Master in Actuarial and Financial Sciences

MASTER THESIS

Title: On the Drivers of Lapse Rates in Life Insurance

Author: Mohnish Vasudev Raheja Bajaj

Advisor: Dr. Antonio Alegre Escolano

Academic Year: 2016-17

UNIVERSITAT DE BARCELONA
Facultat d'Economia i Empresa

Màster de Ciències Actuarials i Financeres

Faculty of Economics and Business

Universitat de Barcelona

Master thesis

Master in Actuarial and Financial Sciences

On the Drivers of Lapse Rates in Life Insurance

Author:

Mohnish Vasudev Raheja Bajaj

Advisor:

Dr. Antonio Alegre Escolano

The contents of this document are the sole responsibility of the author, who declares that he has not committed plagiarism and that all references to other authors have been expressed in the text.

ABSTRACT: Lapse risk is the largest non-financial risk which life insurance companies are faced with. Lapse refers to the contractual disruption of an insurance policy before its maturity. The intention of this paper is to gain an insight into the behavior of lapse rates by identifying the most significant variables which drive lapse rates. The study consists of two approaches: a theoretical approach where current literature on lapses is reviewed and an empirical approach where real lapse data is modelled with generalized linear models. Findings include inflation, external rates of return, internal rates of return, and lagged lapse rates as the main drivers of lapse rates.

Key words: Lapse, Surrender, Generalized Linear Models, Solvency II, Technical Provisions, Life Insurance

Table of Contents

List of Tables

List of Figures

1. Introduction

Improving risk management has been a top priority for insurance companies, especially after the global financial turmoil which took place in the year 2008. Due to their complexity, insurance products are subject to many different risks[1] which have to be managed optimally. The increasing importance of risk management has been addressed by European regulators with the implementation of Solvency II. Solvency II is a regulatory directive that has been implemented in the year 2016 which establishes a series of quantitative and qualitative requirements that European insurance companies must follow in order to protect policyholders and guarantee the stability of the financial system as a whole. Deriving from these qualitative requirements, insurers are required to provision a part of their own funds in order to address the uncertainty that they face due to the risks they assume.

In this paper the concept of lapse risk is addressed. According to EIOPA (2011, a), lapse risk is one of the three largest risks faced by life insurance companies, along with market and credit risk. Therefore, it is the largest "insurance risk" (non-financial) which companies are faced with.

In the traditional sense, lapses consist of the termination of an insurance contract because the policyholder has failed to comply with his/her obligations (e.g. premium payment). However, a broader definition has been acquired in most academic literature as the term lapse includes the "traditional" definition of lapses as well as the surrender option. The surrender option is an option embedded in many life insurance products which allows policyholders to voluntarily terminate their contracts before their maturity in exchange for a cash-value payment.

In recent years, lapses have gained importance due to the financial turmoil, as policyholders have begun to exercise their surrender options and many policies have lapsed in the traditional sense, due to policyholder's going through times of financial distress and not being able to comply with their contractual obligations.

The uncertainty of lapses is what poses a great risk for insurance companies, as they are unknown events which can have a direct impact on the financial state of the insurer. Lapse risk is aggravated in the current economic context: post-crisis with very low and even negative interest rates. Unexpected lapses can endanger the financial stability of any insurer as it can force the sale of assets with long durations which could cause losses. Likewise, early unexpected lapses can lead the insurer to suffer from heavy losses if the policy's acquisition costs are not covered in time. This scenario will usually lead to an increased premium for future policyholders, both for new policyholders and for renewals of current policyholders. A complete breakdown of the importance of understanding lapse rates will be provided in Section 3.1.1.

Due to the large risk which lapses suppose for insurance companies, it is of vital importance to understand the drivers and motivation behind lapses, as this would allow for a better predictive capacity and an estimation of more accurate assumptions when dealing with actuarial calculations (e.g. pricing, underwriting, reserving). This would allow for a better management of lapse risk. Specifically, understanding lapse dynamics can lead to an improved management of lapse risk which can consist in modifying Asset Liability Management (ALM) strategies, underwriting strategies, and product design. Moreover, understanding lapses is not

[1] More information regarding these risks has been provided in Section 2.3.

only important for insurance companies but is also important for regulatory bodies as it will allow for the establishment of more accurate capital requirements.

The main objective of this paper is to identify the main variables which drive lapse rates, in order to gain an insight into the behavior of lapse rates and its inherent dynamics.

In order to achieve this objective, available literature on lapse rates and policyholder behavior will be reviewed, detailing the main concepts and conclusions derived from previous studies. Subsequently, an empirical study on real lapse data of a company which operates in the Spanish market will be carried out. Lapse rates will be modelled with the help of generalized linear models. Possible explanatory variables which are expected to have an influence on lapses will be included in the study and will be detailed in Section 6.1.

Lapse Rates will be modelled separately, differentiating between five different product types: Individual Savings, Group Savings, Individual Protection, Group Protection, and Unit-Linked products. The rationale behind this division is that lapse rates differ vastly between product types, as each product covers different policyholder needs and therefore might lead to distinct policyholder behavior. This division will allow for testing between different product types, as these products have different inherent features which might lead to the identification of different lapse drivers. Additionally, behavioral differences between groups and individuals can also be tested, as it is assumed that group policies will be handled by professional investors who tend to act more rationally from a financial point of view.

In this paper a top to down approach will be followed, commencing with a broader view of the question and progressively going into more detail. Concretely, the study is presented as follows:

- In Section 2 the concept and origin of insurance are explained. Subsequently, life insurance is defined and the different types of products within life insurance are presented, along with the risks which accompany these products.
- Section 3 covers the concept of lapse risk, starting with the definition of lapses and the importance of understanding lapse behavior, detailing the possible impacts which lapses can have on a life insurance company and preventive measures which can be put in place to reduce lapse risk. Furthermore, Solvency II is presented, detailing the balance sheet and capital requirements and the impact that lapse risk has over these elements.
- Section 4 includes an analysis of policyholder behavior, detailing the possible drivers of lapses which have been found in literature along with a brief explanation of behavioral economics in the context of lapses. The concept of dynamic policyholder behavior is introduced, concretely detailing the concept of dynamic lapses.
- In Section 5 a brief theoretical explanation of generalized linear models is provided.
- Section 6 details the data which has been used in the study as well as the complete method and procedure which has been followed in order to estimate predictive models for each product category.
- In Section 7 the obtained results will be discussed and analyzed.
- Section 8 will include the conclusions of the present study, along with the possible limitations which have been encountered and motives for future research.

2. Insurance

This sections aims to provide an overview of the life insurance business. The section will commence with an explanation of the concept and the origin of insurance, followed by an indication of the main types of life insurance, and it will be concluded by detailing the main risks associated to life insurance companies.

2.1 Concept & Origin

Fundamentally, insurance is a system which consists of participants pooling economic resources together in order to hedge the risk of incurring in an uncertain loss when a pre-determined event takes place. Each participant must pay a fee, also known as a premium, in order to protect themselves from incurring in losses if the event takes place. These premiums are then pooled together, successfully transferring the risk from the individual to the group, as the group will share the losses incurred by the participants. The volume of premiums will depend on two factors: the likelihood or probability of the event taking place and the expected cost of said event.

The origin of insurance goes back to the 3rd millennia BC, when Chinese merchants distributed their cargo between several smaller ships to avoid a complete loss of goods. In the case of an accident, all participating merchants would assume the losses equally by ceding part of their goods to the affected merchant. Life Insurance is introduced by the Romans in 600 BC, when they created burial clubs or “collegia” which consisted of members paying premiums to cover for their burial expenses in case of death and, in some cases, a pension for their families.

In modern times, insurance companies act as the intermediaries of this system as they collect premiums from participants or policyholders and pay for the corresponding claims or benefits. As defined by the International Financial Reporting Standards (IFRS), an insurance contract is a “contract under which one party (the **insurer**) accepts significant insurance risk from another party (the **policyholder**) by agreeing to compensate the policyholder if a specified uncertain future event (the **insured event**) adversely affects the policyholder.” Insurance companies assume the risks brought forward by each policyholder as they are convinced that the premiums received by them will be sufficient to deal with the possible claims, which will only affect a limited amount of policyholders. This will only be true if the number of policyholders is large enough to be supported by the law of large numbers, which dictates that the larger the number of policyholders independently exposed to the loss, the higher the probability that the actual losses will be equal to the expected losses. To illustrate, an example of the law of large numbers would be as follows: Assume a fair coin where the probability of landing on heads and tails is equal. If we are to toss this coin 100.000 times, the heads-to-tails ratio would be pretty equal, but if we were to only toss the coin 10 or 20 times, we might see that the ratio might be far from equal due to a poor sample size.

Generally, insurance is divided into two lines: life insurance and non-life insurance (also known as property & casualty Insurance or general insurance). Life insurance generally consists of long-term contracts where the loss triggering event is linked with the life, health or disability of an individual, while non-life insurance consists of short-term contracts where the loss triggering event is linked with the damage or loss of property. In this paper, only life insurance will be discussed.

2.2 Life Insurance

As Boros (2014) states, we can divide life insurance into two main categories:

a) Investment Contracts: Where the goal is capital growth, as paid premiums are invested in assets and a fixed or variable rate is given to the policyholder. In Investment Insurance, we can differentiate between two main types of products: Savings with Profit Participation and Unit-Linked.

a.1) Savings Contracts: Consist of mid to long-term contracts, where the single or recurrent premiums paid by the policyholder are compounded at a guaranteed rate. In order to stay competitive, many companies offer profit participation in these types of products. In these cases, the profit which is derived from the investment of premiums in assets is distributed between the policyholders and the shareholders of the company. Depending on the contract terms, at maturity policyholders can receive their payments in form of a lump sum payment or as annuity payment. Depending on the contract terms, if the policyholder is not alive at the maturity of the contract, the benefits may be transferred to the designated beneficiaries.

a.2) Unit-Linked Contracts: The capital which is contributed by the policyholder is invested into mutual funds. Companies generally offer a variety of mutual funds with diverse risk profiles and the policyholder must select the fund most appropriate for their needs. A risk-seeking policyholder will choose a mutual fund with a higher allocation in equity, while a risk-averse policyholder will choose a mutual fund with a higher allocation in fixed income. Policyholders can generally withdraw the account value at any moment. In this type of product, companies obtain profits mainly through fees and commissions.

The main difference between savings insurance and unit-linked insurance lies within the assumed investment risk: in savings products the investment risk lies with the insurance company, while in unit-linked products this risk lies with the policyholder. This is due to the fact that savings products generally guarantee a minimum return which the insurance company is liable for, whereas traditional unit-linked products do not offer such guarantee (there are some exceptions). Essentially, unit-linked policyholders gain transparency and flexibility when compared with traditional savings policyholders, but end up assuming a greater risk in exchange (Munich RE, 2000).

b) Protection Contracts: Where a benefit is paid contingent on the occurrence of a pre-determined event. Depending on the contract, pre-determined events can include, but are not limited to: death, disability, or severe illness of the insured. Protection policies can also offer profit participation, but in this case the profit would be derived from technical surplus (lower-than-expected claims) and not from a financial return. These type of products can also be known as risk contracts.

The scope of this project will be limited to savings with profit participation, unit-linked, and protection policies.

2.3 Risks associated to Life Insurance

Risk is defined as the probability that the actual outcome is not equal to the expected outcome. In life insurance, the main risks which companies face can be categorized as following[2]:

1) Market Risk: Market risk is considerably one of the most important risks faced by life insurance companies. Market risks consists of the risk of loss or of adverse change in the financial situation of a company, resulting, directly or indirectly, from fluctuations in the level and in the volatility of market prices of assets, liabilities and financial instruments.
In Solvency II, Market risk is calculated as an aggregate of various risks:

a) **Equity risk:** The sensitivity of the values of assets, liabilities and financial instruments to changes in the level or in the volatility of market prices of equities.
b) **Interest Rate Risk:** The sensitivity of the values of assets, liabilities and financial instruments to changes in the term structure of interest rates, or in the volatility of interest rates.
c) **Spread Risk:** The sensitivity of the values of assets, liabilities and financial instruments to changes in the level or in the volatility of credit spreads over the risk-free interest rate term structure[3].
d) **Real-estate risk:** The sensitivity of the values of assets, liabilities and financial instruments to changes in the level or in the volatility of market prices of real estate.
e) **Currency risk:** The sensitivity of the values of assets, liabilities and financial instruments to changes in the level or in the volatility of currency exchange rates.

2) Credit Risk: The risk of loss or of adverse change in the financial situation, resulting from fluctuations in the credit standing of issuers of securities (e.g. corporate bonds), counterparties (e.g. reinsurance contracts) and any debtors (e.g. mortgages).

3) Liquidity Risk: The risk that (re)insurance companies are unable to realize investments and other assets in order to settle their financial obligations when they fall due. That is to say, it is the risk stemming from the lack of marketability of an investment that cannot be bought or sold quickly enough to prevent or minimize a loss.

4) Operational Risk: Operational risk consists of the risk of loss arising from inadequate or failed internal processes, personnel or systems, or from external events. Operational risk excludes all the financial risk which a company has taken on, expecting a financial return.
Operational risk comprises the following risks:

a) **Legal risk:** The possibility that lawsuits, adverse judgements from courts, or contracts that turn out to be unenforceable, disrupt or adversely affect the operations or condition of an insurer. The result may lead to unplanned additional payments to policyholders or that contracts are settled on an unfavorable basis, e.g. unrecoverable reinsurance.
b) **Model risk:** The risk that a model is not giving correct output due to a misspecification or a misuse of the model.
c) **Business Risk:** Unexpected changes to the legal conditions to which insurers are subject, changes in the economic and social environment, as well as changes in

[2] Categories and definitions are derived from Solvency II 2009 Directive (Directive 2009/138/EC), Solvency II Glossary (CEA - Groupe Consultatif, 2007) and Corrigan et. al (2009).
[3] In this case, credit spread refers to the difference in yield between the risk-free interest rate term structure and debt securities (government bonds, corporate bonds, and similar assets) considering the same maturity.

business profile and the general business cycle. Business risks include strategic risk[4], management risk[5] and reputational risk[6].

d) **Expense Risk:** The risk of a change in value caused by the fact that the timing and/or the amount of expenses incurred differs from those expected, e.g. assumed for pricing basis.

5) Life Underwriting[7] Risk: Life Underwriting risk tends to be very material for traditional life insurance companies. It consists of the risk of loss or of adverse change in the value of insurance liabilities, due to inadequate pricing and provisioning assumptions. Specifically, it is the risk of a change in value due to a deviation of the actual claims payments from the expected amount of claims payments (including expenses).

The components of Life Underwriting risk are:

a) **Longevity Risk:** The risk of loss, or of adverse change in the value of insurance liabilities, resulting from changes in the level, trend, or volatility of mortality rates, where a **decrease** in the **mortality rate** leads to an increase in the value of insurance liabilities. Pension schemes and annuities are greatly exposed to longevity risk, as they tend to guarantee lifetime benefits for policyholders. This risk has become very material in developed countries as life expectancy has increased greatly.

b) **Mortality Risk:** The risk of loss, or of adverse change in the value of insurance liabilities, resulting from changes in the level, trend, or volatility of mortality rates, where an **increase** in the **mortality rate** leads to an increase in the value of insurance liabilities. Contracts which greatly assume this type of risk are protection policies which pay an agreed benefit contingent on the death of the insured

c) **Morbidity/Disability Risk:** The risk of loss, or of adverse change in the value of insurance liabilities, resulting from changes in the level, trend or volatility of **disability, sickness and morbidity rates**. As in the case of mortality risk, morbidity risk is also assumed by protection policies which are contingent on disability. It is necessary to indicate that there are two aspects associated to morbidity risk: the risk that the morbidity rates will be greater than expected and the risk that the duration of the disability will be longer than expected.

d) **Revision Risk:** The risk of loss, or of adverse change in the value of insurance liabilities, resulting from fluctuations in the level, trend, or volatility of the **revision rates** applied to annuities, due to changes in the legal environment or in the state of health of the person insured.

e) **Life-Expense Risk:** The risk of loss, or of adverse change in the value of insurance liabilities, resulting from changes in the level, trend, or volatility of the **expenses incurred** in servicing insurance or reinsurance contracts.

f) **Life-Catastrophe Risk:** The risk of loss, or of adverse change in the value of insurance liabilities, resulting from the **significant uncertaint**y of pricing and provisioning assumptions related to **extreme or irregular events**.

---

[4] Strategic risk: The risk of a change in value due to the inability to implement appropriate business plans and strategies, make decisions, allocate resources, or adapt to changes in the business environment.

[5] Management risk: The risk associated with an incompetent management or a management with criminal intentions

[6] Reputational risk: The risk that adverse publicity regarding an insurer's business practices and associations, whether accurate or not, will cause a loss of confidence in the integrity of the institution.

[7] Underwriting consists of evaluating the risk and exposure of a potential insured, and subsequently determining the potential coverage for the policyholder and the premium to be paid in exchange. Underwriting is used to detect and prevent adverse selection.

g) **Lapse Risk:** The risk of loss, or of adverse change in the value of insurance liabilities, resulting from changes in the level or volatility of the rates of **policy lapses, terminations, renewals and surrenders.**

The scope of this paper will be limited to Lapse Risk which will be detailed further in the next section.

3. Lapse Risk

This chapter aims to give the reader insight on the concept of lapses and the risk that they can potentially entail for a life insurance company. Furthermore, the impact of lapses under the new Solvency II regulations will be discussed.

3.1 Lapses

According to the Solvency II Glossary[8], a lapse consists in the expiration of all rights and obligations under an insurance contract if the policyholder fails to comply with certain obligations required to uphold those, e.g. premium payments. If the insurer does not receive the premium payment within a defined grace period[9], the policy will be considered as lapsed. Under this definition, lapses are generally considered an involuntary cancellation of the contract by the policyholder. However, it should be noted that the definition of lapses is not consistent throughout literature and can vary depending on the context.

In this paper, a broader definition of lapses will be considered. Lapses will be defined as the contractual disruption of a policy before its maturity, i.e., expiration date. Under this definition, lapses will also include surrenders. A surrender is an option which is embedded in most life insurance products which consists of a voluntary early termination of the contract in exchange for a cash-value payment. As stated by Bacinelllo (2001), a surrender option is essentially an American put option which entitles its owner (the policyholder) to sell back the contract to the issuer (the insurance company) at the cash surrender value. The policyholder has the right, but not the obligation, to exercise the surrender option. We can generally differentiate between two types of surrender values. First, where the surrender value is equivalent to the mathematical reserve of the policy. Second, where the surrender value is equivalent to the market value of the provisions which have been set aside for the policy, i.e., the value of the assets in which the premiums have been invested in. The latter is mandatory for specific products in Spain, namely, for a product named PPA or Plan de Previsión Asegurado. PPA is a long-term savings insurance contract which offers a guaranteed return and is meant to complement or replace pension schemes as policies cannot be surrendered until retirement[10]. PPA products can only be surrendered before retirement if they are being transferred to another company’s fund, and in that case they are to be transferred at market value. In either cases, surrender charges or fees may apply depending on the policy’s conditions.

[8] Solvency II Glossary (CEA - Groupe Consultatif, 2007)

[9] Period immediately after the due date, where the failure to meet contractual obligations is waived, provided that the obligation is satisfied during the time frame.

[10] The surrender option might be available before retirement if the policyholder has been unemployed for a long time or is suffering from a severe illness.

3.1.1 Importance on Understanding Lapse Behavior

Why is it important to understand the dynamics of lapse behavior? Lapses suppose a very material risk to insurance companies, mainly due to five factors:

1) Surrenders present a great liquidity risk for insurance companies, as they have a heavy impact on cash flows due to the surrender value which is paid to the policyholder. If an insurance company does not have enough cash to payout the surrender value, they will be obligated to liquidate assets in order to meet with their commitments, even if it means selling assets with unrealized losses (e.g. selling a long-duration bond which has dropped in price due to an increase in interest rates). Due to the liquidity risk presented by lapses, companies are often forced to buy more flexible assets which, in exchange, offer lower returns.

