1.0160	0.31516
$STKMT_1$	-13.2187	6.39257	-2.0678	0.04457
$STKMT_2$	-11.9578	6.33367	-1.8880	0.06564
$STKMT_3$	0.326938	7.90628	0.0414	0.96720
$STKMT_4$	-1.46795	7.84347	-0.1872	0.85240
$STKMT_5$	4.75479	7.11465	0.6683	0.50743
$STKMT_6$	-2.29154	8.35824	-0.2742	0.78524
$STKMT_7$	-10.8819	7.8551	-1.3853	0.17294
$STKMT_8$	-5.54426	7.50939	-0.7383	0.46425
INT_1	-3.68414	4.67189	-0.7886	0.43459
INT_2	0.269765	4.04558	0.0667	0.94714
INT_3	1.95036	3.62085	0.5386	0.59285
INT_4	-2.38693	4.14159	-0.5763	0.56733
INT_5	3.56991	3.85475	0.9261	0.35944
INT_6	-1.8486	3.77735	-0.4894	0.62700
INT_7	-11.0788	2.80535	-3.9492	0.00028
INT_8	1.30791	3.05398	0.4283	0.67055
GDP_1	1.00787	1.1587	0.8698	0.38911
GDP_2	-0.614742	1.18603	-0.5183	0.60683
GDP_3	-0.269388	1.27261	-0.2117	0.83333
GDP_4	0.730291	1.41856	0.5148	0.60926
GDP_5	0.269501	1.40004	0.1925	0.84824
GDP_6	-0.850572	1.30771	-0.6504	0.51880
GDP_7	-1.22833	1.49041	-0.8242	0.41430
GDP_8	-1.07741	1.19863	-0.8989	0.37362

Mean dependent var	20.90344	S.D. dependent var	23.78087
Sum squared resid	16899.30	S.E. of regression	19.59783
R-squared	0.675193	Adjusted R-squared	0.320858
F(48, 44)	1.905520	P-value(F)	0.016167
rho	-0.020167	Durbin-Watson	1.986591

All lags of AVLP	F(8, 44) = 1.1239 [0.3664]
All lags of INFL	F(8, 44) = 1.4907 [0.1882]
All lags of UNEMPL	F(8, 44) = 1.0152 [0.4388]
All lags of STKMT	F(8, 44) = 1.61 [0.1496]
All lags of INT	F(8, 44) = 2.9104 [0.0107]
All lags of GDP	F(8, 44) = 0.60477 [0.7688]
All vars, lag 8	F(6, 44) = 0.73815 [0.6217]