2) In the case of recurring premium policies, lapses can lead to a loss of future profits as the insurance company will not receive expected future premiums.

3) An early policy lapse might lead to the insurer being unable to fully recover their acquisition costs (initial expenses) which include the cost of procuring, underwriting and issuing new business. Since the insurance company has to pay for these expenses at the time of issue of the contract but earns profits during the life of the contract, an early lapse might cause losses. (Kuo et al., 2003)

4) Lapses may lead to adverse selection as policyholders who are considered riskier, i.e., policyholders who have a higher chance of receiving a benefit from a protection policy, tend to have lower lapse rates. For example, in the context classic life insurance which pays out a benefit on the death of the insured, policyholders who have adverse health or other insurability issues tend to have lower lapse rates than healthy policyholders. (Bluhm, 1982) This indicates that lapses can lead to adverse selection as only riskier individuals would be left in the insurer's portfolio, generating greater claims than expected, especially if early lapse rates are high.

5) A mass lapse scenario can suppose an elevated reputational risk, which can lead to contagion in terms of lapses along with a negative impact over new business. (Eling and Kochanski, 2012)

It is worth taking into account that there are some cases where surrenders could be beneficial for insurance companies, i.e., surrender of an old policy with a very high guaranteed rate.

Due to these reasons, it is of vital importance to understand the drivers and motivation behind lapses, as this would allow for a better predictive capacity and an estimation of more accurate assumptions when dealing with actuarial calculations (e.g. pricing, underwriting, reserving), allowing for a better management of lapse risk. To go into further detail, policyholder behavior will be discussed in Section 4.

3.1.2 Preventive Measures to Mitigate Lapse Risk

Companies have developed many strategies to mitigate or deal with lapse risk. Most of these strategies are implemented in the product development phase, as they involve special product features which discourage early surrenders or protect them from losses in the event of a surrender. Some examples of product characteristics which are likely to disincentive surrenders are:

- Surrender charges.
- Significant tax advantages that are lost on surrender.
- Adjusted participation rates, i.e., in line with market rates.
- Terminal bonus schemes that are lost on surrender. Terminal bonus consists of a postponed profit participation payment which is made when the policyholder reaches a certain age or dies.
- Products with a significant biometric insurance component. Policyholders might not be able to find their current coverage at the same price at another insurance company, as their risk profile might have changed.

An example of a product feature which would protect an insurer from losses in the event of a surrender would be the implementation of Market Value Adjustment (MVA) clauses, which modify surrender charges in order for them to fluctuate depending on market rates. An example would be if an insurer is forced to sell a bond with unrealized losses due to a hike in interest rates, leading to an increased surrender charge for the policyholder. Inversely, if the insurer would have been forced to sell a bond with unrealized gains, it should also lead to a reduced surrender charge for the policyholder. Another example would be the aforementioned surrender charges, which can serve a double purpose as they disincentive surrenders as well as help insurers recover their acquisition costs in case of an early surrender. For this very reason many insurance products include surrender charges which decrease with time, as acquisition costs are recovered.

3.2 Solvency II[11]

In order to protect policyholders, solvency margin requirements have been in place in Europe since the early 1970s. The objective of the solvency margin is to make sure that insurers are able to fulfill their insurance contracts, even under extremely adverse circumstances. These requirements were deemed insufficient by the third-generation European insurance directives which were adopted in the 1990s, thus leading to the implementation of Solvency I in 2002.

The capital requirements under Solvency I consisted in a fixed percentage of technical provisions. The main issue detected was the fact that the calculation was volume-based and not risk –based, therefore punishing low-risk based companies as they would have to maintain the same regulatory capital as high-risk based companies with the same amount of technical provisions. Consequently, there was no incentive to improve the risk management of an entity, and as a result, policyholders were not completely protected from adverse scenarios. Due to these shortcomings detected in Solvency I, a new risk-based supervisory system was introduced in 2009 named Solvency II.

Under this risk-based supervisory system, capital requirements are intended to be aligned with the underlying risks of the company[12]. Solvency II aims to enhance policyholder protection as well guarantee the stability of the financial system as a whole.

Solvency II represents a comprehensive framework for risk and capital management (Conwill, 2016) which encompasses three main pillars:

- **Pillar 1** covers the **quantitative requirements**, by detailing the principles and methods which should be followed when valuating assets, liabilities, own funds, the Minimum

[11] The contents of this section have been derived from Vandenabeele (2014).
[12] Risks faced by life insurance companies have been described in Section 2.3

Capital Requirement and the Solvency Capital Requirement. These will be detailed in the following subsections.

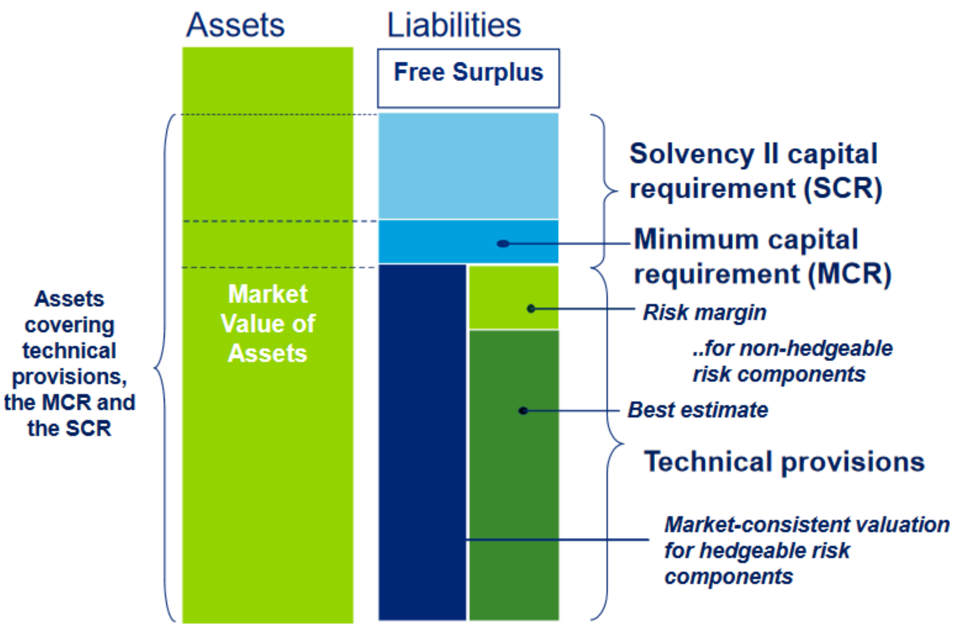
- **Pillar 2** sets out the **qualitative requirements & rules on supervision** which must be met, specifically stating how the entity must be organized in terms of governance.
- **Pillar 3** provides the **reporting & disclosure requirements**, describing the information that should be disclosed publicly along with the information that should be additionally reported to the supervisory entity.

The three pillar system is an inspiration from Basel II, a similar framework in terms of risk and capital management established for banking entities in an attempt to making sure they remain solvent, especially under adverse circumstances.

In the following three subsections, the principal components of the new balance sheet which has been implemented by Solvency II will be discussed.

To illustrate, the balance sheet under Solvency II can be seen as follows:

Figure 1: Balance Sheet under the Solvency II framework



Source: Chládková (2014)

3.2.1 Solvency Capital Requirement

The main risk an insurance company faces is insolvency, i.e., the risk that they will not be able to meet their obligations towards policyholders. A company can be led to insolvency due to adverse events affecting either the assets or the liability side of the balance sheet, mainly due to the different risks which have been mentioned in section 2.3. In order to guarantee the company's solvency, the concept of capital is introduced, defined as the elements on the liability side of the balance sheet that entail no obligations to outside creditors and therefore, can serve as a buffer in adverse circumstances. (Ortiz, 2016) Under this definition, capital can also be named Own Funds or Net Asset Value (NAV), which is defined as:

$$NAV_t = A_t - L_t$$

where A_t is equivalent to the market value of assets and L_t is equivalent to the market value of liabilities.

Under the Solvency II framework, entities are required to have two types of capital: the Minimum Capital Requirement and the Solvency Capital Requirement.

The **Minimum Capital Requirement (MCR)** reflects the minimum level of security required for policyholders. MCR should be calculated as the Value-at-Risk (VaR) of the own funds subject to a confidence level of 85% over a one year period, i.e., the amount of capital required to limit the probability of insolvency over one year to 15%. If an entity breaches the MCR level, it would suppose an unacceptable level of risk for policyholders and the regulator could prohibit the entity from selling new business until the MCR is met.

The **Solvency Capital Requirement (SCR)** reflects the amount of capital an entity should possess in order to reasonably guarantee policyholders that they will remain solvent. It should be calculated as the VaR of the NAV subject to a confidence level of 99.5% over a one year period, therefore corresponding with the amount of own funds necessary to cope with the worst annual loss expected to occur in the next two-hundred years. The SCR should be calibrated to ensure that all the quantifiable risks which the insurance company is exposed to are taken into account and it should also consider risk-mitigation techniques, i.e. diversification of risks. (CEIOPS, 2009). If an entity breaches the SCR level, they would be subjected to intensified supervision and they might be required to take measures in order to meet the SCR within six months.

From a mathematical point of view, one of the definitions[13] which Christiansen and Niemeyer (2014) offer of the SCR is:

$$SCR := VaR_{0.995}(NAV_0 - v(0,1) * NAV_1)$$

where v(0,1) is the discount factor for a one-year period.

Since the calculation of the SCR is quite complex, EIOPA has established a Standard Model which takes a modular approach, assessing each risk individually by applying predefined shocks on own funds and then aggregating all of them, allowing for diversification between risks. Entities are allowed to elaborate their own internal models to calculate the regulatory capital, but they must go through a series of tests and the model must be approved by the local regulator. Entities are also allowed to use partial internal models which consist of a combination of the previous two models, using internal models for some calculations and the standard model for the rest.

Under the standard model, the capital requirement for lapse risk is quantified as the loss of the Net Asset Value under the most adverse of three scenarios:

1) **Lapse-Up Shock**: consists of a permanent increase in 50% of the base lapse rates, although shocked rates shall not exceed 100%.
2) **Lapse-Down Shock**: consists of a permanent decrease in 50% of the base lapse rate, although shocked rates shall not decrease in more than 20% in absolute terms.

[13] It is to be noted that they also include other possible interpretations of the SCR.

3) **Mass-Lapse Shock**: consists of a one-time shock where 40% or 70% (depending on the type of business which is pursued by the insurance entity) immediately surrender.

Mathematically, the following formula can be obtained:
Let $BL(t)$ represent the base lapse rate and $SL(t)$ represent the shocked lapse rate, where $SL_{up}(t)$ represents shock #1, $SL_{down}(t)$ represents shock #2, and $SL_{mass}(0)$ represents shock #3:

$$SL_{up}(t) = Min(BL(t) * (1 + 0.5); 100\%)$$

$$SL_{down}(t) = Max(BL(t) * (1 - 0.5); BL(t) - 20\%)$$

$$SL_{mass}(0) = 40\%$$

Let the capital requirement for lapse risk be represented by $SCR_{lapse-risk}$ and the impact on own funds be denoted as $NAV(shock) = NAV - (NAV|shock)$ where $(NAV|shock)$ represents the Net Asset Value which has been stressed:

$$SCR_{lapse-risk} = Max(NAV\left(SL_{up}(t)\right); NAV\big(SL_{down}(t)\big);\ NAV\big(SL_{mass}(0)\big); 0)$$

3.2.2 Assets

In Insurance companies, assets mostly consist of investments which have been made with the premiums received by policyholders in order to back the technical provisions as well as the regulatory capital requirements. Investments can include government bonds, corporate bonds, equity, real estate, etc. Under Solvency II regulation, "assets shall be valued at the amount for which they could be exchanged between knowledgeable willing parties in an arm's length transaction[14]", or in other words, assets should be market valued.

Surrenders will directly affect the company's assets as they will have to dispose of cash or investments to pay policyholders their corresponding surrender values.

3.2.3 Liabilities

Liabilities are mainly composed of technical provisions which represent the obligations of the insurance company towards policyholders and beneficiaries. Similarly to assets, liabilities must also be market valued as their value should be equal to the amount any (re)insurance entity would have to pay if they were to transfer their (re)insurance obligations immediately to another (re)insurance entity. Under this definition, EIOPA[15] expresses two possible methods which should be followed when valuating technical provisions:

1) **Calculation of Technical Provisions as a whole.** In this case, technical provisions would be equivalent to the market value of the replicating portfolio, composed of assets which can reliably replicate future liability cash flows. Since it is very unlikely to find financial instruments that can reliably replicate cash-flows that depend on the probability of policyholders exercising contractual options (e.g. surrenders), mortality

[14] A transaction between two related or affiliated parties that is conducted as if they were unrelated, so that there is no question of a conflict of interest.
[15] European Insurance and Occupational Pensions Authority, the European regulatory institution which is responsible for the Solvency II framework.

rates, morbidity rates, operational risk, and other non-hedgeable risks[16], this approach is not commonly used.

2) **Calculation of Technical Provisions (TP) as the sum of the Best Estimate of Liabilities (BEL) and a Risk Margin (RM).** This is the most common approach and the one that is expected to be used by default.

$$Technical\ Provisions = BEL + RM$$

Following up on the components of the second approach:

Best Estimate of Liabilities (BEL) consists of the present value of expected future cash flows, which should be discounted using a risk free yield-curve. The projected cash flows should take into account all the future cash-in flows (premiums) and cash-out flows (benefits, expenses, etc.) which would be required to settle insurance obligations. It is to be noted that cash-inflows related to investment returns should not be included when calculating the BEL. On another note, BEL must be calculated gross of reinsurance and without any security margins. Since future cash flows are uncertain, realistic assumptions[17] must be made regarding their developments. Assumptions can be split into two categories: economic assumptions and non-economic assumptions. Economic assumptions are assumptions about variables such as risk-free interest rates or market inflation, which can be derived from information which is available in the financial markets. Non-economic assumptions are assumptions regarding variables which are not observable in the financial markets such as lapse rates or mortality rates. Non-economic assumptions are generally derived from the company's experience along with the sector's experience.

The following formula would be obtained for the calculation of the Best Estimate of Liabilities:

$$BEL = \sum_{t=1}^{n} \frac{CF(t)}{(1 + i(t))^t}$$

where $CF(t)$ is the total cash-flow (outflows minus inflows) and $i(t)$ represents the risk-free yield curve.

The calculation of TP must also include the value of financial guarantees[18] and any contractual options[19] which are included in (re)insurance policies. Examples of contractual options include the surrender option, the annuity conversion option (conversion of a lump sum benefit into an annuity benefit), and the extended coverage option (extension of the policy without further proof of health on behalf of the policyholder) while some examples of financial guarantees would be contracts with a guaranteed capital or a guaranteed investment return. Since the options must be voluntarily triggered by policyholders, while guarantees are generally automatic, realistic assumptions regarding policyholder behavior must be set, determining the likelihood that the options will be triggered.

[16] According to the Solvency II Glossary, a hedgeable risk is a risk associated with an asset or an obligation that can be effectively neutralized by buying or selling a market instrument whose value is expected to change in such a way as to offset the change in value of the asset or liability caused by the presence of the risk.

[17] Realistic or Best Estimate Assumptions should adequately represent the underlying uncertainty of cash-flows.

[18] A financial guarantee is present when there is the possibility to pass losses to the insurer or to receive additional benefits as a result of the evaluation of financial variables. In the case of guarantees, the trigger is generally automatic and thus not dependent of a deliberate decision of the policyholder / beneficiary.

[19] Contractual options are defined as a right to change the benefit, to be taken at the choice of its holder (generally the policyholder), on terms that are established in advance. Thus, in order to trigger an option, a deliberate decision of its holder is necessary.

Options & Guarantees can be separated into two: Intrinsic Value & Time Value. The intrinsic value of options and guarantees is captured in the basic calculation of the BEL, which is calculated with a projection of cash-flows using a deterministic scenario[20]. The Time Value of O&G (TVOG) are more complicated to obtain as they normally require the implementation of stochastic techniques. The most common approach used to value TVOG consists of a stochastic Monte Carlo valuation, where a predetermined number of scenarios (e.g. 1000) are generated based on a set of assumptions. The objective of the scenario generation is to correctly capture the volatilities which are implied in the financial markets. Once the scenarios have been generated, the cash flows are projected for each scenario. Finally, a stochastic BEL is obtained by calculating an average of the stochastic results. Consequently, the TVOG is obtained by applying the following formula:

$$O\&G = Max(BEL^S - BEL; 0)$$

where BEL^S is the average BEL obtained from the stochastic projections and BEL is the BEL obtained from the deterministic projection. The inclusion of O&G would modify the formula for TP: $Technical\ Provisions = BEL + O\&G + RM$

Risk Margin (RM) is defined as the cost of providing an amount of own funds which is equal to the SCR necessary to support the (re)insurance obligations over their lifetime. Establishing a risk margin is only required for non-hedgeable risks, which include underwriting risks, operational risks and financial risks that are not hedgeable as their risk cannot be removed through the financial markets. The risk margin must ensure that the value of technical provisions is equal to the market value, that is, the amount any (re)insurance company would have to pay if they were to transfer their (re)insurance obligations immediately to another (re)insurance company. Risk Margin can be formulaically represented as:

$$Risk\ Margin = CoC * \sum_{t=0}^{n} \frac{SCR(t)}{(1 + i(t))^t} *$$

where CoC is the cost of capital, n is the run-off period of the current (re)insurance obligations, i(t) represents the risk-free yield curve, and SCR(t) is the estimated SCR which is considered necessary for the current (re)insurance obligations. Currently, the cost of capital established by Solvency II is 6%.

4. Policyholder Behavior[21]

In this section, the reader will be given an insight behind the thought process which most policyholders go through before lapsing or surrendering their policies.

Policyholder behavior refers to decisions policyholders take concerning the exercise of financial options or guarantees which are included in their insurance contracts. Modelling policyholder behavior refers to calibrating a set of assumptions regarding these decisions, which has traditionally been done using historical experience.

In the recent years understanding and modelling policyholder behavior has become of great importance, especially due to the importance it is given by Solvency II. Campbell et. al (2014)

[20] In the case of Solvency II, the deterministic scenario consists of the risk-free yield-curves provided by EIOPA. (Volatility Adjustments or Matching Adjustments may be applied)
[21] The contents of this section are derived from Campbell et. al (2014) and Lombardi et. al (2012).

indicate that the main drivers which have led to an increased attention towards policyholder behavior are:

- Product innovation which has led to the creation of products which offer more flexibility to policyholders, allowing them to adjust the product to their needs.
- The sale of investment policies which give policyholders decision-making power regarding the allocation of funds. (e.g. Unit-linked)
- Increased volatility in financial markets.
- Development of complex and thorough financial reporting and regulatory solvency standards (e.g. Solvency II).
- Information & Technology Era: Information is now easily available to policyholders and can greatly influence their behavior.

Modelling policyholder behavior has a great impact on the projection of future cashflows, thus greatly impacting Asset Liability Management (ALM), product design, pricing, reserving, risk management and capital management.

As Campbell et. al (2014) indicate, we need to move beyond concluding “a random 3 percent of our policyholders will lapse this year” to understanding the underlying decisions that led to these lapses and identify which 3 percent of the population it is likely to be, and how their decisions could be influenced, so as to understand what can be done today to change their behaviors and how behaviors might evolve in the future under different scenarios. To be able to model policyholder behavior correctly, policyholders must not be viewed as a model point such as “50 year-old single female, no medical history, smoker” but as member of a household and society, considering their education and social upbringing as well as emotional and cultural aspects which can potentially lead to different thought processes when making decisions. For this very reason, many new approaches have been made towards modelling policyholder behavior such as the study of behavioral economics, predictive modelling and behavioral simulation.

Two major problems have been detected with the traditional approaches which have been used to explain policyholder behavior. First, models have been estimated assessing behavior at an aggregate level and not at a policyholder level. This will be one of the limitations of the present study and will be a principle area for future research, as lapse data was only available at a product level due to privacy reasons. Second, traditional techniques tend to assume that individuals will act rationally, when the reality reflects otherwise as policyholders’ decisions are subjected to the impact of social, intellectual and emotional factors. The following three new approaches intend to address these two drawbacks faced by the traditional techniques.

Behavioral economics refers to the study of customers’ actual (instead of rational) decision making, in an effort to identify the social, economic and cognitive factors which influence their decisions. (Lombardi et al. 2012) This will be discussed in the next subsection, explicitly investigating the motives which lead individuals to lapse or surrender their policies.

Predictive modelling can be defined as the use of data, algorithms, and statistical techniques in order “to make inferences or identify meaningful relationships, and the use of these relationships to better predict future events”. (Batty et. al, 2010). Predictive modeling will be explained in Section 5, as the intention of this paper is to model lapse rates with the help of predictive models such as Generalized Linear Models (GLM).

Behavioral simulation[22] consists in using an agent-based model which uses artificial intelligence in order to simulate policyholders' behavior and predict lapses at a policyholder level. If needed, data can also be analyzed at an aggregated level.

4.1 Main drivers for Lapses & Behavioral Economics

The aim of this subsection is to thoroughly investigate the main divers which lead policyholders to lapse or surrender their policy, beginning with a study of theoretical approaches which have been made throughout literature and following up with empirical studies.

Theoretically, if a policyholder's decision to lapse is optimal from a financial point of view, the main driver for lapsation should be value maximization, i.e., the maximization of wealth. To illustrate, the definition of the surrender option as an American-put option will be retaken. If the surrender option is in-the-money, if exercised it should lead to a profit for the policyholder. Therefore, if the policyholder is driven by value maximization, the contract should be surrendered. In reality, an obstacle is encountered in this aspect as policyholders' decisions regarding lapsation can be far from optimal under a financial point of view, mainly because the behavior of a policyholder is not only influenced by financial factors but is also affected by emotional, psychological and cognitive factors. The need to understand these factors has led to the emergence of behavioral economics in the insurance industry. Some examples of drivers that lead to financially suboptimal surrenders can include: the need for liquidity, change in the policyholder's insurable interest needs (e.g. children grow up and are financially stable and independent), "information constraints" (Christelis et al., 2010), etc.

Campbell et. al (2014) present different psychological factors and biases which are potentially believed to drive lapses. These are represented under four categories of Behavioral Economics:

1) **Decision shortcuts**: Policyholders may lack the skillset which is required to financially valuate insurance products and their embedded options. Therefore, they rely on decision shortcuts and heuristics, i.e., focusing only on a single aspect of a complex problem (Bauer et. al, 2015). Some decision shortcuts which have been detected are:
    - *Relative choices*: When thinking about lapsing a policy, an insured is likely to compare their current policy with other products in the market in order to determine if their product is worth keeping, as it might be hard for a policyholder to determine the "fair value" of a product. Product comparison is based on how attractive the product seems compared to other products, therefore implying that product marketing can have a great impact on policyholder behavior.
    - *Mental accounting*: Individuals commonly create artificial budgets differentiating between saving and spending, which tend to lead to irrational decisions. For example, an individual might be saving funds at a very low rate of return (e.g. child's college fund) but at the same time might be borrowing funds at a very high rate of return (e.g. loan to purchase a vehicle). Since the individual feels that the college fund is "untouchable" due to mental accounting, he or she borrows money at a very high interest-rate when the logical and cheaper method would have been to use the child's college fund to purchase the car and return the money into the fund in the same manner as the loan would have been paid. In terms of insurance, lapses will vary depending on how policyholders view premiums: as

[22] For more information regarding Behavioral Simulation, please see Lombardi et. al (2012).

"savings" or as "expenses". If viewed as "savings", lapses should be lower in comparison.

2) **Value assessments:** The value which customers associate to insurance products plays a big factor in their decision-making process. This perception of value is more easily influenced by emotions rather than by a rational assessment. Some examples of value assessments are:
   - *Hyperbolic discounting*: Individuals tend to behave as if they apply an increasing discount rate to events which are supposed to occur in the future, therefore valuing events in the near future much more than events which will take place in far out in the future. This behavior could lead to an increase in surrenders surrender value would be obtained now, contrary to benefits which would be obtained later on.
   - *Love of free*: When a product is presented as "free", consumers perceive it as more valuable, even though rationally it might not be. If a product has the same technical cost for a consumer in terms of present value of premiums, they will prefer the product which offers a "free" component (e.g. product which pays for itself with investment returns).

3) **Emotional impacts:** Emotions play a big role in policyholder's decision making regarding life insurance. For example, in the case of policies which pay-out on the death of the insured (protection policies), policyholders pay for all the premiums but do not live to see any of the benefits. Another example are savings contracts such as pension schemes, where policyholders save money in order to not be a burden for their family members. Examples of emotional impacts can include:
   - *Risk aversion*: The main reason consumers purchase insurance is because they are not willing to take risks, i.e., they are risk-adverse.
   - *Loss aversion*: Individuals are willing to hold on to "losers", as they are not ready to assume their loss and move on. (E.g. holding on to a stock which has fallen 10% in hopes that it will bounce back up even though it is likely to drop). For example, considering unit-linked products with no guarantees which have incurred in losses, policyholders might want to hold on to the products for longer instead of realizing their losses.
   - *Self-control facilitation:* Some people are willing to accept less freedom in order to nudge themselves to saving. For this very reason, some individuals are willing to accept the restraints which surrender charges impose, as they feel it will press them towards saving and will make it easier to have self-control.
4) **Social impacts:** Policyholders might be greatly influenced by their social environment when making decisions regarding their policies. For example, receiving advice from a friend or family member regarding or reading a newspaper article can greatly influence policyholder's behavior. With the help of the internet and social networks, information travels very quickly and can easily result in the *bandwagon effect*, i.e., contagion of emotions which can lead to common decisions such as mass lapses of an insurance product.

Throughout empirical literature, many different variables and hypothesis have been used in an effort to correctly explain the behavior of lapses. However, two well-known hypothesis stand-out: the Emergency Fund Hypothesis (EFH) and the Interest Rate Hypothesis (IRH).

The **Emergency Fund Hypothesis** declares that policyholders consider the surrender value as an emergency fund to be used in times of financial distress (Outreville, 1990). Also, in the case of policies with recurring premiums, policyholders would not be able to fulfill future payments and would be inclined towards the termination of the policy. To test this hypothesis, the explanatory variable unemployment rate is studied, analyzing if indeed it is a significant predictor variable for lapse rates.

On the other hand, the **Interest Rate Hypothesis** states that increase an increase in market interest rates will lead to an increase in lapses, as policyholders view interest rates as an opportunity cost for holding an insurance policy. (Dar and Dodds, 1989) Moreover, an increase in interest rates will lead to a decrease in premiums and there is a higher chance that a new policy will offer the same coverage at a lower price. This will induce policyholders to lapse or surrender their actual policies in search of new policies with higher returns or lower cost. (Kuo et al., 2003). To test this hypothesis, the significance of interest rates as an explanatory variable of lapse rates will have to be verified, although some variations include the companies crediting rate[23], competitor's crediting rates, and the differential between interest rates and crediting rates as alternative explanatory variables. Campbell et. al (2014) clarify the IRH by stating that lapse rates are positively correlated with external rates of return (e.g. market interest rate or stock returns), and add that lapse rates are negatively correlated with internal rates of return (e.g. high minimum guaranteed crediting rates)

Results in empirical literature regarding these hypotheses have not been consistent. Martin and Kochanski (2012) find that the interest rate hypothesis has found more significance when lapse rates have been studied at an aggregate level, but Bauer et. al (2015) find that the emergency fund hypothesis receives greater support from studies which have been elaborated at a household or policyholder level.

Considering other empirical studies, Fang and Kung (2012) find that when policyholders are young, a big portion of their lapses are mainly due to unobserved idiosyncratic factors which are uncorrelated with health, income and bequest[24] motives. Idiosyncratic factors consist of shocks which are peculiar or particular to an individual, and therefore are random and quite difficult to predict. When policyholders grow older, the impact of these specific idiosyncratic shocks diminishes as lapses are mainly driven by income, health or bequest motives. Health and income factors have a higher impact on young policyholders' lapses, while bequest motives gain significance as policyholders age.

Sirak (2015) argues that individuals with higher income and wealth are less likely to lapse. Additionally, he observes that a transition from an employed state to an unemployment state can increase lapse rates in over 75%. Mullohad and Finke (2014) find that individuals with better numeracy skills have a lower probability of lapsing their contracts. With a study on the German market, Kiesenbauer (2011) discovers that buyer confidence, yield offered by the contracts, and GDP growth are the main economic drivers for lapses. Additionally, Kiesenbauer also includes company characteristic's to his study, finding that distributional channels, company age and the crediting rate are the most relevant company characteristics which help explain lapses. Kim (2005) finds that policy age, unemployment rates, and the lagged difference between the market rate and the crediting rate are significant explanatory variables when trying to predict lapse rates.

[23] A crediting rate is a policy's contractual rate of return, which can be variable or fixed amount and might be subject to a minimum guaranteed rate.
[24] Bequest motives are the need individuals feel to leave behind savings for their heirs or next of kin in the event of their passing, i.e., transfer of wealth between generations.

Considering classic life insurance, i.e., protection policies which cover the contingency of death, He (2011) and Finkenstein et al. (2005) find that policyholders with higher mortality risk are less likely to lapse, leading to adverse selection as described in Section 3.1.1. On the same note, Bauer et al. (2014) find a positive correlation between survivability and surrenders, thus proving the presence of asymmetric information, as policyholders are bound to know more about their health than insurance companies.

To summarize, lapses are very complex to study and correctly predict as they are dependent on many factors such as[25]:

1) Value maximization: policyholder's rationality from a financial point of view
2) Policyholder's financial situation
3) Policyholder's health
4) Contract specific features (e.g. policy age)
5) Regulator's decisions
6) State of the economy
7) Interest rate levels
8) Policyholder's savings and protection needs
9) Available alternative investment and protection opportunities
10) Policyholder's perception of the company
11) Company's reputation

Further conclusions and results regarding studies on lapses can be found in Martin and Kochanski (2012) and Bauer et al. (2014).

4.2 Dynamic Policyholder Behavior

In the present day, policyholders and financial markets are becoming increasingly connected, and information about new products are spread out much more quickly. The critical aspect of this fast-evolving area is thus the dynamic nature of policyholders' behavior, which results in relationships which are not simple aggregations of underlying stable processes. As a consequence, the modeling of such events in particular challenging and dynamic assumptions must be flexible enough to be adapted to the policy's characteristics or to external risk factors. (Barsotti et. al, 2016)

As stated by the CEIOPS[26] (2010), policyholder behavior should not be assumed independent from the financial markets. This will introduce what is currently known as dynamic policyholder behavior (DPHB), the modelling of assumptions on how policyholders will behave in the future, which are assumed to vary according to one or more factors which are unknown at the outset of an actuarial projection (Milliman, 2013). It represents an assumption structure where the values of the assumption vary throughout the projection of the model, in reaction to specific characteristics. The term "dynamic" reflects how policyholders will react to altering external factors such as market interest rates. The dynamic aspect has gained a lot of importance in the past years, especially due to the increasing availability of information which allows policyholders to make informed decisions.

Solvency II requirements state that dynamic policyholder behavior should be assumed and indicates the fundamental principles which should be followed, but no explicit methodology is

[25] See Conwill et. al (2013).

[26] CEIOPS (Committee of Insurance and Occupational Pensions Regulators) is the regulatory institution which EIOPA has replaced in 2011.

provided. As stated by the Article 26 of Commission Delegated Regulation 2015/35: when calibrating policyholder behavior assumptions, the analysis of past policyholder behavior should take into account "how beneficial the exercise of the options was and will be to the policy holders under circumstances at the time of exercising the option" and the "influence of past and future economic conditions".

Dynamic policyholder behavior assumptions mainly impacts three components of an insurance company:

- **Liabilities**: Technical provisions are mainly impacted through the variation of the Options&Guarantees, which can vary significantly with the incorporation of dynamic assumptions due to the projection of stochastic scenarios.
- **Capital Requirements**: Any variation in a company's technical provisions will affect its own funds as well, thus leading to a variation in the Minimum Capital Requirement (MCR) and Solvency Capital Requirement (SCR).
- **Asset Liability Management (ALM):** Due to the changes in valuation of liabilities, assets might have to be managed differently in order for the insurance company to meet with its obligations.

As most insurance risk, dynamic policyholder behavior is generally unhedgeable due to market incompleteness, i.e., no product which successfully hedges this type of risk is available in the market.

In terms of dynamic policyholder behavior, the option which is most commonly used and studied is the lapse or surrender option. However, there are other options that can be investigated as well such as the lump-sum option, annuity option, paid-up option, resumption option, deferment option and extension option. Dynamic lapses will be discussed in the next subsection.

4.3 Dynamic Lapses

According to Cerchiara et. al (2009), lapse rates can be divided into rational and irrational components. The rational component represents lapses which have been driven by a comprehensive following of the financial markets by the policyholder while the irrational component represents all lapses which are independent from the evolution of the financial markets. The rational component considers that policyholders are able to value their lapse option in comparison with the financial market and will exercise their options accordingly. The irrational component will cover all the fluctuations which are not explained by the financial market and will include explanatory variables such as policy age, gender, policyholder's age, product type, etc. Similarly, Barsotti et. al (2016) distinguish between structural and temporary lapses, where structural lapses correspond with the baseline risk, i.e., underlying average lapse risk, and temporary lapses are related to disturbances which are driven by the policyholder's rationality. Structural lapses cover idiosyncratic factors which are policyholder specific such as a policyholder's need for liquidity which is quite difficult to predict but is assumed to follow a stable and independent process in a large portfolio. On the other hand, temporary lapses are scenario specific and depend on the policyholder's rationality such as the valuation of the moneyness[27] of the surrender option. In accordance with Eling and Kochanski (2012), these two components will be referred to as deterministic (irrational or structural) and dynamic (rational or temporary) lapses.

[27]Moneyness refers to state of money of the option. An option can be in-the-money, at-the-money, and out-the-money.

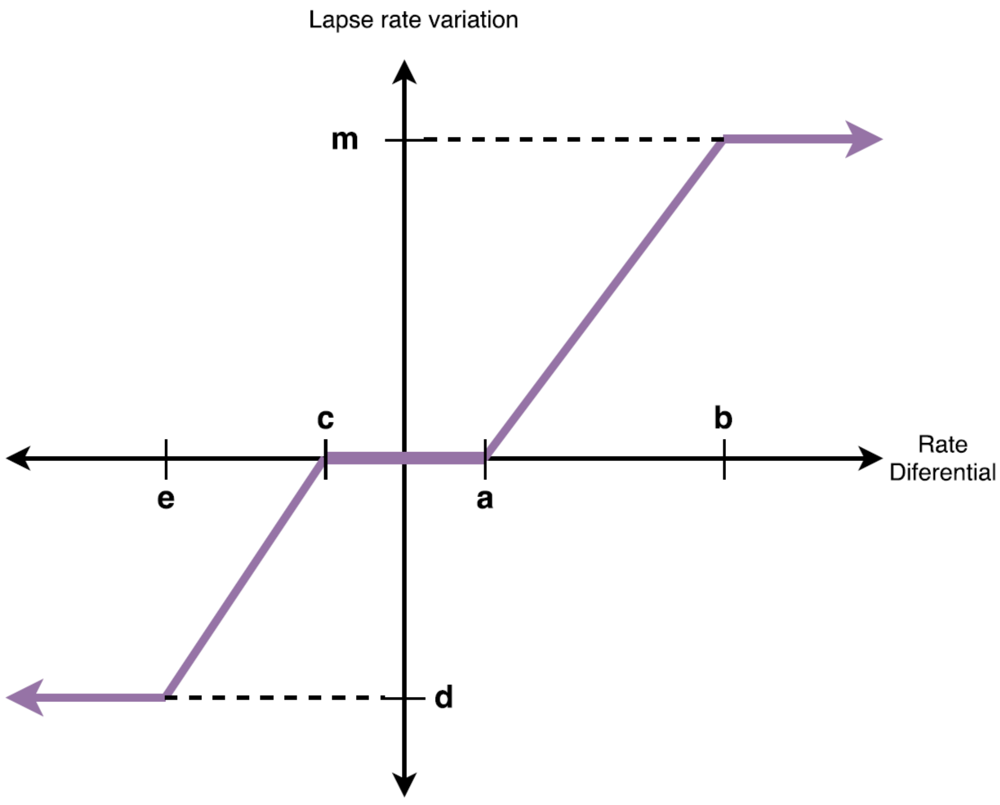
The most common approach used by life insurance companies considers that deterministic lapses will be assumed at the beginning of the projection of cash flows and will not vary throughout the projection, while dynamic lapse assumptions will modify the deterministic component in order to capture the effect of a changing economic environment.

As dynamic lapses are generally only assumed for savings products, the differential or delta between market interest rates and the policy's credited rate is commonly assumed as the most representative driver of dynamic lapses in literature and in practice as well. The market interest rate is supposed to represent the opportunity cost which the policyholder sustains for owning the policy as it represents the return he or she would obtain if an alternate investment were made. In this case, the dynamic lapses are meant to reflect the moneyness of the guarantees through a variation of lapses. If the crediting rates are not competitive, lapse rates will increase as alternative investment opportunities offer higher returns. Inversely, if crediting rates are higher than external rates of return, policyholders will want to stay and lapse rates will diminish. The first case would be considered out-of-the-money while the second one would be considered in-the-money.

Even though the differential between the credited interest rate and the market rate may not completely represent the value which the lapse option has for the policyholder, it is an indicator which is relatively easy to understand and comprehend for policyholders. (Milliman, 2013).

The model which is most commonly used by life insurers is the following:

Figure 2: Dynamic Lapse Function



Source: Own Elaboration

This model considers 6 parameters which have to be calibrated. The dynamic lapse function will modify the deterministic (base) lapse rate depending on the differential between the reference market rate (RMR) and the policy credited rate (CR), thus increasing lapses if the market rate is higher than the credited rate and reducing lapses if the market rate is lower than the credited rate. The dependency between the differential rate and lapses will be calibrated through the implementation of six parameters:

1) Cap for lapse increases (m): Maximum lapse increase.
2) 1st trigger – Increase (a): Differential rate value (RMR-CR) where lapse rates will begin to increase.
3) 2nd trigger – Increase (b): Delta value where lapse rates stop increasing and remain at the cap, i.e., point where a marginal increase in the delta value will no longer increase lapse rates.
4) Floor for lapse decreases (d): Maximum lapse decrease.
5) 1st trigger – Decrease (c): Differential value which causes lapse rates to start decreasing.
6) 1st trigger – Decrease (e): Delta value where lapse rates stop decreasing and remain at the floor value, i.e., point where a marginal decrease in the delta value will no longer decrease lapse rates.

The idea behind the first triggers is that small values of the differential rate do not heavily affect lapse rates, as policyholders are not likely to react to small market movements. Second triggers are in place because, after a certain gap between market rate and credited rate, lapses are no longer affected marginally and will remain at a minimum or maximum value.

With the help of these parameters, the base lapse rate (BL) will be adjusted in order to capture the moneyness of the product in the current market state.