Appendix B

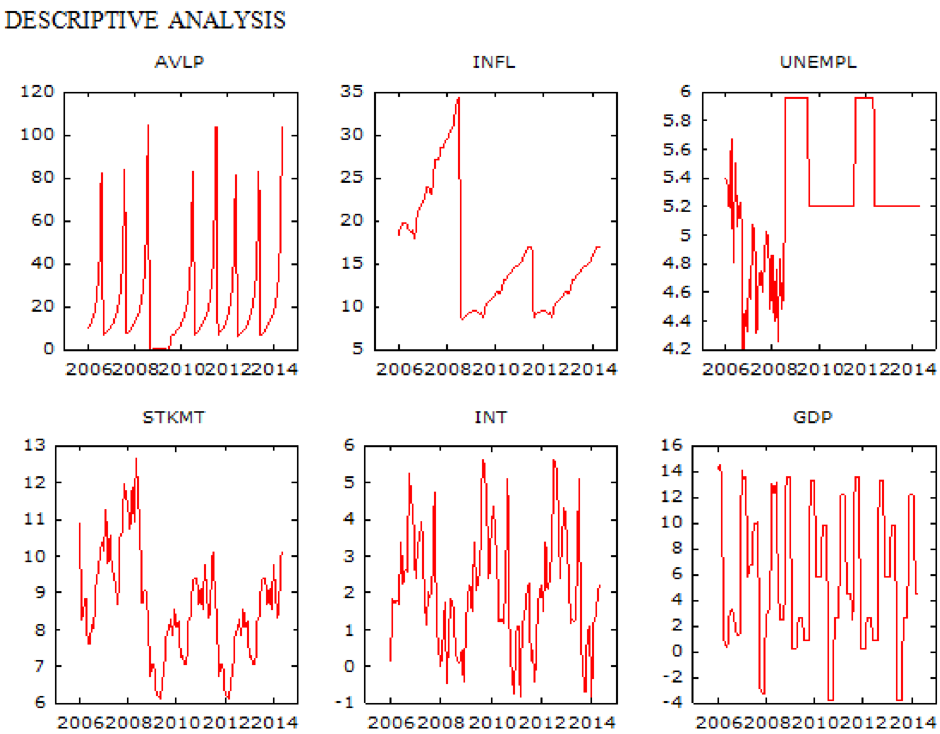


Figure 5.1: Time Series Analysis

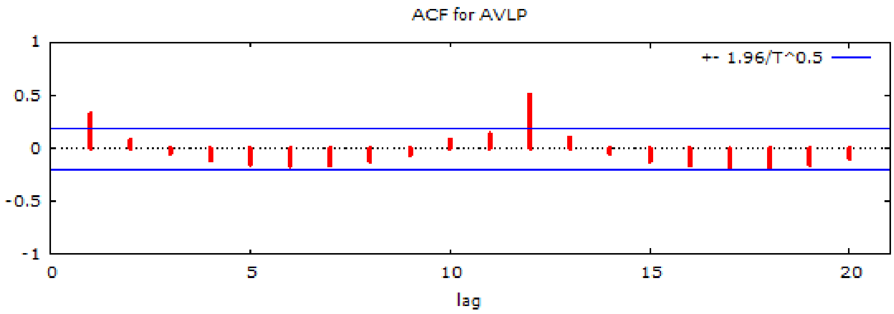
AVLP	INFL	UNEMPL	STKMT	INT	GDP	
1.0000	0.1440	-0.0966	0.0550	-0.0958	0.0778	AVLP
	1.0000	-0.7699	0.8068	-0.0799	0.0719	INFL
		1.0000	-0.6849	-0.1795	-0.0717	UNEMPL
			1.0000	-0.1229	0.0348	STKMT
			1.0000	-0.1229	0.0348	STKMT
				1.0000	-0.2291	INT
					1.0000	GDP

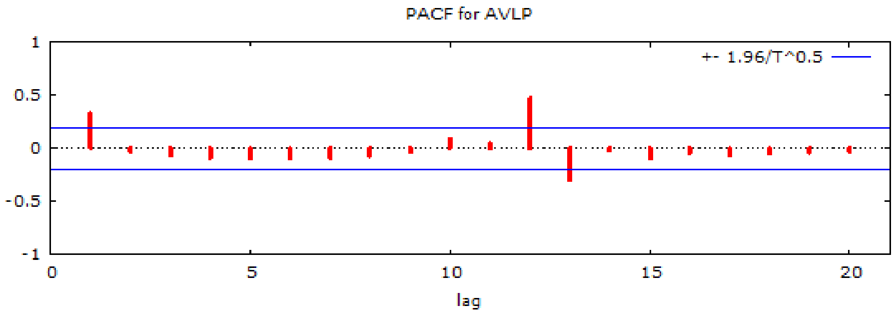
Table 5.3: Summary Statistics

Variable	Mean	Median	Minimum	Maximum
AVLP	21.4275	12.8200	0.100000	105.060
INFL	16.0064	14.5000	8.60000	34.3500
UNEMPL	5.25139	5.20000	4.20000	5.96000
STKMT	8.66010	8.56000	6.15000	12.6900
INT	2.03772	1.97000	-0.820000	5.63000
GDP	5.66535	4.50000	-3.80000	14.4800
Variable	Std. Dev.	C.V.	Skewness	Ex. kurtosis
AVLP	23.7753	1.10957	2.11455	3.98896
INFL	6.81784	0.425944	1.02222	0.120487
UNEMPL	0.460137	0.0876220	0.0340246	-0.309763
STKMT	1.47269	0.170055	0.421127	-0.218091
INT	1.59830	0.784355	0.407470	-0.364234
GDP	5.59240	0.987124	0.150202	-1.20963

Table 5.4: Correlation coefficients

AVLP	INFL	UNEMPL	STKMT	INT	GDP	
1.000	0.1440	-0.0966	0.0550	-0.0958	0.0778	AVLP
	1.0000	-0.7699	0.8068	-0.0799	0.0719	INFL
		1.0000	-0.6849	-0.1795	-0.0717	UNEMPL
			1.0000	-0.1229	0.0348	STKMT
				1.0000	-0.2291	INT
					1.0000	GDP





Similar charts could be a shown for the other variables

References

Adhikari, R., & Agrawal, R. K. (2013). An Introductory Study on Time Series Modeling and Forecasting. arXiv preprint arXiv:1302.6613.

Akaike, H. (1974). A new look at the statistical model identification.... Page(s):716 - 723; ISSN : 0018-9286; DOI: 10.1109/TAC.1974.1100705 ... Issue Date : Dec1974

Anderson, C.J. (2010). Introduction to Generalized Linear Models (GLMs). Retrieved October 05, 2011 from http://www.ed.uiuc.edu/courses/edpsy490AT/lectures/4glm1-ha- online.pdf.

Bauer, D., Gao, J., Moenig, T., Ulm, R. E., & Zhu, N. (2015). Policyholder Exercise Behavior in Life Insurance: The State of Affairs. Society of Actuaries' (SOA) Center of Actuarial Excellence (CAE) Research Grant.

Bozdogan, H. (2000). Akaike's information criterion and recent developments in information complexity. Journal of mathematical psychology, 44(1), 62-91.

Carson, J. & Dumm, R. (2000). The relationship of insurer characteristics and life insurance surrender values. Journal of Financial Service Professionals, 54 (5), 86-90.

Cochrane, H. J. (1997). Time Series for Macroeconomics and Finance. Graduate School of Business, University of Chicago.