Let $Delta = RMR - CR$. Assuming that $e \leq c \leq a \leq b$, the value of the adjusted lapse rate will be equal to:

$$AL = \begin{cases} BL * (1+m), & \text{if Delta} \geq b \\ BL, & \text{if c} \leq Delta \leq a \\ BL * (1+d), & \text{if Delta} \leq e \\ BL * (1+r), & \text{otherwise} \end{cases}$$

Where

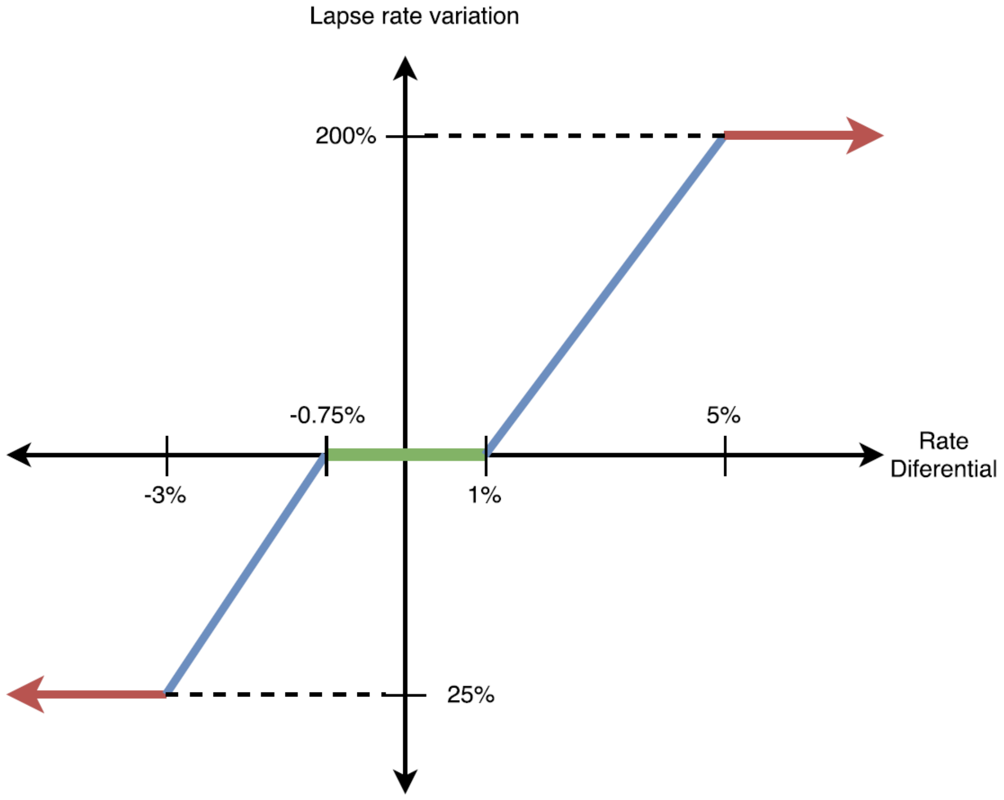
$$r = \begin{cases} m * \left(\frac{Delta-a}{b-a}\right), & \text{if Delta} \geq a \\ d * \left(\frac{Delta-c}{e-c}\right), & \text{otherwise} \end{cases}$$

To illustrate, an example will be presented. Assuming that the base lapse rate is equal to 10% and that the following values are assigned the six dynamic lapse parameters:

- Cap for lapse increases (m) = 200%
- 1st trigger – Increase (a) = 1%
- 2nd trigger – Increase (b) = 5%

- Floor for lapse decreases (d) = -75%
- 1st trigger – Decrease (c) = -0,75%
- 2nd trigger – Decrease (e) = -3%

Figure 3: Dynamic Lapse Example



Source: Own Elaboration

Considering five different scenarios, the following adjusted lapse rates are obtained:

Table 1 : Dynamic Lapse Example

Scenario	RMR	CR	Delta	Factor	BL	AL
1	6%	3%	3%	2,2	10%	22%
2	4%	5%	-1%	0,75	10%	7,5%
3	2%	1,75%	0,25%	1	10%	10%
4	9%	2%	7%	3	10%	30%
5	2%	6%	-4%	0,25	10%	2,5%

Source: Own Elaboration

The factor refers to the number the base lapse is multiplied by in order to obtain the adjusted lapse rate. According to the parameters established in this example, the maximum lapse rate is

30% while the minimum lapse rate is 2,5%. Depending on the value of delta, the adjusted value of lapses will move in between these two rates.

In literature, no explicit methodology can be found regarding the estimation of these specific dynamic lapse parameters. This traditional approach has some serious limitations as it is fairly simple and will not be able to completely capture the underlying complexity of dynamic lapses, but it is a useful starting point towards more advanced techniques.

As stated in section 4.2, a way to quantitatively determine the impact dynamic lapses have is mainly through the Options&Guarantees component of the technical provisions. It is also worth noting that the assumption of dynamic lapses can have a very material impact over a company's risk management, as it successfully links market risk (e.g. interest rate risk) with insurance risk.

5. Predictive Modelling: Generalized Linear Models

Predictive modelling consists of the use of complicated algorithms and statistical techniques on a particular data set in hopes of determining possible variables which can explain the movements of a dependent variable, as well as determining co-relationships between explanatory variables.

The most commonly used predictive models in Life Insurance are the Generalized Linear Model (GLM) and the Classification and Regression Tree (CART). In this section the theoretical aspects of Generalized Lineal Models will be presented, as they the models which will be applied in the empirical analysis of this study.

Generalized Linear Models are a broad class of regression models which have been popularized by McCullagh and Nelder (1982).

The following explanation of GLM has been derived from PennState (n.d., b) and Bolancé (2016).

There are three components to any GLM:

1) Random Component – refers to the probability distribution of the response variable (Y). Instead of always assuming the normality of residuals, GLMs include the possibility of working with the distributions which are included in the exponential family. Namely, the most used distributions include the normal, binomial, Poisson and gamma distributions.

2) Systematic Component - specifies the explanatory variables (X_1, X_2, ... X_k) in the model, more specifically their linear combination in creating the so called *linear predictor*; e.g., $\beta_0 + \beta_1 x_1 + \beta_2 x_2$ as seen in linear regressions.

3) Link Function, η or g(μ) - specifies the link between random and systematic components. It says how the expected value of the response relates to the linear predictor of explanatory variables; e.g., $\eta = g(E(Y_i)) = E(Y_i)$ for linear regression, or $\eta = logit(\pi)$ for logistic regression. The most commonly used link functions include: identity, logit, log and inverse.

These models generalize linear regressions, as a GLM with a normal distribution and an identity link function is equivalent to a linear regression.

6. Methodology

The aim of this section is to instruct the reader on the complete methodology which has been used in this study, outlining the data which has been used and the procedure which has been followed in the modelling of lapse rates.

6.1 Data

In this subsection, the data which has been used in the modeling of lapses will be described.

6.1.1 Dependent Variable

The dependent variable of the present study will be the **lapse rate.** This response variable has been modelled with the objective of identifying independent variables which are significantly explanatory in explaining future lapse behavior, allowing for an accurate prediction of future lapse rates and setting of reliable assumptions for projection purposes.
Quarterly lapse rate data has been obtained from a life insurance company which is operating in the **Spanish market.** The lapse rate has been derived by the following formula:

$$Quarterly\ Lapse\ Rate = 100\ x\ \frac{Number\ of\ Policies\ Lapsed\ During\ Quarter}{Number\ of\ Policies\ Exposed\ to\ Lapse\ During\ Quarter}$$

Lapse Rates have been modelled separately depending on the product type[28], differentiating between five different product types:

- Individual Savings
- Group Savings
- Individual Protection
- Group Protection
- Unit-Linked

The term individual refers to policies which have been established with an individual, while group refers to policies which have been established with a collective or group of individuals. Group policies are normally contracted by companies.

The rationale behind this division is that lapse rates differ vastly between product types, as different products cover different needs and may have completely different characteristics.

Furthermore, this division will allow to test for behavioral differences between groups and individuals, as it is assumed that group policies will be handled by professional investors who tend to act more rationally from a financial point of view.

6.1.2 Independent Variables

The explanatory variables of this study are mainly composed of economic variables and company specific variables (e.g. crediting rate).

[28] Research has verified that dividing the data into these five different groups leads to better results.

Since only aggregated lapse data is available[29], policyholder characteristics (e.g. age, gender, income) and contract-specific characteristics (e.g. policy age) cannot be included in the study of lapse rates.

1) Unemployment Rate: The rate considers the percentage of unemployed individuals which belong to the employable population[30]. Unemployment rate has been the most studied variable throughout empirical literature in the context of the aforementioned emergency fund hypothesis (Outreville, 1990). Quarterly data has been obtained from the Spanish National Institute of Statistics (INE).

2) Inflation Rate: Inflation refers to the rate at which prices of products and services increase. Inflation rate can also be considered as an important indicator of economic growth due to the fact that when the demand of products and services increase, prices are likely to increase as well. Quarterly data has been obtained from the Spanish National Institute of Statistics (INE).

3) GDP Growth: More specifically, the considered data is the quarterly growth (%) of the Spanish gross domestic product. The gross domestic product refers to the market value of all the final products and services which have been produced during a specific period (in this case quarterly). The development of GDP will be a good indicator for economic growth and can also be used to test the emergency fund hypothesis. Quarterly data has been obtained from the Spanish National Institute of Statistics (INE).

4) Δ Risk Free Rate (Δ RFR) = **Risk Free Rate – Crediting Rate**

This predictor is a differential between two different variables:

The first variable, risk free rate (RFR), has been obtained from EuroStat, the EU's statistical institution. The data corresponds to the "Zero-coupon yield curve spot rate" of the Euro Zone which has been derived from government bonds which have been classified as AAA[31] and can be considered risk-free. Yield curves have been obtained quarterly and each curve contains annual spot rates for a 30 year period, i.e., thirty different spot rates for thirty different possible maturity years. For each product, the interest rate which corresponds to the maturity year that is equivalent to the duration of the product has been selected. Annual rates have been transformed into quarterly rates to maintain uniformity with the rest of the variables of this study.

The second variable, crediting rate, has been obtained at a product level from the same life insurance company which has provided the lapse rate data. The crediting rate refers to the internal rate of return which policyholders receive from the savings component of the premiums to the insurance undertaking. Generally, crediting rates can vary whenever the company prefers and will normally be representative of the returns which the company is earning in the market. Depending on the contract, a minimum crediting rate can be established and is generally referred to as minimum guaranteed rate.

This differential is meant to quantify the opportunity costs which policyholders withstand for owning insurance policies. If the differential is positive, policyholders would be able to obtain

---

[29] Data at a policyholder level was not available due to confidentiality issues.
[30] Employable population refers to individuals which belong to the labor force and are working or actively looking for work.
[31] AAA is the highest possible rating which can be assigned to a bond. It indicates that the counterparty will easily be able to meet its financial commitments

higher returns investing in risk-free assets and would logically lapse their policies in order to access these risk-free returns. Inversely, if the differential is negative, policyholders are obtaining higher returns from the policy than they would be able to from risk-free assets and therefore, should be less likely to lapse. This variable will be able to test of the aforementioned interest rate hypothesis (Kuo et. al, 2003).

As the crediting rate is only available for savings products, **Δ RFR** will be equivalent to the risk free rate for protection and unit-linked products.

5) Δ Reference Market Rate (Δ RMR) = Reference Market Rate – Crediting Rate

This variable is very similar to the previous one, as the only difference is the use of the reference market rate instead of the risk free rate.

The reference market rate (RMR) is considered as it incorporates a return considering "riskier" assets such as the Spanish governmental bonds[32], these being a common investment of competitor life insurance companies. Therefore, it is considered that Spanish governmental bonds are a fair approximation to life insurance competitor's crediting rates. Quarterly data is obtained from the web page "Investing.com". The data corresponds to historic annual returns of Spanish governmental bonds, including results of bonds with maturities of three, five, ten and fifteen years. For each product, the variable RMR will consider the returns of the bond with a maturity which is closest to the duration of the product has been selected. For the purpose of this study, the annual rates have been transformed into quarterly rates.

In the same manner as ΔRFR, Δ RMR will correspond with the reference market rate for protection and unit-linked products as no crediting rate is offered for these type of products.

6) IBEX 35 Returns

Returns of the stock market are considered as they are the most common "risky" assets which consumers purchase and can be contemplated as an alternative to life insurance savings products. Stock market returns of general indexes are also a great indicator of economic growth. The Spanish indicator of stock market performance is the IBEX 35, an index which includes the weighted[33] performance of the 35 largest companies in the Spanish stock market. Closing prices of IBEX 35 have been obtained from Yahoo Finance. Quarterly returns have been calculated using the logarithmic approach, which considers returns as:

$$Quarterly\ Return = Ln[\frac{P_t}{P_{t-1}}]$$

where Pt is the closing price of the index for the considered quarter and Pt-1 is the closing price of the index for the previous quarter.

All the data has been obtained quarterly, ranging from 2004 to 2016. The observations from the year 2004 have been left out from the study in order to use lagged variables. Four lags of each independent variable have been introduced, in order to account for a complete year and capture potential effects of seasonality (e.g. more surrenders in a certain quarter due to tax incentives). Moreover, four lagged variables of lapse rates have been introduced in the study

[32] As of 31/03/2017, Standard & Poor's credit rating for Spanish bonds is BBB+.

[33] Stock returns of companies are weighted by their market capitalization, i.e., the market value of the company's shares.

as well, with the objective of capturing trend and seasonality effects. The rationale behind lagged explanatory variables is that individuals may take time to react to external factors and take their corresponding decisions. Additional lags have not been introduced in order to not reduce an already small sample of observations.

To recap, the independent variables which have been included in this study are the following:

Table 2: Possible Explanatory Variables

Possible Explanatory Variables						
Lapse(-1)	Unemployment	Inflation	GDP	ΔRMR	ΔRFR	IBEX35
Lapse(-2)	Unemployment(-1)	Inflation(-1)	GDP(-1)	ΔRMR(-1)	ΔRFR(-1)	IBEX35(-1)
Lapse(-3)	Unemployment(-2)	Inflation(-2)	GDP(-2)	ΔRMR(-2)	ΔRFR(-2)	IBEX35(-2)
Lapse(-4)	Unemployment(-3)	Inflation(-3)	GDP(-3)	ΔRMR(-3)	ΔRFR(-3)	IBEX35(-3)
	Unemployment (-4)	Inflation(-4)	GDP(-4)	ΔRMR(-4)	ΔRFR(-4)	IBEX35(-4)

Source: Own Elaboration

The parenthesis and the number within symbolize a lagged variable, where (-2) for example represents a lagged variable from two quarters ago.

6.2 Method & Procedure

The aim of the present study is to obtain a simple and interpretable model which will allow for a clear understanding of the underlying relationship between lapse rates and the aforementioned independent variables.

6.2.1 Variable selection

As seen in the previous subsection, the study includes over 35 explanatory variables, yielding a very complex model which includes too many variables for a small amount of observations. In order to avoid overfitting[34], model selection criteria will be applied in order to reduce the amount of explanatory variables and select the most significant ones. For this, the procedure GLMSELECT will be used on SAS (Statistical Analysis Software). GLMSELECT procedure consists of a model selection method which allows the user to identify which combination of parameters provides the best model for the selected data. This procedure includes five model selection methods[35]: backward selection, forward selection, stepwise selection, LASSO selection and LAR selection, all of which have been used in the present study.

For the first three selection methods, also known as the traditional methods, a selection criterion must be specified. The selection criterion customizes how the different parameters are selected and will, by default[36], also indicate when the process should come to a halt. In this study, the following selection criteria will be used in the GLMSELECT procedure:

[34] Overfitting can occur if a model has too many parameters, resulting in an overly complex model which will correctly predict the underlying data which has been used to specify the model but will not be very accurate when applied to external data. This phenomenon takes place because when the model predicts specific noise or idiosyncrasies instead of the underlying relationship between the predictor and response variables.
[35] For more information on these selection methods, please refer to the Cohen (2006).
[36] The procedure also allows for users to set a STOP criterion different from the selection criterion.

AICC: The Akaike Information Criterion (AIC) is a measure of the relative quality of a statistical model. Note that AIC cannot be used as an absolute measure; it provides no value for a sole model. It can only be used to compare two or more different models which have been fit to a given set of data. A lower value of the AIC statistic indicates a better model. In this case, AICc is used, a correction of AIC which is meant to be used when sample sizes are small compared to the number of estimated parameters as AICc tends to select more parsimonious models, i.e., models that captures all the characteristics of the data using the minimum number of parameters possible.

Cross-Validation: The data is randomly divided into k-equal parts. The procedure leaves out one part for validation purposes and fits the model with the k-1 remaining parts. Consequently, the fitted model is used to predict the part of data which was initially left out and the prediction error is obtained. This process is repeated until all parts of the data have been left out and predicted at least once. Once all k parts have been estimated, the prediction error is combined and the statistic PRESS (Predicted Residual Sum of Squares) is obtained. When this selection method is chosen, only parameters which reduce the PRESS are included in the model.

Adjusted R^2: The coefficient of determination or R^2 is a statistical measure which indicates how well a model is able to predict future outcomes, as it represents the proportion of variance of the response variable which can be predicted by the selected predictor variables. The problem with R^2 is that it increases whenever a predictor is added to the model (never decreases) and, therefore, is likely to take large values for over fitted models which might be modeling random noise. For this reason, Adjusted R^2 is used instead. Adjusted R^2 is a modified version of R^2 which takes into account the number of explanatory variables in the model and penalizes the addition of predictors which are not helpful in describing the variance of the response variable.

For example, if we consider backward selection with AICc as the selection criterion, the procedure will begin the study by including all of the explanatory variables along with an intercept in the model. The parameters whose removal yields a decrease in AICc are successively removed from the model in a gradual manner. These effects are removed until no more variables that reduce the AICc can be removed from the model.

Eleven[37] different models will be obtained for each product type and only one of them must be selected. In order to do so, the following statistics will be analyzed for each model: AICc, SBC, Adjusted R^2, number of parameters, number of significant[38] parameters, and the existence of multicollinearity between the selected variables. These statistics are meant to be an indicator of the goodness of fit of the model. Schwarz Bayesian Criterion (SBC) is a model selection criterion which is very similar to the AIC, as it indicates the relative quality of a model for a given set of data. Similarly to the AIC, the model with the lowest SBC value should be selected. SBC is known to be more restrictive than the AIC as it applies higher penalties to complex models, i.e., models with a high number of parameters. On another hand, multicollinearity is a phenomenon that occurs when two or more explanatory variables are highly correlated. Multicollinearity can lead to misleading or inaccurate estimation of parameter coefficients and, consequently, inaccurate predictions. The existence of multicollinearity will be assessed with the support of the Variance Inflation Factor (VIF) which is defined as:

[37] For each traditional selection method (stepwise, forward, and backward) three models will be obtained for each selection criterion (AICc, Adjusted R^2, Cross-Validation). Additionally, LASSO and LAR models will be obtained.

[38] A parameter is considered statistically significant if the null hypothesis is rejected, as it indicates that a relationship exists between the parameter and the studied variable. The significance level which has been considered for this study is 5%.

$$VIF = \frac{1}{1 - R_k^2}$$

where "R_k^2 is the R^2-value obtained by regressing the k^{th} predictor on the remaining predictors. Note that a variance inflation factor exists for *each of the k predictors* in a multiple regression model. "(PennState, n.d., a)

The variance inflation factors have been calculated in SAS using the regression procedure, PROC REG.

The priority was to select the model with the lowest AICc and SBC values, and the highest Adj. R^2 value along with the least number of parameters possible, where most of them should be individually statistically significant and not present multicollinearity. As it is hard to find a model which meets all these requirements due to trade-offs between some of these statistics, a table with a summary of these statistics will be presented and the most adequate model will be selected. As stated at the beginning of this section, the objective is to obtain a parsimonious model which incorporates the minimum amount of parameters necessary to adequately explain the behavior of the data.

6.2.2 Model selection

Once the variables have been selected, the model is ready to be estimated in the framework of generalized linear models.

As stated in the previous section, when working with generalized linear models a theoretical distribution for the response variable and a link function must be established. Since the current study intends to model aggregate lapse data in the form of percentages, the normal distribution and the beta distribution have been considered adequate and were tested. The normal distribution was considered because it is always a great starting point due to its inherent characteristics and because it is relatively simple to calculate and comprehend. On the other hand, the beta distribution was selected as a potential candidate because it delimits data to the interval (0, 1) which is where percentages lie and it allows for skewness which is not possible when assuming normality. It should be noted that variables that are equal to 0 or 1 are not included in the study when considering the Beta distribution. The link functions which have been considered in this study are identity, log, logit and complementary log-log.

If the study were based on a policyholder level, where the lapse variable is binary as it can only take the values 0 or 1, the most adequate distributions would be the Poisson or the negative binomial regression (Cox and Lin, 2006). As this is not the case, these distributions have discarded.

In order to estimate the generalized linear models, the GLIMMIX procedure will be used in SAS as it incorporates the beta distribution[39], unlike the commonly used GENMOD procedure. Even though GLIMMIX allows for the modelling of generalized linear mixed models, only generalized linear models will be estimated in this study. Parameters will be estimated by maximum likelihood.

[39] For more information on how the Beta distribution is modelled under the framework of Generalized Linear Models, please refer to Ferrari and Cribari-Neto (2004).

In accordance with Kim (2005) and Kiesenbauer (2011), RMSE and MAPE will be calculated for each of the potential models. The root mean square error (RMSE) is calculated as:

$$\frac{1}{\sqrt{n}}\sqrt{\sum_{k=1}^{n}(y_k-\hat{y}_k)^2}$$

while mean absolute percentage error (MAPE) is calculated as:

$$\frac{1}{n}\sum_{k=1}^{n}\frac{|y_k-\hat{y}_k|}{y_k}$$

RMSE is a relative indicator of error, as it is represented in same units as the predicted variable, while MAPE is considered an absolute indicator of error as the error is defined percentually.

The model will be selected according to the AICc criterion, SBC criterion, as well as the RMSE and MAPE statistics, where a smaller value of these statistics is an indicator of a better relative model.

7. Results & Analysis

In this section the results which have been obtained will be presented and analyzed. To avoid repetitiveness, the whole procedure will only be discussed for Individual Risk products and a summary table including the selected model for all product types will be presented and analyzed at the end of this section.

The procedure which has been detailed in the previous subsection will be followed in the same order. First, in order to avoid overfitting and obtain a parsimonious model, the specific explanatory variables which will be included in the model are selected using the GLMSELECT procedure in SAS. Five different types of regression will be applied, using three different selection criteria for the first three types of regression. The results are presented in the following table:

Table 3: Model Selection for Individual Protection

Regression	Selection Criteria	Adjusted R^2	AICC	SBC	# Param.	# Sig. Param.	Multicol.
Backward	Cross-Validation	0.675	-218.77	-445.45	35	1	YES
Forward		0.713	-460.30	-499.41	9	7	YES
Stepwise		**0.720**	**-463.47**	**-503.24**	**8**	**7**	**NO**
Backward	Adjusted R^2	0.800	-417.80	-481.97	24	12	YES
Forward		0.696	-455.37	-493.99	10	8	YES
Stepwise		0.757	-449.18	-489.64	16	9	YES
Backward	AICC	0.749	-447.73	-488.20	16	16	YES
Forward		0.676	458.29	-498.89	7	3	NO
Stepwise		0.662	-459.59	-502.28	5	4	NO
LASSO	LASSO	-	-	-	-	-	-
LAR	LAR	-	-	-	-	-	-

Source: Own Elaboration

The objective is to select the model which can adequately explain the behavior of lapse rates with the minimum number of parameters possible. Additionally, no multicollinearity must exist between the predictor variables.

As shown in Table 3, the methods LASSO and LAR were not able to select an adequate model for this specific product type. In this case the best available model is the one obtained from the Stepwise regression, using cross-validation as the selection criteria. This model has been chosen because, even though it does not have the highest adjusted R^2, it has the lowest AICc and SBC values, along with the lowest number of parameters where the majority are statistically significant. Moreover, the model does not present multicollinearity. The stepwise regression yields the following model under the framework of linear regressions:

Table 4: Selected Model for Individual Protection

Parameter Estimation						
Parameter	**DF**	**Coefficient**	**Std. Error**	**t-statistic**	**Pr > |t|**	**VIF**
Intercept	1	-0.02029	0.00693	-2.93	0.0056	0
Inflation (-1)	1	-0.27026	0.07987	-3.38	0.0016	1.94
Inflation (-3)	1	-0.24189	0.09799	-2.47	0.0179	2.95
Lapse (-1)	1	0.62921	0.13584	4.63	<.0001	2.40
Lapse (-4)	1	0.52619	0.12893	4.08	0.0002	1.61
GDP (-4)	1	0.45938	0.10483	4.38	<.0001	1.58
IBEX35 (-3)	1	0.00985	0.00657	1.5	0.1414*	1.10
Δ RMR (-1)	1	1.38582	0.26994	5.13	<.0001	2.01

Source: Own Elaboration

Even though the IBEX35 (-3) is not considered statistically significant individually, it does provide value as an additional component of the aggregate of predictor variables as its removal returns a model with a higher AIC and a lower Adjusted R^2. Consequently, it is considered appropriate to maintain this parameter in the modelling of lapse rates. Apart from this variable, all other variables are considered statistically significant and additionally, they do not present multicollinearity as their variance inflation factors are low (smaller than 10). As stated in Section 5, a linear regression is a specific case of generalized linear models, where the distribution is Gaussian (normal). Coincidentally, this statement will be proved in the next step.

Once the explanatory variables have been identified, the specification of the Generalized Linear Model can take place. Namely, two factors must be selected when using generalized linear models: a theoretical distribution for the response variable and a link function for the predictors. As discussed in the previous subsection, the model will be tested for the normal and beta distribution, using four different links: identity, log, logit and the complementary log-log function. Testing results for individual protection products are provided below in Table 5.

Table 5: Generalized Linear Model Selection for Individual Protection

Distribution	Link	AICc	SBC	RMSE	MAPE
Normal	**Identity**	**-375.25**	**-363.15**	**0.0038**	**0.0835**
	Log	-368.74	-356.64	0.0041	0.0925
	Logit	-369.01	-356.91	0.0041	0.0920
	Complementary Log-Log	-368.88	-356.77	0.0041	0.0922
Beta	Identity	-368.8	-356.69	0.0039	0.0846
	Log	-359.8	-347.7	0.0042	0.0915
	Logit	-360.15	-348.04	0.0042	0.0913
	Complementary Log-Log	-359.97	-347.87	0.0042	0.0914

Source: Own Elaboration

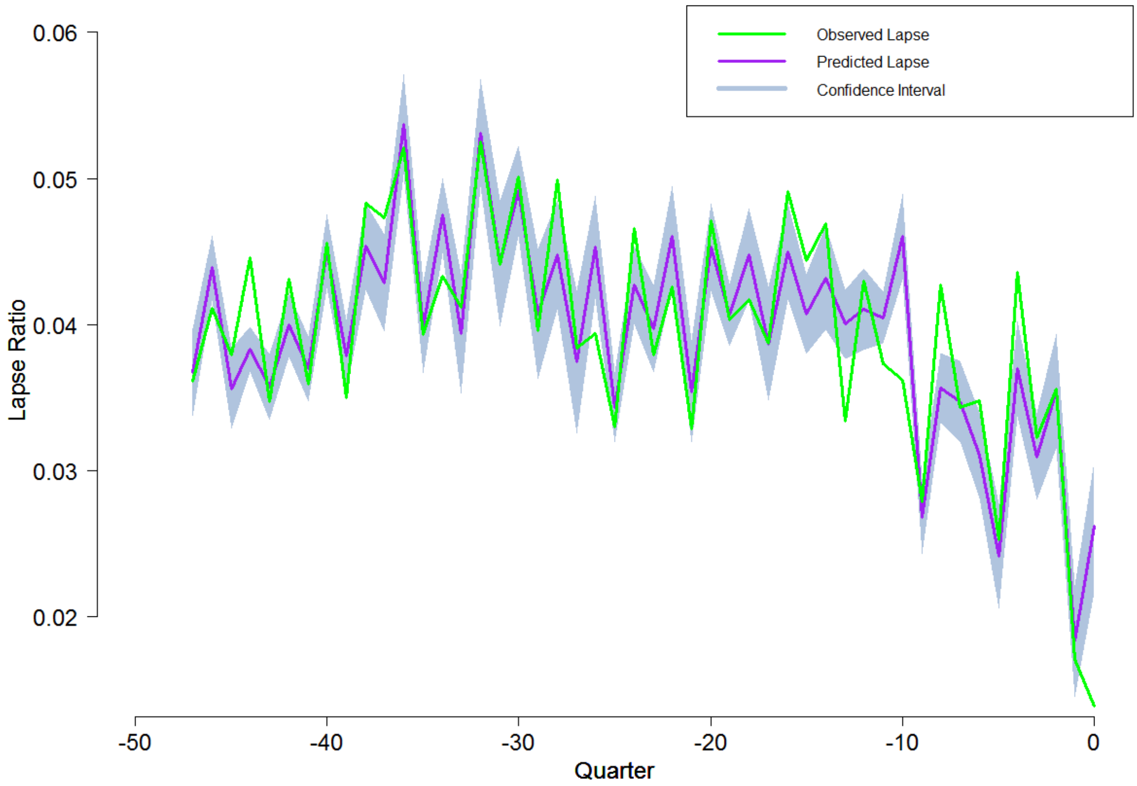
In this case it is fairly easy to select the most appropriate combination of distribution a link function as the normal distribution with the identity log is the model with the lowest AICc, SBC, RMSE and MAPE, making it the best relative model considering these statistics.

The proposed model for lapse rates in this case is:
$\hat{Y}$ = -0.02029 - 0.2703 * Inflation (-1) - 0.2419 * Inflation (-3) + 0.6292 * Lapse (-1) + 0.5262 * Lapse (-4) + 0.4594 * GDP (-4) + 0.009854 * IBEX35 (-3) + 1.3858* Δ RMR (-1)

As foreshadowed, this is the exact same model which was yielded by the stepwise regression when running the GLMSELECT procedure.

The observed lapse rates and the predicted lapse rates for Individual Protection products are illustrated below in Figure 4.

Figure 4: Predicted Lapses for Individual Protection



Source: Own Elaboration

For the most part the observed lapse is contained in the confidence interval and predictions of the lapse rate seem to be quite accurate, except for some upward jumps which are not completely captured by the predictive model.

Due to the possession of a very limited sample size, validation of the model with external observations is not possible as all the data has been used to fit the model.

The steps which have been described for the product type Individual Protection have been applied to each product type as well. Table 6 and 7 illustrate the final results which have been obtained for each product type.

Table 6: Summary of Selected Models per Product Group

Product type	Distribution	Link function	MAPE	RSME	Adj. R^2
Group Protection	Beta	Identity	0.1396	0.00420	0.8414
Group Savings	Beta	Log	0.6149	0.01257	0.3257
Individual Protection	Normal	Identity	0.0835	0.00383	0.7203
Individual Savings	Beta	Identity	0.2573	0.00735	0.7539
Unit Linked	Beta	Log	0.3437	0.03402	0.5131

Source: Own Elaboration

It should be noted that the Adjusted R^2 does not correspond to the estimated generalized linear model, it corresponds to the multinomial linear regression which has been fitted in the previous step when executing the GLMSELECT procedure. It has been included to give an estimated indicator of the prediction power of the model.

Interestingly, four out of the five products have been fit with the Beta distribution and only the log or identity link function have been selected.

Both MAPE and Adjusted R^2 are indicators of the predictive power of the model. They are negatively correlated, as when the adjusted R^2 increases, MAPE will decrease. Concretely, MAPE describes the predictive accuracy of the model when applied to the data which was used for fitting, indicating the % of error in the predictions. In this case, the MAPE is pretty large for most product types. Specifically, MAPE is very high for Group Savings, mainly due to the fact that the Group Savings portfolios are very small as policies are normally contracted directly with other entities and only one policy exists per entity.

Generally speaking, the low predictive power of these models can mainly be attributed to working with a small set of observations as well as working with an aggregate set of data. Even though the aggregation of data might eliminate the noise and result in a "cleaner" set of data, the underlying causes of lapse behavior which are inherent to each policyholder might be lost, thus making it harder to completely identify the drivers of lapses.

Table 7: Proposed Model per Product Group

Product type	Proposed model
Group Protection	$\hat{Y}= -0.022 + 0.23 * \text{Inflation}(-3) - 0.12 * \text{Lapse}(-2) + 0.69 * \text{Lapse}(-4) + 0.02 * \text{IBEX35}(-1) - 1.31 * \Delta\text{RFR} + 1.50 * \Delta\text{RFR}(-4) + 1.15 * \Delta\text{RMR}(-2)$
Group Savings	$\hat{Y} = \text{Exp}[-2.82 + 12.26 * \text{Inflation}(-3) - 17.56 * \text{Lapse}(-3) + 1.10 * \text{IBEX35}(-2)]$
Individual Protection	$\hat{Y} = -0.020 - 0.27 * \text{Inflation}(-1) - 0.24 * \text{Inflation}(-3) + 0.63 * \text{Lapse}(-1) + 0.53 * \text{Lapse}(-4) + 0.46 * \text{GDP}(-4) + 0.01 * \text{IBEX35}(-3) + 1.39 * \Delta\text{RMR}(-1)$
Individual Savings	$\hat{Y} = 0.003 - 0.034 * \text{Inflation}(-1) + 0.119 * \text{Inflation}(-2) + 0.596 * \text{Lapse}(-1) + 0.264 * \text{Lapse}(-4) - 0.001 * \text{IBEX35} - 0.064 * \text{IBEX35}(-2) + 0.212 * \Delta\text{RFR}(-2)$
Unit Linked	$\hat{Y} = \text{Exp}[-2.91 - 7.38 * \text{Inflation}(-1) + 2.93 * \text{Lapse}(-1) - 12.49 * \text{GDP}(-2) + 116.28 * \Delta\text{RFR}(-2) + 73.85 * \Delta\text{RMR}(-2) - 132.27 * \Delta\text{RMR}(-3)]$

Source: Own Elaboration

The proposed model for each product type is shown in Table 7, taking into account the link function which has been used for each model. Most of the models are made up of lagged variables, as only three "unlagged" variables have been incorporated, although it is true that the majority (80%) of included variables are lagged. This may indicate the importance of taking into account the time which policyholders need to react and take decisions. Furthermore, lagged variables of inflation and previous lapse rates are the only variables which are included in every model. The inclusion of lagged lapse rates indicates that lapse rates follow a trend, while the incorporation of inflation as a significant driver of lapse rates is not surprising as it is a very important indicator of economic growth and changing environments. Moreover, it is curious to see that the only model which does not incorporate the stock index IBEX35's returns is the Unit-Linked product category, whose returns are directly linked with equity. ΔRFR is considered in Group Protection, Individual Savings and Unit Linked while ΔRMR is considered in Group Protection, Individual Protection and Unit Linked. Group Savings is the only product type which does not include ΔRMR or ΔRFR which is interesting as group policies are normally handled by professional investors and rational behavior would be expected. This behavior might be attributable to t to the lack of data quality for this product which has been mentioned and therefore, no comment can be made on this aspect. GDP growth is only considered in Individual Protection and Unit-Linked products. Noticeably, unemployment has not been included in any of the five models as a significant driver.

Although the insertion of more than one lag of a same variable does not induce multicollinearity in these specific models, at a conceptual level it does not allow the reader to fully view the relationship between the explanatory variable and the dependent variable, as the coefficients might counter act against each other. As seen in Table 7, some variables have been included more than once in a same model with different lags and have had different mathematical signs in their coefficients (e.g. Δ RFR in Group Protection). In order to fully understand the relationship of these variables with lapse rates, less significant additional lags of a same variable have been removed from the model. The obtained models are represented in Table 8.

Table 8: Simplified Proposed Models for Theoretical Interpretation

Product type	Modified model	Removed Variables
Group Protection	$\hat{Y}$= -0.006 + 0.106 * Inflation (-3) + 0.814 * Lapse(-4) + 0.016 * IBEX35 (-1) + 0.533 * Δ RFR (-4) + 0.892* Δ RMR (-2)	Lapse (-2) & Δ RFR
Group Savings	$\hat{Y}$ = Exp[-2.8191 + 12.2646* Inflation (-3) - 17.5665 * Lapse (-3) + 1.1045 * IBEX35 (-2)]	N/A
Individual Protection	$\hat{Y}$ = -0.009 - 0.4878 * Inflation (-1)+ 0.36 * Lapse (-1) + 0.36 * GDP (-4) + 0.009 * IBEX35 (-3) + 1.75 *Δ RMR (-1)	Inflation (-3) & Lapse (-4)
Individual Savings	$\hat{Y}$ =0.003 + 0.193 * Inflation (-2) + 0.826 * Lapse (-1) - 0.006 * IBEX35 + 0.146 * Δ RFR (-2)	Inflation (-1), Lapse (-4), & IBEX35 (-2)
Unit Linked	$\hat{Y}$ = Exp [-3.19 - 10.50 * Inflation (-1) + 2.68 * Lapse (-1) - 0.23 * GDP(-2) + 88.36*Δ RFR (-2)]	ΔRMR (-2) & ΔRMR (-3) [40]

[40] Both ΔRMR have been removed as they are less significant than ΔRFR. When modelled together they produce counter acting coefficients but if modelled individually, they both produce positive coefficients.

Source: Own Elaboration

It should be noted that the models presented in Table 8 table perform worse predictions than the ones estimated previously and have been elaborated for purely theoretical reasons.

Some findings which can be extracted from this table include:

- An increase in Δ RFR results in an increase of lapses for Savings products. For Savings products, Δ RFR is the differential between the risk free rate and the crediting rate.
- Furthermore, Δ RMR and Δ RFR are also positively correlated with lapses for Protection and Unit Linked products.

According to these findings, the Interest Rate Hypothesis proposed by (Dar and Dodds, 1989) would be accepted for this given set of data as policyholder seem to value interest rates or the differential between external and internal rates of return as a cost of opportunity for holding an insurance policy. As this cost of opportunity rises and the insurance policy loses competitiveness, policyholders lapse their contracts.

As mentioned previously, unemployment has not been included as a significant variable for any of the five models. Therefore, the Emergency Fund Hypothesis proposed by Outreville (1990) cannot be accepted nor it can be rejected as the effects proposed by unemployment may have been introduced through other variables such as Inflation or GDP growth. These two variables cannot be used to solely test the emergency fund hypothesis as they also might affect the interest rate hypothesis (e.g. decrease in interest rates may lead to an increase in inflation since individuals are able to borrow money at a cheaper rate).

Other findings include:

- Inflation drives lapses in a positive manner for Savings and Group Protection products as an increase in inflation would lead to an increase in lapses while the opposite occurs for Individual Protection and Unit-Linked products.
- IBEX35 returns are positively correlated with lapses group policies and individual protection, while it is negatively correlated for Individual Savings. It should be noted that the relative weight of IBEX35 is low in these four models, as the associated coefficients are pretty small in comparison to the coefficients of the rest of parameters.
- GDP is positively correlated with lapses for Individual protection while it is negatively correlated for Unit Linked.

No significant difference can be found between product groups or between group and individual policies.

Judging on coefficient sizes and previous analysis, lagged lapse rates, inflation rates and rates of return (external & internal) seem to be the most important drivers of lapse rates.