Cox S. H. & Lin Y. (2006). Annuity lapse rate modeling: Tobit or not tobit?. Retrieved October 30, 2011 http://www.soa.org/files/pdf/Cox%20Linn%20paper%2011-15-06.pdf.

Cummins, J. & M. Weiss (2009). Convergence of Insurance and Financial Markets. Hybrid and Securitized Risk-Transfer Solutions, 76(3), 493-545.

Eling, M., & Kiesenbauer, D. (2013). What Policy Features Determine Life Insurance Lapse? An Analysis of the German Market, Journal of Risk and Insurance, DOI: 10.1111/j.1539-6975.2012.01504.x.

Eling, M., & Kochanski, D. (2012). Research on Lapse in Life Insurance: What Has Been Done and What Needs to Be Done? University of St Gallen Working Papers on Risk Management and Insurance.

Fang, H. & Kung, E. (2012). Why Do Life Insurance Policyholders Lapse? The Roles of Income, Health and Bequest Motive Shocks, Working Paper. The University of Pennsylvania.

Fier, G. S. & Liebebberg, P. A. (2012). Life Insurance Lapse Behavior.

Gatzert, N., Hoermann, G., & Schmeiser, H. (2009). The Impact of the Secondary Market on Life Insurers' Surrender Profits," Journal of Risk and Insurance, 76(4), 887-908.

Guszcza, J. (2005). CART from A to B. Retrieved October 30, 2011 from http://www.casact.org/education/specsem/f2005/handouts/cart.ppt.

Hardin J., & Hilbe, J. (2001). Generalized Linear Models and Extensions. Stata Press.

Inderst, R., & Sirak, A.S. (2014) Income and Unemployment Effects on Life

Insurance Lapse, Working Paper, University of Frankfurt.

Jarque, C. M. & Bera, K. A. (1987). A test for normality of observations and regression residuals. International Statistical Review, 55, 163-172.

Kaas R., Goovaerts M., Dhaene J., Denuit M. (2001). Modern actuarial risk theory. Boston: Kluwer Academic Publishers.

Kestel, S. (2013). Lecture Notes on the Time Series Analysis. Department of Empirical Research and Econometrics, Albert-Ludwigs University Freiburg

Loisel, S., Maume-Deschamps, V. & Milhaud, X. (2010). Surrender triggers in life insurance: Classification and risk predictions. Working paper.

Luetkepohl. H. (2011). Vector autoregressions. EUI Working papers ECO 2011/30.

Mauer, L. & Holden, N. (2007). Determinants of the lapse rate in life insurance operating companies. Review of Business Publisher. St. John's University College of Business Administration, Vol. 28 No. 1. ISSN: 0034-6454

McCullagh, P., & Nelder, J.A. (1989). Generalized Linear Models, (2nd Edition). CRC Press.

Michories, Z. C. (2011). Modelling Lapse rate. Investigating the variables that drive lapse rate. Department of Finance and Accounting, University of Twente.

Mojekwu N. J. (2011). Study of Modes of Exit of Life - Insurance Policy Holders in Nigeria: Trends and Patterns. International Business Research, 4 (3), 182-187 doi:10.5539/ibr.v4n3p182

Outreville, J. F. 1990. Whole-life Insurance Lapse Rates and the Emergency Fund Hypothesis. Insurance: Mathematics and Economics, 9, 249-255.

Pfaff, B. (2007). Using the vars package. Available at http://ftp.uni bayreuth.de/math/statlib/

Quinn, B.G., (1980). Order determination for a multivariate autoregression. Journal of the Royal Statistical Society Series B, 42, 182-185.

Rocco, R., Cerchiara, R. R., Edwards, M. & Gambini, A. (n.d.). Generalized Linear Models In Life Insurance: Decrements And Risk Factor Analysis Under Solvency II.

Russell, T. D., Fier, G. S., Carson, M. J., & Dumm, E. R. (2013). An Empirical Analysis of Life Insurance Policy Surrender Activity. Journal of Insurance Issues, 36(1), 35-57. Western Risk and Insurance Association. Stable URL: http://www.jstor.org/stable/41946336 Accessed: 25-06-2015 10:46 UTC

Schwarz, G., (1978). Estimating the dimension of a model. The Annals of Statistics 6, 461-464

Stock, H. J., & Watson, W. M. (2001). Vector autoregressions

The Geneva Association (August, 2012). Surrenders in the Life Insurance Industry and their Impact on Liquidity.

The Messenger (2011). Lapse rates in a principle base world. Archival material Transamerica Reinsurance Risk Management Newsletter.

Truyens, P. (1996). Life- Insurance in Africa. Past Products and Future Strategies.African Insurance Organisation,Mombasa Kenya, Novermber 27-28, pp. 33-37.

Verguts, T., & Storms, G. (2004). Assessing the informational value of parameter estimates in cognitive models. Behavior Research Methods, Instruments, & Computers, 36 (1), 1-10.

Watson Wyatt (2007): A Practitioner's Guide to Generalized Linear Models

Zhang, G. P. (2007). A neural network ensemble method with jittered training data for time series forecasting. Information Sciences 177, 5329-5346.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50,159-175.