8. Conclusions

The objective of the present study was to gain an insight into the behavior of lapse rates and its inherent dynamics. This understanding is provided by the identification of the main variables which drive lapse rates and can be used to accurately predict lapses.

In order to reach this objective, the present study has been divided into two parts: The first part consisted of a theoretical approach, where main literature on lapse rates has been studied along with a brief description of the behavioral economics which can be applied to the understanding of lapse rates. The second part consisted of an empirical approach, where real lapse data from a company operating in the Spanish market was studied with the help of generalized linear models. Possible explanatory variables which were expected to have an influence on lapse rates were tested. Namely, the studied variables were: previous lapse rates, unemployment rate, growth of the gross domestic product, inflation rate, differential between market interest rate and crediting rate, and the differential between the risk free rate and the crediting rate

From these variables, inflation rate, lagged lapse rates and the two differentials between external and internal rates of return have proven to be the most important drivers of lapses in this study.

Considering the theoretical approach, two main hypotheses are theorized in empirical literature: the emergency fund hypothesis (Outreville, 1990) and the interest rate hypothesis (Dar and Dodds, 1989). The first declares that policyholders consider the surrender value which is available on lapsation as an emergency fund to be used in times of financial distress. The second states that interest rates are viewed as an opportunity cost for owning a life insurance policy. Through the empirical approach, the interest rate hypothesis which has been theorized by Dar and Dodds is accepted, as findings show that lapse rates increase when:

- Reference market rates (RMR) increase
- Risk free rates (RFR) increase
- The differential between RMR and crediting rate increases.
- The differential between RFR and crediting rate increases.

These four scenarios represent an increasing opportunity cost for the policyholder.

Findings cannot accept nor reject the emergency fund hypothesis as the unemployment rate, which is the variable which has been used throughout literature to test this hypothesis, has not been considered significant enough to be included in the study but variables such as the growth of the gross domestic product or inflation rate are included and these might capture part of the effects provided by the unemployment rate.

Limitations of the study & areas for future research

The most important limitation that has been encountered with the present study is in reference to the available data. Due to privacy reasons, lapse data was only available at an aggregate product level and not at a policyholder level, resulting in a small sample of data. For this reason, only economic and company specific explanatory variables could be included in the study. These variables allow us capture a great part of the dynamics of lapses, but an important part still remains unexplained. It is possible that this unexplained component is due to idiosyncratic shocks or the small amount of data, but it is believed that if the study is approached with data at a policyholder level, results will improve as policyholder specific characteristics such as age, gender, policy age, etc. along with interactions between these variables and the economic variables can be included to the study as well.

The present study demonstrates that a large proportion of lapse rates can be explained with economic and company specific variables, but it leaves room for improvement. An extension of the present study with policyholder lapse data is encouraged.

Another area for future research which has been discovered in the present paper is dynamic policyholder behavior, namely dynamic lapses. As this concept is still fairly new and is arising a lot of interest the past few years, it is a very promising subject. Retaking the dynamic lapse model which was presented in Section 4.3 and currently used by most insurance companies: the proposal of a theoretical formula which is able to fit empirical data to the six dynamic lapse parameters of this model would be a very interesting area for future research as no explicit methodology is found in literature.

9. Appendices

Appendix A: SAS Script

Data Import

```
/*Import Data & Sorted by Product type*/
%web_drop_table(WORK.IMPORT);

FILENAME REFFILE '/folders/myfolders/Data_NEW.xlsx';

PROC IMPORT DATAFILE=REFFILE
	DBMS=XLSX
	OUT=WORK.IMPORT;
	GETNAMES=YES;
RUN;

PROC CONTENTS DATA=WORK.IMPORT; RUN;

%web_open_table(WORK.IMPORT);

proc sort data=work.import;
by product_type;
run;
```

Selection of Explanatory Variables

```
/*******Individual Savings******/

/*Backward CV*/

proc glmselect data=work.import plots=criteria seed=968843041;
where product_type="Individual Savings" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=backward (select=CV);
run;

proc reg data=work.import;
where product_type="Individual Savings";
model lapse_ratio= Unemployment   Unemployment_lag1  Inflation_lag2  Inflation_lag4
LAPSE_RATIO_lag1  LAPSE_RATIO_lag2  GDP_Growth_lag1  IBEX35Returns_lag1
DELTA_LAG3  DELTA_LAG4   / tol vif collin;
run;

/*Forward CV*/

proc glmselect data=work.import plots=criteria seed=968843041;
where product_type="Individual Savings" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
```

```
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=forward (select=CV);
run;

proc reg data=work.import;
where product_type="Individual Savings";
model lapse_ratio= Unemployment   Unemployment_lag2  Unemployment_lag3
  Unemployment_lag4  Inflation_lag2  LAPSE_RATIO_lag1  LAPSE_RATIO_lag2
  GDP_Growth_lag1  IBEX35Returns_lag2  DELTA_LAG2  DELTA_LAG4    / tol vif collin;
run;

/*Stepwise CV*/

proc glmselect data=work.import plots=criteria seed=968843041;
where product_type="Individual Savings" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=stepwise (select=CV);
run;

proc reg data=work.import;
where product_type="Individual Savings";
model lapse_ratio= Unemployment   Unemployment_lag1  Inflation_lag2  Inflation_lag4
LAPSE_RATIO_lag1  LAPSE_RATIO_lag2  GDP_Growth_lag1  IBEX35Returns_lag1
DELTA_LAG1  DELTA_LAG2  DELTA_LAG3  DELTA_LAG4    / tol vif collin;
run;

/*Backward Adjusted R-squared*/

proc glmselect data=work.import plots=criteria ;
where product_type="Individual Savings" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=backward (select=ADJRSQ);
run;

proc reg data=work.import;
where product_type="Individual Savings";
model lapse_ratio= Unemployment   Unemployment_lag1  Unemployment_lag2
Unemployment_lag3  Unemployment_lag4  Inflation_lag1  Inflation_lag2  Inflation_lag4
LAPSE_RATIO_lag1  LAPSE_RATIO_lag2  LAPSE_RATIO_lag3  LAPSE_ratio_lag4
GDP_Growth  GDP_Growth_lag1  GDP_Growth_lag2  GDP_Growth_lag3
 IBEX35Returns_lag2  DELTA_LAG2  DELTA_LAG4  DELTA_RMR  DELTA_RMR_lag1
   / tol vif collin;
run;

/*Forward Adjusted R-squared*/

proc glmselect data=work.import plots=criteria ;
where product_type="Individual Savings" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
```

```
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=forward (select=ADJRSQ);
run;

proc reg data=work.import;
where product_type="Individual Savings";
model lapse_ratio= Inflation   Inflation_lag2  Inflation_lag3  LAPSE_RATIO_lag1
LAPSE_RATIO_lag2  LAPSE_ratio_lag4
IBEX35Returns  IBEX35Returns_lag2  IBEX35Returns_lag3  DELTA  DELTA_LAG1
DELTA_LAG2  DELTA_LAG3  DELTA_LAG4    / tol vif collin;
run;

/*Stepwise Adjusted R-squared*/

proc glmselect data=work.import plots=criteria ;
where product_type="Individual Savings" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=stepwise (select=ADJRSQ);
run;

proc reg data=work.import;
where product_type="Individual Savings";
model lapse_ratio= Inflation   Inflation_lag2  Inflation_lag3  LAPSE_RATIO_lag1
LAPSE_RATIO_lag2  LAPSE_ratio_lag4
IBEX35Returns  IBEX35Returns_lag2  IBEX35Returns_lag3  DELTA  DELTA_LAG1
DELTA_LAG2  DELTA_LAG3  DELTA_LAG4    / tol vif collin;
run;

/*Backward AICC*/

proc glmselect data=work.import plots=criteria ;
where product_type="Individual Savings" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=backward (select=AICC);
run;

proc reg data=work.import;
where product_type="Individual Savings";
model lapse_ratio= Unemployment   Unemployment_lag1  Unemployment_lag2
Unemployment_lag3  Unemployment_lag4  Inflation_lag1  Inflation_lag2  Inflation_lag4
LAPSE_RATIO_lag1  LAPSE_RATIO_lag2  LAPSE_RATIO_lag3  LAPSE_ratio_lag4
GDP_Growth GDP_Growth_lag1  IBEX35Returns_lag2  DELTA_LAG2  DELTA_LAG4
DELTA_RMR_lag1    / tol vif collin;
run;

/*Forward AICC*/
```

```
proc glmselect data=work.import plots=criteria ;
where product_type="Individual Savings" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=forward (select=AICC);
run;

proc reg data=work.import;
where product_type="Individual Savings";
model lapse_ratio= Inflation   Inflation_lag2  Inflation_lag3  LAPSE_RATIO_lag1
LAPSE_ratio_lag4  IBEX35Returns   IBEX35Returns_lag2  DELTA_LAG1  DELTA_LAG2
DELTA_LAG4    / tol vif collin;
run;

/*Stepwise AICC*/

proc glmselect data=work.import plots=criteria ;
where product_type="Individual Savings" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=stepwise (select=AICC);
run;

proc reg data=work.import;
where product_type="Individual Savings";
model lapse_ratio= Inflation   Inflation_lag2  Inflation_lag3  LAPSE_RATIO_lag1
LAPSE_RATIO_lag2  LAPSE_ratio_lag4
IBEX35Returns  IBEX35Returns_lag2  IBEX35Returns_lag3  DELTA  DELTA_LAG1
DELTA_LAG2  DELTA_LAG3  DELTA_LAG4    / tol vif collin;
run;

/*LASSO*/

proc glmselect data=work.import plots=criteria ;
where product_type="Individual Savings" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=lasso;
run;

proc reg data=work.import;
where product_type="Individual Savings";
model lapse_ratio= Inflation_lag1  Inflation_lag2
LAPSE_RATIO_lag1  LAPSE_RATIO_lag2  LAPSE_ratio_lag4  IBEX35Returns
IBEX35Returns_lag2  DELTA_LAG2    / tol vif collin;
run;
```

```
/*LAR*/

proc glmselect data=work.import plots=criteria ;
where product_type="Individual Savings" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=lar;
run;

proc reg data=work.import;
where product_type="Individual Savings";
model lapse_ratio= Inflation_lag1  Inflation_lag2
LAPSE_RATIO_lag1  LAPSE_RATIO_lag2  LAPSE_ratio_lag4  IBEX35Returns
IBEX35Returns_lag2  DELTA_LAG2    / tol vif collin;
run;

/*******Group Risk******/

/*Backward CV*/

proc glmselect data=work.import plots=criteria seed=968843041 ;
where product_type="Group Risk" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=backward (select=CV);
run;

proc reg data=work.import;
where product_type="Group Risk";
model lapse_ratio= Unemployment_lag1   Unemployment_lag2  Unemployment_lag3
Unemployment_lag4  Inflation  Inflation_lag1  Inflation_lag2  Inflation_lag3  Inflation_lag4
LAPSE_RATIO_lag1  LAPSE_RATIO_lag2  LAPSE_RATIO_lag3  LAPSE_ratio_lag4
GDP_Growth  GDP_Growth_lag1  GDP_Growth_lag2  GDP_Growth_lag3  GDP_Growth_lag4
 IBEX35Returns  IBEX35Returns_lag1  IBEX35Returns_lag2  IBEX35Returns_lag3
IBEX35Returns_lag4  DELTA  DELTA_LAG1  DELTA_LAG2  DELTA_LAG3  DELTA_LAG4
DELTA_RMR  DELTA_RMR_lag1  DELTA_RMR_lag2  DELTA_RMR_lag3  DELTA_RMR_lag4
   / tol vif collin;
run;

/*Forward CV*/

proc glmselect data=work.import plots=criteria seed=968843041;
where product_type="Group Risk" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=forward (select=CV) ;
run;
```

```
proc reg data=work.import;
where product_type="Group Risk";
model lapse_ratio= Inflation_lag2   LAPSE_RATIO_lag2  LAPSE_ratio_lag4  GDP_Growth_lag1
  GDP_Growth_lag2  IBEX35Returns_lag1  IBEX35Returns_lag4  DELTA_LAG2
DELTA_LAG4  DELTA_RMR_lag2  DELTA_RMR_lag4
   / tol vif collin;
run;

/*Stepwise CV*/

proc glmselect data=work.import plots=criteria seed=968843041;
where product_type="Group Risk" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=stepwise (select=CV);
run;

proc reg data=work.import;
where product_type="Group Risk";
model lapse_ratio= Inflation_lag2 LAPSE_RATIO_lag2 LAPSE_ratio_lag4 GDP_Growth_lag2
DELTA_LAG4 DELTA_RMR DELTA_RMR_lag2    / tol vif collin;
run;

/*Backward Adjusted R-squared*/

proc glmselect data=work.import plots=criteria ;
where product_type="Group Risk" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=backward (select=ADJRSQ) ;
run;

proc reg data=work.import;
where product_type="Group Risk";
model lapse_ratio= Unemployment   Unemployment_lag1  Unemployment_lag2
Unemployment_lag3  Unemployment_lag4  Inflation_lag1  Inflation_lag2  Inflation_lag4
LAPSE_RATIO_lag1  LAPSE_RATIO_lag2  LAPSE_RATIO_lag3  LAPSE_ratio_lag4
GDP_Growth  GDP_Growth_lag1  GDP_Growth_lag2  GDP_Growth_lag3
 IBEX35Returns_lag2  DELTA_LAG2  DELTA_LAG4  DELTA_RMR  DELTA_RMR_lag1
  / tol vif collin;
run;
/*Forward Adjusted R-squared*/

proc glmselect data=work.import plots=criteria ;
where product_type="Group Risk" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
```

```
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=forward (select=ADJRSQ);
run;

proc reg data=work.import;
where product_type="Group Risk";
model lapse_ratio= Inflation   Inflation_lag2  LAPSE_RATIO_lag2  LAPSE_ratio_lag4
IBEX35Returns_lag1  IBEX35Returns_lag3  DELTA  DELTA_LAG4  DELTA_RMR_lag2
DELTA_RMR_lag3    / tol vif collin;
run;

/*Stepwise Adjusted R-squared*/

proc glmselect data=work.import plots=criteria ;
where product_type="Group Risk" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=stepwise (select=ADJRSQ);
run;

proc reg data=work.import;
where product_type="Group Risk";
model lapse_ratio= Inflation   Inflation_lag2  LAPSE_RATIO_lag2
LAPSE_ratio_lag4  IBEX35Returns_lag1  IBEX35Returns_lag3  DELTA  DELTA_LAG4
DELTA_RMR_lag2    / tol vif collin;
run;

/*Backward AICC*/

proc glmselect data=work.import plots=criteria ;
where product_type="Group Risk" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=backward (select=AICC);
run;

proc reg data=work.import;
where product_type="Group Risk";
model lapse_ratio= Inflation_lag3   LAPSE_RATIO_lag2
LAPSE_ratio_lag4  IBEX35Returns_lag1  DELTA  DELTA_LAG4  DELTA_RMR_lag2
   / tol vif collin;
run;

/*Forward AICC*/

proc glmselect data=work.import plots=criteria ;
where product_type="Group Risk" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
```

```
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=forward (select=AICC);
run;

proc reg data=work.import;
where product_type="Group Risk";
model lapse_ratio= Inflation   LAPSE_RATIO_lag2  LAPSE_ratio_lag4
IBEX35Returns_lag1  DELTA  DELTA_LAG4  DELTA_RMR_lag2  DELTA_RMR_lag3
   / tol vif collin;
run;

/*Stepwise AICC*/

proc glmselect data=work.import plots=criteria ;
where product_type="Group Risk" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4
          / selection=stepwise (select=AICC);
run;

proc reg data=work.import;
where product_type="Group Risk";
model lapse_ratio= Inflation   Inflation_lag2  LAPSE_RATIO_lag2
LAPSE_ratio_lag4  IBEX35Returns_lag1  DELTA  DELTA_LAG4  DELTA_RMR_lag2
   / tol vif collin;
run;

/*LASSO*/

proc glmselect data=work.import plots=criteria ;
where product_type="Group Risk" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=lasso;
run;

proc reg data=work.import;
where product_type="Group Risk";
model lapse_ratio= LAPSE_ratio_lag4   DELTA_RMR_lag3    / tol vif collin;
run;

/*LAR*/

proc glmselect data=work.import plots=criteria ;
where product_type="Group Risk" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
```

```
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=lar;
run;

proc reg data=work.import;
where product_type="Group Risk";
model lapse_ratio= Inflation_lag1  Inflation_lag2
LAPSE_RATIO_lag1  LAPSE_RATIO_lag2  LAPSE_ratio_lag4  IBEX35Returns
IBEX35Returns_lag2  DELTA_LAG2   / tol vif collin;
run;

/*******Group Savings******/

/*Backward CV*/

proc glmselect data=work.import plots=criteria seed=968843041 ;
where product_type="Group Savings" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=backward (select=CV);
run;

proc reg data=work.import;
where product_type="Group Savings";
model lapse_ratio= Unemployment   Unemployment_lag1  Unemployment_lag3
Unemployment_lag4  Inflation_lag2  Inflation_lag4  LAPSE_RATIO_lag3  GDP_Growth_lag1
GDP_Growth_lag3    IBEX35Returns  IBEX35Returns_lag1  IBEX35Returns_lag2
IBEX35Returns_lag3  IBEX35Returns_lag4  DELTA  DELTA_LAG1  DELTA_LAG2
DELTA_LAG4  DELTA_RMR_lag2  DELTA_RMR_lag3  DELTA_RMR_lag4    / tol vif collin;
run;

/*Forward CV*/

proc glmselect data=work.import plots=criteria seed=968843041;
where product_type="Group Savings" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=forward (select=CV) ;
run;

proc reg data=work.import;
where product_type="Group Savings";
model lapse_ratio= Unemployment   Unemployment_lag1  Unemployment_lag2
Unemployment_lag3  LAPSE_RATIO_lag3     IBEX35Returns_lag2
/ tol vif collin;
run;

/*Stepwise CV*/

proc glmselect data=work.import plots=criteria seed=968843041;
where product_type="Group Savings" ;
```

```
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=stepwise (select=CV);
run;

proc reg data=work.import;
where product_type="Group Savings";
model lapse_ratio= Unemployment   Unemployment_lag1  Unemployment_lag2
Unemployment_lag3  LAPSE_RATIO_lag3     IBEX35Returns_lag2
/ tol vif collin;
run;

/*Backward Adjusted R-squared*/

proc glmselect data=work.import plots=criteria ;
where product_type="Group Savings" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=backward (select=ADJRSQ) ;
run;

proc reg data=work.import;
where product_type="Group Savings";
model lapse_ratio= Unemployment   Unemployment_lag1  Unemployment_lag3
Unemployment_lag4  Inflation_lag2  Inflation_lag4  LAPSE_RATIO_lag3  GDP_Growth_lag1
GDP_Growth_lag3     IBEX35Returns  IBEX35Returns_lag1  IBEX35Returns_lag2
IBEX35Returns_lag3  IBEX35Returns_lag4  DELTA  DELTA_LAG1  DELTA_LAG2
DELTA_LAG3  DELTA_RMR  DELTA_RMR_lag2  DELTA_RMR_lag3  DELTA_RMR_lag4
/ tol vif collin;
run;

/*Forward Adjusted R-squared*/

proc glmselect data=work.import plots=criteria ;
where product_type="Group Savings" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=forward (select=ADJRSQ);
run;

proc reg data=work.import;
where product_type="Group Savings";
model lapse_ratio= Inflation_lag3   Inflation_lag4  LAPSE_RATIO_lag3     IBEX35Returns_lag2
/ tol vif collin;
run;


/*Stepwise Adjusted R-squared*/
```

```
proc glmselect data=work.import plots=criteria ;
where product_type="Group Savings" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=stepwise (select=ADJRSQ);
run;

proc reg data=work.import;
where product_type="Group Savings";
model lapse_ratio= inflation_lag3  Inflation_lag4  LAPSE_RATIO_lag3    IBEX35Returns_lag2 /
tol vif collin;
run;

/*Backward AICC*/

proc glmselect data=work.import plots=criteria ;
where product_type="Group Savings" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=backward (select=AICC);
run;

proc reg data=work.import;
where product_type="Group Savings";
model lapse_ratio= Unemployment   Unemployment_lag1  LAPSE_RATIO_lag3
IBEX35Returns_lag2  DELTA_LAG2  DELTA_LAG3 / tol vif collin;
run;

/*Forward AICC*/

proc glmselect data=work.import plots=criteria ;
where product_type="Group Savings" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=forward (select=AICC);
run;

proc reg data=work.import;
where product_type="Group Savings";
model lapse_ratio= Inflation_lag3   LAPSE_RATIO_lag3    IBEX35Returns_lag2
/ tol vif collin;
run;

/*Stepwise AICC*/

proc glmselect data=work.import plots=criteria ;
where product_type="Group Savings" ;
```

```
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=stepwise (select=AICC);
run;

proc reg data=work.import;
where product_type="Group Savings";
model lapse_ratio= Inflation_lag3   LAPSE_RATIO_lag3    IBEX35Returns_lag2
/ tol vif collin;
run;

/*LASSO*/

proc glmselect data=work.import plots=criteria ;
where product_type="Group Savings" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4
            / selection=lasso;
run;

proc reg data=work.import;
where product_type="Group Savings";
model lapse_ratio=   / tol vif collin;
run;

/*LAR*/

proc glmselect data=work.import plots=criteria ;
where product_type="Group Savings" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=lar;
run;


proc reg data=work.import;
where product_type="Group Savings";
model lapse_ratio=   / tol vif collin;
run;

/*******Individual Risk******/

/*Backward CV*/

proc glmselect data=work.import plots=criteria SEED=928728001 ;
where product_type="Individual Risk" ;
```

```
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=backward (select=CV);
run;

proc reg data=work.import;
where product_type="Individual Risk";
model lapse_ratio= Unemployment   Unemployment_lag1  Unemployment_lag2
Unemployment_lag3 Unemployment_lag4  Inflation  Inflation_lag1  Inflation_lag2  Inflation_lag3
Inflation_lag4  LAPSE_RATIO_lag1 LAPSE_RATIO_lag2  LAPSE_RATIO_lag3
LAPSE_ratio_lag4  GDP_Growth  GDP_Growth_lag1  GDP_Growth_lag2 GDP_Growth_lag3
GDP_Growth_lag4  IBEX35Returns  IBEX35Returns_lag1  IBEX35Returns_lag2
IBEX35Returns_lag3 IBEX35Returns_lag4  DELTA  DELTA_LAG1  DELTA_LAG2
DELTA_LAG3  DELTA_LAG4  DELTA_RMR  DELTA_RMR_lag1  DELTA_RMR_lag2
DELTA_RMR_lag3  DELTA_RMR_lag4    / tol vif collin;
run;

/*Forward CV*/

proc glmselect data=work.import plots=criteria SEED=928728001;
where product_type="Individual Risk" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=forward (select=CV) ;
run;

proc reg data=work.import;
where product_type="Individual Risk";
model lapse_ratio= Inflation_lag1   Inflation_lag3  LAPSE_RATIO_lag1  LAPSE_ratio_lag4
GDP_Growth_lag4  IBEX35Returns_lag3  DELTA_RMR_lag1  DELTA_RMR_lag2
   / tol vif collin;
run;

/*Stepwise CV*/

proc glmselect data=work.import plots=criteria SEED=928728001;
where product_type="Individual Risk" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=stepwise (select=CV);
run;

proc reg data=work.import;
where product_type="Individual Risk";
model lapse_ratio= Inflation_lag1   Inflation_lag3  LAPSE_RATIO_lag1
LAPSE_ratio_lag4  GDP_Growth_lag4  IBEX35Returns_lag3  DELTA_RMR_lag1
   / tol vif collin;
run;
```

```
/*Backward Adjusted R-squared*/

proc glmselect data=work.import plots=criteria ;
where product_type="Individual Risk" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=backward (select=ADJRSQ) ;
run;

proc reg data=work.import;
where product_type="Individual Risk";
model lapse_ratio= Unemployment   Unemployment_lag2  Unemployment_lag3
Unemployment_lag4  Inflation_lag1 Inflation_lag2  Inflation_lag3  LAPSE_RATIO_lag1
LAPSE_RATIO_lag3  LAPSE_ratio_lag4  GDP_Growth GDP_Growth_lag1  GDP_Growth_lag4
IBEX35Returns  IBEX35Returns_lag1  IBEX35Returns_lag2 IBEX35Returns_lag4  DELTA
DELTA_LAG3  DELTA_LAG4  DELTA_RMR  DELTA_RMR_lag1  DELTA_RMR_lag3
  / tol vif collin;
run;

/*Forward Adjusted R-squared*/

proc glmselect data=work.import plots=criteria ;
where product_type="Individual Risk" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=forward (select=ADJRSQ);
run;

proc reg data=work.import;
where product_type="Individual Risk";
model lapse_ratio= Inflation_lag1   LAPSE_RATIO_lag1  LAPSE_RATIO_lag3
LAPSE_ratio_lag4  DELTA  DELTA_RMR  DELTA_RMR_lag1  DELTA_RMR_lag2
DELTA_RMR_lag3    / tol vif collin;
run;

/*Stepwise Adjusted R-squared*/

proc glmselect data=work.import plots=criteria ;
where product_type="Individual Risk" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=stepwise (select=ADJRSQ);
run;

proc reg data=work.import;
where product_type="Individual Risk";
```

```
model lapse_ratio= Unemployment_lag1  Unemployment_lag3 Inflation_lag1 Inflation_lag2
Inflation_lag3 LAPSE_RATIO_lag1 LAPSE_ratio_lag4 GDP_Growth_lag1 IBEX35Returns
IBEX35Returns_lag2 IBEX35Returns_lag4 DELTA_LAG2 DELTA_LAG4 DELTA_RMR
DELTA_RMR_lag1   / tol vif collin;
run;

/*Backward AICC*/

proc glmselect data=work.import plots=criteria ;
where product_type="Individual Risk" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3 unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1 gdp_growth_lag2 gdp_growth_lag3 gdp_growth_lag4  ibex35returns
ibex35returns_lag1 ibex35returns_lag2 ibex35returns_lag3 ibex35returns_lag4  delta
delta_lag1 delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4 / selection=backward (select=AICC);
run;

proc reg data=work.import;
where product_type="Individual Risk";
model lapse_ratio= Unemployment  Unemployment_lag2 Unemployment_lag3
Unemployment_lag4
Inflation_lag1 Inflation_lag2 Inflation_lag3 GDP_Growth_lag1 IBEX35Returns
IBEX35Returns_lag2 DELTA_LAG3 DELTA_LAG4 DELTA_RMR DELTA_RMR_lag1
DELTA_RMR_lag3   / tol vif collin;
run;

/*Forward AICC*/

proc glmselect data=work.import plots=criteria ;
where product_type="Individual Risk" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3 unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1 gdp_growth_lag2 gdp_growth_lag3 gdp_growth_lag4  ibex35returns
ibex35returns_lag1 ibex35returns_lag2 ibex35returns_lag3 ibex35returns_lag4  delta
delta_lag1 delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4 / selection=forward (select=AICC);
run;

proc reg data=work.import;
where product_type="Individual Risk";
model lapse_ratio=Inflation_lag1  LAPSE_RATIO_lag1 LAPSE_RATIO_lag3
LAPSE_ratio_lag4 DELTA DELTA_RMR_lag2   / tol vif collin;
run;

/*Stepwise AICC*/

proc glmselect data=work.import plots=criteria ;
where product_type="Individual Risk" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3 unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1 gdp_growth_lag2 gdp_growth_lag3 gdp_growth_lag4  ibex35returns
ibex35returns_lag1 ibex35returns_lag2 ibex35returns_lag3 ibex35returns_lag4  delta
delta_lag1 delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4 / selection=stepwise (select=AICC);
run;
```

```
proc reg data=work.import;
where product_type="Individual Risk";
model lapse_ratio= Inflation_lag1   LAPSE_RATIO_lag3  DELTA  DELTA_RMR_lag2
   / tol vif collin;
run;

/*LASSO*/

proc glmselect data=work.import plots=criteria ;
where product_type="Individual Risk" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=lasso;
run;

/*LAR*/

proc glmselect data=work.import plots=criteria ;
where product_type="Individual Risk" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4
         / selection=lar;
run;

/*******Unit Linked******/

/*Backward CV*/

proc glmselect data=work.import plots=criteria SEED=968843041 ;
where product_type="Unit Linked" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=backward (select=CV);
run;

proc reg data=work.import;
where product_type="Unit Linked";
model lapse_ratio= Unemployment   Unemployment_lag1  Unemployment_lag2
Unemployment_lag3  Unemployment_lag4  Inflation
 Inflation_lag1  Inflation_lag2  Inflation_lag3  Inflation_lag4  LAPSE_RATIO_lag1
LAPSE_RATIO_lag2  LAPSE_RATIO_lag3   LAPSE_ratio_lag4  GDP_Growth
GDP_Growth_lag1  GDP_Growth_lag2  GDP_Growth_lag3  GDP_Growth_lag4
IBEX35Returns  IBEX35Returns_lag1 IBEX35Returns_lag2  IBEX35Returns_lag3
IBEX35Returns_lag4 DELTA  DELTA_LAG1  DELTA_LAG2  DELTA_LAG3  DELTA_LAG4
 DELTA_RMR  DELTA_RMR_lag1  DELTA_RMR_lag2  DELTA_RMR_lag3
DELTA_RMR_lag4 / tol vif collin;
```

```
run;

/*Forward CV*/

proc glmselect data=work.import plots=criteria SEED=968843041;
where product_type="Unit Linked" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4 / selection=forward (select=CV)  ;
run;

proc reg data=work.import;
where product_type="Unit Linked";
model lapse_ratio= Inflation_lag1   LAPSE_RATIO_lag1  GDP_Growth_lag4
IBEX35Returns_lag3  DELTA_LAG2  DELTA_RMR_lag1  DELTA_RMR_lag3
  / tol vif collin;
run;

/*Stepwise CV*/

proc glmselect data=work.import plots=criteria SEED=968843041;
where product_type="Unit Linked" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4 / selection=stepwise (select=CV);
run;

proc reg data=work.import;
where product_type="Unit Linked";
model lapse_ratio= Inflation_lag1   LAPSE_RATIO_lag1  GDP_Growth_lag4
IBEX35Returns_lag3  DELTA_LAG2  DELTA_RMR_lag1  DELTA_RMR_lag3
  / tol vif collin;
run;

/*Backward Adjusted R-squared*/

proc glmselect data=work.import plots=criteria ;
where product_type="Unit Linked" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4 / selection=backward (select=ADJRSQ)  ;
run;

proc reg data=work.import;
where product_type="Unit Linked";
model lapse_ratio= Unemployment   Unemployment_lag2  Unemployment_lag4  Inflation_lag2
Inflation_lag4  LAPSE_RATIO_lag1  LAPSE_RATIO_lag3
```

```
LAPSE_ratio_lag4  GDP_Growth_lag1  GDP_Growth_lag3  GDP_Growth_lag4  IBEX35Returns
IBEX35Returns_lag1  IBEX35Returns_lag2  DELTA
 DELTA_LAG2  DELTA_RMR  DELTA_RMR_lag1  DELTA_RMR_lag3 / tol vif collin;
run;

/*Forward Adjusted R-squared*/

proc glmselect data=work.import plots=criteria ;
where product_type="Unit Linked" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4 / selection=forward (select=ADJRSQ);
run;

proc reg data=work.import;
where product_type="Unit Linked";
model lapse_ratio= Inflation_lag1   LAPSE_RATIO_lag1  GDP_Growth_lag2  DELTA_LAG2
DELTA_RMR_lag2  DELTA_RMR_lag3
   / tol vif collin;
run;

/*Stepwise Adjusted R-squared*/

proc glmselect data=work.import plots=criteria ;
where product_type="Unit Linked" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4 / selection=stepwise (select=ADJRSQ);
run;

proc reg data=work.import;
where product_type="Unit Linked";
model lapse_ratio= Inflation_lag1   LAPSE_RATIO_lag1  GDP_Growth_lag2  DELTA_LAG2
DELTA_RMR_lag2  DELTA_RMR_lag3
   / tol vif collin;
run;

/*Backward AICC*/

proc glmselect data=work.import plots=criteria ;
where product_type="Unit Linked" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4 / selection=backward (select=AICC);
run;

proc reg data=work.import;
where product_type="Unit Linked";
```

```
model lapse_ratio= Unemployment   Unemployment_lag2  Unemployment_lag4
Inflation_lag2  LAPSE_RATIO_lag1  GDP_Growth_lag1  IBEX35Returns_lag1
DELTA  DELTA_LAG2  DELTA_RMR_lag1  DELTA_RMR_lag3    / tol vif collin;
run;

/*Forward AICC*/

proc glmselect data=work.import plots=criteria ;
where product_type="Unit Linked" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4  / selection=forward (select=AICC);
run;

proc reg data=work.import;
where product_type="Unit Linked";
model lapse_ratio= Inflation_lag1   LAPSE_RATIO_lag1  GDP_Growth_lag2  DELTA_LAG2
DELTA_RMR_lag2  DELTA_RMR_lag3    / tol vif collin;
run;


/*Stepwise AICC*/

proc glmselect data=work.import plots=criteria ;
where product_type="Unit Linked" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4 / selection=stepwise (select=AICC);
run;


proc reg data=work.import;
where product_type="Unit Linked";
model lapse_ratio= Inflation_lag1   LAPSE_RATIO_lag1  GDP_Growth_lag2  DELTA_LAG2
DELTA_RMR_lag2  DELTA_RMR_lag3    / tol vif collin;
run;


/*LASSO*/

proc glmselect data=work.import plots=criteria ;
where product_type="Unit Linked" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4 / selection=lasso;
run;
```

```
/*LAR*/

proc glmselect data=work.import plots=criteria ;
where product_type="Unit Linked" ;
model lapse_ratio= unemployment unemployment_lag1 unemployment_lag2
unemployment_lag3  unemployment_lag4 inflation inflation_lag1 inflation_lag2 inflation_lag3
inflation_lag4 lapse_ratio_lag1 lapse_ratio_lag2 lapse_ratio_lag3 lapse_ratio_lag4 gdp_growth
gdp_growth_lag1  gdp_growth_lag2  gdp_growth_lag3 gdp_growth_lag4   ibex35returns
ibex35returns_lag1  ibex35returns_lag2 ibex35returns_lag3  ibex35returns_lag4   delta
delta_lag1  delta_lag2 delta_lag3 delta_lag4 delta_rmr delta_rmr_lag1 delta_rmr_lag2
delta_rmr_lag3 delta_rmr_lag4 / selection=lar;
run;
```

<u>Model Selection</u>

```
/* Macro to calculate Mean Absolute Percentage Eror and Root Mean Squared Error */
/* Outputs to data set, log, and macro variable */
%macro mape_rmse(
	dataset /* Data set which contains the actual and predicted values */,
	actual /* Variable which contains the actual or observed valued */,
	predicted /* Variable which contains the predicted value */
	);
%global mape rmse; /* Make the scope of the macro variables global */
data &dataset;
   retain square_error_sum abs_error_sum;
   set &dataset
      end=last /* Flag for the last observation */
      ;
   error = &actual - &predicted; /* Calculate simple error */
 if &actual ne 0 then do;   abs_error_per=abs(error)/&actual; end; else do abs_error_per=0; end;
/*Calculate absolute percentual error */
   square_error = error * error; /* error^2 */
   if _n_ eq 1 then do;
      /* Initialize the sums */
      square_error_sum = square_error;
      abs_error_sum = abs_error_per;
      end;
   else do;
      /* Add to the sum */
      square_error_sum = square_error_sum + square_error;
      abs_error_sum = abs_error_sum + abs_error_per;
   end;
   if last then do;
      /* Calculate RMSE and MAPE and store in SAS data set. */
      mape = abs_error_sum/_n_;
      rmse = sqrt(square_error_sum/_n_);
      /* Write to SAS log */
      put 'NOTE: ' mape= rmse=;
      /* Store in SAS macro variables */
      call symput('mape', put(mape, 20.10));
      call symput('rmse', put(rmse, 20.10));
      end;
run;
%mend;

/* Model Selection */

/*Group Protection */
```

```
proc reg data=work.import;
where product_type="Group Risk";
model lapse_ratio= Inflation_lag3   LAPSE_RATIO_lag2
LAPSE_ratio_lag4  IBEX35Returns_lag1  DELTA  DELTA_LAG4  DELTA_RMR_lag2
   /  tol vif collin;
run;


proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Group Risk";
model lapse_ratio= Inflation_lag3   LAPSE_RATIO_lag2  LAPSE_ratio_lag4
IBEX35Returns_lag1  DELTA  DELTA_LAG4  DELTA_RMR_lag2    /   dist=normal link=identity
solution ;
 output out=gmxout1 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout1, lapse_ratio, pred);


proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Group Risk";
model lapse_ratio= Inflation_lag3   LAPSE_RATIO_lag2  LAPSE_ratio_lag4
IBEX35Returns_lag1  DELTA  DELTA_LAG4  DELTA_RMR_lag2    /  dist=normal link=log
solution ;
 output out=gmxout2 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout2, lapse_ratio, pred);

proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Group Risk";
model lapse_ratio= Inflation_lag3   LAPSE_RATIO_lag2  LAPSE_ratio_lag4
IBEX35Returns_lag1  DELTA  DELTA_LAG4  DELTA_RMR_lag2    /   dist=normal link=logit
solution ;
 output out=gmxout3 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout3, lapse_ratio, pred);

proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Group Risk";
model lapse_ratio= Inflation_lag3   LAPSE_RATIO_lag2  LAPSE_ratio_lag4
IBEX35Returns_lag1  DELTA  DELTA_LAG4  DELTA_RMR_lag2    /   dist=normal link=cloglog
solution ;
 output out=gmxout4 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout4, lapse_ratio, pred);


proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Group Risk";
model lapse_ratio= Inflation_lag3   LAPSE_RATIO_lag2  LAPSE_ratio_lag4
IBEX35Returns_lag1  DELTA  DELTA_LAG4  DELTA_RMR_lag2    / dist=beta link=identity
solution ;
 output out=gmxout5 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout5, lapse_ratio, pred);
```

```
proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Group Risk";
model lapse_ratio= Inflation_lag3   LAPSE_RATIO_lag2  LAPSE_ratio_lag4
IBEX35Returns_lag1  DELTA  DELTA_LAG4  DELTA_RMR_lag2    /   dist=beta link=log
solution ;
 output out=gmxout6 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout6, lapse_ratio, pred);

proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Group Risk";
model lapse_ratio= Inflation_lag3   LAPSE_RATIO_lag2  LAPSE_ratio_lag4
IBEX35Returns_lag1  DELTA  DELTA_LAG4  DELTA_RMR_lag2    /  dist=beta link=logit
solution ;
 output out=gmxout7 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout7, lapse_ratio, pred);

proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Group Risk";
model lapse_ratio= Inflation_lag3   LAPSE_RATIO_lag2  LAPSE_ratio_lag4
IBEX35Returns_lag1  DELTA  DELTA_LAG4  DELTA_RMR_lag2    / dist=beta link=cloglog
solution ;
 output out=gmxout8 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout8, lapse_ratio, pred);



/*Group Savings */

proc reg data=work.import;
where product_type="Group Savings";
model lapse_ratio= Inflation_lag3   LAPSE_RATIO_lag3     IBEX35Returns_lag2
/ tol vif collin;
run;

proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Group Savings";
model lapse_ratio= Inflation_lag3   LAPSE_RATIO_lag3     IBEX35Returns_lag2   /
dist=normal link=identity solution ;
 output out=gmxout1 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout1, lapse_ratio, pred);


proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Group Savings";
model lapse_ratio= Inflation_lag3   LAPSE_RATIO_lag3     IBEX35Returns_lag2   /  dist=normal
link=log solution ;
 output out=gmxout2 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout2, lapse_ratio, pred);
```

```
proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Group Savings";
model lapse_ratio= Inflation_lag3   LAPSE_RATIO_lag3     IBEX35Returns_lag2   /
dist=normal link=logit solution ;
 output out=gmxout3 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout3, lapse_ratio, pred);

proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Group Savings";
model lapse_ratio= Inflation_lag3   LAPSE_RATIO_lag3     IBEX35Returns_lag2    /
dist=normal link=cloglog solution ;
 output out=gmxout4 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout4, lapse_ratio, pred);


proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Group Savings";
model lapse_ratio= Inflation_lag3   LAPSE_RATIO_lag3     IBEX35Returns_lag2   / dist=beta
link=identity solution ;
 output out=gmxout5 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout5, lapse_ratio, pred);


proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Group Savings";
model lapse_ratio= Inflation_lag3   LAPSE_RATIO_lag3     IBEX35Returns_lag2  /   dist=beta
link=log solution ;
 output out=gmxout6 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout6, lapse_ratio, pred);

proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Group Savings";
model lapse_ratio= Inflation_lag3   LAPSE_RATIO_lag3     IBEX35Returns_lag2   /  dist=beta
link=logit solution ;
 output out=gmxout7 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout7, lapse_ratio, pred);

proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Group Savings";
model lapse_ratio= Inflation_lag3   LAPSE_RATIO_lag3     IBEX35Returns_lag2  / dist=beta
link=cloglog solution ;
 output out=gmxout8 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout8, lapse_ratio, pred);

/*Individual Protection*/

proc reg data=work.import;
```

```
where product_type="Individual Risk";
model lapse_ratio= Inflation_lag1   Inflation_lag3  LAPSE_RATIO_lag1
LAPSE_ratio_lag4  GDP_Growth_lag4  IBEX35Returns_lag3  DELTA_RMR_lag1
/ tol vif collin;
run;

proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Individual Risk";
model lapse_ratio= Inflation_lag1   Inflation_lag3  LAPSE_RATIO_lag1
LAPSE_ratio_lag4  GDP_Growth_lag4  IBEX35Returns_lag3  DELTA_RMR_lag1  /
dist=normal link=identity solution ;
 output out=gmxout1 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;


proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Individual Risk";
model lapse_ratio= Inflation_lag1   Inflation_lag3  LAPSE_RATIO_lag1
LAPSE_ratio_lag4  GDP_Growth_lag4  IBEX35Returns_lag3  DELTA_RMR_lag1  /
dist=normal link=log solution ;
 output out=gmxout2 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout2, lapse_ratio, pred);

proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Individual Risk";
model lapse_ratio= Inflation_lag1   Inflation_lag3  LAPSE_RATIO_lag1
LAPSE_ratio_lag4  GDP_Growth_lag4  IBEX35Returns_lag3  DELTA_RMR_lag1  /
dist=normal link=logit solution ;
 output out=gmxout3 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout3, lapse_ratio, pred);

proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Individual Risk";
model lapse_ratio= Inflation_lag1   Inflation_lag3  LAPSE_RATIO_lag1
LAPSE_ratio_lag4  GDP_Growth_lag4  IBEX35Returns_lag3  DELTA_RMR_lag1    /
dist=normal link=cloglog solution ;
 output out=gmxout4 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout4, lapse_ratio, pred);


proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Individual Risk";
model lapse_ratio= Inflation_lag1   Inflation_lag3  LAPSE_RATIO_lag1
LAPSE_ratio_lag4  GDP_Growth_lag4  IBEX35Returns_lag3  DELTA_RMR_lag1   / dist=beta
link=identity solution ;
 output out=gmxout5 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout5, lapse_ratio, pred);


proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Individual Risk";
model lapse_ratio= Inflation_lag1   Inflation_lag3  LAPSE_RATIO_lag1
```

```
LAPSE_ratio_lag4  GDP_Growth_lag4  IBEX35Returns_lag3  DELTA_RMR_lag1 /   dist=beta
link=log solution ;
 output out=gmxout6 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout6, lapse_ratio, pred);

proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Individual Risk";
model lapse_ratio= Inflation_lag1   Inflation_lag3  LAPSE_RATIO_lag1
LAPSE_ratio_lag4  GDP_Growth_lag4  IBEX35Returns_lag3  DELTA_RMR_lag1  /  dist=beta
link=logit solution ;
 output out=gmxout7 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout7, lapse_ratio, pred);

proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Individual Risk";
model lapse_ratio= Inflation_lag1   Inflation_lag3  LAPSE_RATIO_lag1
LAPSE_ratio_lag4  GDP_Growth_lag4  IBEX35Returns_lag3  DELTA_RMR_lag1 / dist=beta
link=cloglog solution ;
 output out=gmxout8 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout8, lapse_ratio, pred);

/*Individual Savings*/

proc reg data=work.import;
where product_type="Individual Savings";
      model lapse_ratio= Inflation_lag1  Inflation_lag2
LAPSE_RATIO_lag1  LAPSE_ratio_lag4  IBEX35Returns  IBEX35Returns_lag2  DELTA_LAG2
/ tol vif collin;
run;

proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Individual Savings";
      model lapse_ratio= Inflation_lag1  Inflation_lag2
LAPSE_RATIO_lag1  LAPSE_ratio_lag4  IBEX35Returns  IBEX35Returns_lag2  DELTA_LAG2
/   dist=normal link=identity solution ;
 output out=gmxout1 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout1, lapse_ratio, pred);


proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Individual Savings";
      model lapse_ratio= Inflation_lag1  Inflation_lag2
LAPSE_RATIO_lag1  LAPSE_ratio_lag4  IBEX35Returns  IBEX35Returns_lag2  DELTA_LAG2
/  dist=normal link=log solution ;
 output out=gmxout2 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout2, lapse_ratio, pred);

proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Individual Savings";
model lapse_ratio= Inflation_lag1  Inflation_lag2
```

```
LAPSE_RATIO_lag1  LAPSE_ratio_lag4  IBEX35Returns  IBEX35Returns_lag2  DELTA_LAG2
/   dist=normal link=logit solution ;
 output out=gmxout3 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout3, lapse_ratio, pred);

proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Individual Savings";
model lapse_ratio= Inflation_lag1  Inflation_lag2
LAPSE_RATIO_lag1  LAPSE_ratio_lag4  IBEX35Returns  IBEX35Returns_lag2  DELTA_LAG2
/   dist=normal link=cloglog solution ;
 output out=gmxout4 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout4, lapse_ratio, pred);


proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Individual Savings";
model lapse_ratio= Inflation_lag1  Inflation_lag2
LAPSE_RATIO_lag1  LAPSE_ratio_lag4  IBEX35Returns  IBEX35Returns_lag2  DELTA_LAG2
/ dist=beta link=identity solution ;
 output out=gmxout5 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout5, lapse_ratio, pred);


proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Individual Savings";
model lapse_ratio= Inflation_lag1  Inflation_lag2
LAPSE_RATIO_lag1  LAPSE_ratio_lag4  IBEX35Returns  IBEX35Returns_lag2  DELTA_LAG2
/   dist=beta link=log solution ;
 output out=gmxout6 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout6, lapse_ratio, pred);

proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Individual Savings";
model lapse_ratio= Inflation_lag1  Inflation_lag2
LAPSE_RATIO_lag1  LAPSE_ratio_lag4  IBEX35Returns  IBEX35Returns_lag2  DELTA_LAG2
/  dist=beta link=logit solution ;
 output out=gmxout7 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout7, lapse_ratio, pred);

proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Individual Savings";
model lapse_ratio= Inflation_lag1  Inflation_lag2
LAPSE_RATIO_lag1  LAPSE_ratio_lag4  IBEX35Returns  IBEX35Returns_lag2  DELTA_LAG2
/ dist=beta link=cloglog solution ;
 output out=gmxout8 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout8, lapse_ratio, pred);
```

```
/*Unit-Linked*/


proc reg data=work.import;
where product_type="Unit Linked";
model lapse_ratio= Inflation_lag1   LAPSE_RATIO_lag1  GDP_Growth_lag2  DELTA_LAG2
DELTA_RMR_lag2  DELTA_RMR_lag3    / tol vif collin;
run;


proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Unit Linked";
model lapse_ratio= Inflation_lag1   LAPSE_RATIO_lag1  GDP_Growth_lag2  DELTA_LAG2
DELTA_RMR_lag2  DELTA_RMR_lag3 /   dist=normal link=identity solution ;
 output out=gmxout1 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout1, lapse_ratio, pred);


proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Unit Linked";
model lapse_ratio= Inflation_lag1   LAPSE_RATIO_lag1  GDP_Growth_lag2  DELTA_LAG2
DELTA_RMR_lag2  DELTA_RMR_lag3  /  dist=normal link=log solution ;
 output out=gmxout2 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout2, lapse_ratio, pred);

proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Unit Linked";
model lapse_ratio= Inflation_lag1   LAPSE_RATIO_lag1  GDP_Growth_lag2  DELTA_LAG2
DELTA_RMR_lag2  DELTA_RMR_lag3 /   dist=normal link=logit solution ;
 output out=gmxout3 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout3, lapse_ratio, pred);

proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Unit Linked";
model lapse_ratio= Inflation_lag1   LAPSE_RATIO_lag1  GDP_Growth_lag2  DELTA_LAG2
DELTA_RMR_lag2  DELTA_RMR_lag3  /   dist=normal link=cloglog solution ;
 output out=gmxout4 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout4, lapse_ratio, pred);


proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Unit Linked";
model lapse_ratio= Inflation_lag1   LAPSE_RATIO_lag1  GDP_Growth_lag2  DELTA_LAG2
DELTA_RMR_lag2  DELTA_RMR_lag3  / dist=beta link=identity solution ;
 output out=gmxout5 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout5, lapse_ratio, pred);


proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Unit Linked";
```

```
model lapse_ratio= Inflation_lag1   LAPSE_RATIO_lag1  GDP_Growth_lag2  DELTA_LAG2
DELTA_RMR_lag2  DELTA_RMR_lag3  /   dist=beta link=log solution ;
 output out=gmxout6 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout6, lapse_ratio, pred);

proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Unit Linked";
model lapse_ratio= Inflation_lag1   LAPSE_RATIO_lag1  GDP_Growth_lag2  DELTA_LAG2
DELTA_RMR_lag2  DELTA_RMR_lag3  /  dist=beta link=logit solution ;
 output out=gmxout7 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout7, lapse_ratio, pred);

proc glimmix data=work.import METHOD=MSPL plots=residualpanel;
where product_type="Unit Linked";
model lapse_ratio= Inflation_lag1   LAPSE_RATIO_lag1  GDP_Growth_lag2  DELTA_LAG2
DELTA_RMR_lag2  DELTA_RMR_lag3 / dist=beta link=cloglog solution ;
 output out=gmxout8 pred(ilink)=pred lcl(ilink)=lower ucl(ilink)=upper;
run;

%mape_rmse(gmxout8, lapse_ratio, pred);
```

Appendix B: LaTeX Script

Dynamic Lapse Formula

```
\usepackage{amsmath}

\begin{document}
\[
   AL=
\begin{cases}
   BL* (1+m),       &  \text{if Delta}\geq b \\
   BL,       &  \text{if c}\leq  Delta}\leq a \\
   BL* (1+d) ,& \text{if Delta}\leq e \\
   BL*(1+r) ,           & \text{otherwise}
\end{cases}
\]
\text{where}

\[
   r=
\begin{cases}
   m * (\frac{Delta-a}{b-a}),       &  \text{if Delta}\geq a \\
   d * (\frac{Delta-c}{e-c}),           & \text{otherwise}
\end{cases}
\]
\end{document}
```

Dynamic Lapse Example Table

```
\documentclass[12pt]{article}
\begin{document}

\begin{center}
 \begin{tabular}{||c c c c c c c||}
 \hline
Scenario & RMR & CR & Delta & Factor & BL & AL \\ [0.5ex]
 \hline\hline
 1 & 6\% & 3\% & 3\% & 2,2 & 10\% & 22\%\\
 \hline
 2 & 4\% & 5\% & -1\%  & 0,75 & 10\% & 7,5\% \\
 \hline
 3 & 2\% & 1,75\% & 0,25\%  & 1 & 10\% & 10\%\\
 \hline
 4 & 9\% & 2\% & 7\%  & 3 & 10\% & 30\% \\
 \hline
 5 & 2\% & 6\% & -4\%  & 0,25 & 10\% & 2,5\%\\
 \hline

\end{tabular}
\end{center}
\end{document}
```

Table Explanatory Variables

```
\documentclass{article}
\usepackage[utf8]{inputenc}
\setlength{\arrayrulewidth}{1mm}
\setlength{\tabcolsep}{2pt}
\renewcommand{\arraystretch}{1.5}
\centering
\small
\usepackage{amsmath}

\begin{document}

	\begin{tabular}{ |c c c c c c c|    }
		\hline
		\multicolumn{7}{|c|}{Possible Explanatory Variables} \\
		\hline
		Lapse(-1)  & Unemployment   & Inflation  & GDP & $\Delta$RMR &$\Delta$RFR
&IBEX35 \\
		Lapse(-2) & Unemployment(-1)  & Inflation(-1) & GDP(-1) & $\Delta$RMR(-1) &
$\Delta$RFR(-1) &IBEX35(-1) \\
		Lapse(-3) & Unemployment(-2) & Inflation(-2) & GDP(-2) & $\Delta$RMR(-2)
&$\Delta$RFR(-2) &IBEX35(-2) \\
		Lapse(-4) & Unemployment(-3) & Inflation(-3) & GDP(-3) & $\Delta$RMR(-3)
&$\Delta$RFR(-3) &IBEX35(-3) \\
		  & Unemployment (-4) & Inflation(-4) & GDP(-4) & $\Delta$RMR(-4)
&$\Delta$RFR(-4) &IBEX35(-4) \\
```

```
            \hline
        \end{tabular}
\end{document}
```

Appendix C: R Script

Predicted Individual Protection Graph

```
ir <- read.csv("K:/TFM Mohnish/Misc/IR.csv", sep="\t", dec=".", header=T)

attach(ir)

head(ir)

install.packages('ggplot2')

library(ggplot2)

X <- as.character(Actual.Lapse)

X <- as.double(gsub("%","", X))/100

PREDX<- as.character(Predicted.Lapse)

PREDX <- as.double(gsub("%","", PREDX))

UC.Band

LC.Band

op <- par(cex.main = 1.5, mar = c(5, 6, 4, 5) + 0.1, mgp = c(3.5, 1, 0), cex.lab = 1.5,

 font.lab = 2, cex.axis = 1.5, bty = "n", las = 1)

plot(1,1, xlab = "", ylab = "", type = "n", xlim = c(-50, max(Q_COUNT)), ylim = c(0.01, 0.06), axes
= FALSE)

axis(1)

axis(2)

polygon(c(Q_COUNT, rev(Q_COUNT)), c(UC.Band, rev(LC.Band)), col = "lightsteelblue", border
= NA, lwd=30)

lines(Q_COUNT, PREDX, lwd =3, col="purple")

lines(Q_COUNT, X, lwd =3, col="green")

mtext("Quarter", side = 1, line = 2.5, cex = 1.5)
```

```
mtext("Lapse Ratio", side = 2, line = 3.7, cex = 1.5, las = 0)

legend("topright", c("Observed Lapse","Predicted Lapse","Confidence Interval"),
lty=c(1,1,1),lwd=c(3,3,5),col=c("green","purple","lightsteelblue"))
```

10. Bibliography

Bacinello, A. R. (2003). *Fair valuation of a guaranteed life insurance participating contract embedding a surrender option.* Journal of risk and insurance, 70(3), 461-487.

Barsotti, F., Milhaud, X., & Salhi, Y. (2016). *Lapse risk in life insurance: Correlation and contagion effects among policyholders' behaviors.* Insurance: Mathematics and Economics, 71, 317-331.

Batty, M., Tripathi, A., Kroll, A., Wu, C., Moore, D., Stehno, C., Lau, L., Guszcza, J. & Katcher, M. (2010). *Predictive Modeling for Life Insurance. Ways Life Insurers Can Participate in the Business Analytics Revolution* (Deloitte Development LLP, New York).

Bauer, D., Russ, J., & Zhu, N. (2014). *Adverse Selection in Secondary Insurance Markets: Evidence from the Life Settlement Market.* In Proceedings of the NBER Insurance Workshop–2014.

Bauer, D., Gao, J., Moenig, T., Ulm, E. R., & Zhu, N. (2015). *Policyholder Exercise Behavior in Life Insurance: The State of Affairs.*

Bermúdez, L., Navarro, F. E., & Torres, J. L. P. (2004). *Introducció al món de les assegurances* (Vol. 76). Edicions Universitat Barcelona.

Bluhm, W. F. (1982). *Cumulative antiselection theory.* Transactions of the Society of Actuaries, 34, 215-231.

Bolancé, C. (2016). *Modelos de Tarificación a Priori*, lecture notes distributed in Applied Statistical Models, Universitat de Barcelona.

Boros, D. (2014). *On Lapse risk factors in Solvency II.* KTH Royal Institute of Technology

Campbell, J., Chan, M., Li, K., Lombardi, L., Lombardi, L., & Purushotham, M. (2014). *Modeling of policyholder behavior for life insurance and annuity products.* Society of Actuaries.

CEIOPS (2009). *Advice on Technical Provisions: Actuarial and statistical methodologies to calculate the best estimate.*

Chládková, D. (2014). *Pojistný matematik a Solventnost II.* Deloitte.

Christiansen, M. C., & Niemeyer, A. (2014). *Fundamental definition of the solvency capital requirement in solvency II.* ASTIN Bulletin: The Journal of the IAA, 44(3), 501-533.

Christelis, D., Jappelli, T., and Padula, M. (2010) *Cognitive abilities and portfolio choice*, European Economic Review 54, 18-38.

Clark, D., Kent, J., & Morgan, Ed. (2013). *Dynamic Policyholder Behaviour & Management Actions Survey Report.* Milliman

Cohen, R. A. (2006, March). *Introducing the GLMSELECT procedure for model selection*. In Proceedings of the Thirty-first Annual SAS Users Group International Conference. Cary, NC: SAS Institute Inc.

Corrigan, J., Decker, J., Hoshino, T., Delft, L. V., & Verheugen, H. (2009*). Aggregation of risks and Allocation of capital.* Milliman.

Conwill, S. (2016*). Life insurance risks: Observations on Solvency II and the modeling of capital needs.* Milliman.

Cox, S. H., & Lin, Y. (2006). *Annuity lapse rate modeling: Tobit or not tobit.* Society of Actuaries.

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

EIOPA (2014). *Guidelines on the valuation of technical provisions.*

EIOPA (2015,a). *Report on the fifth Quantitative Impact Study (QIS5) for Solvency II.*

EIOPA (2015,b). *Commision Delegated Regulation (EU) 2015/35*. Official Journal of the European Union.

Fang, H., and Kung, E. (2012) *Why Do Life Insurance Policyholders Lapse? : The Roles of Income, Health and Bequest Motive Shocks*, Princeton University.

Ferrari, S., & Cribari-Neto, F. (2004). *Beta regression for modelling rates and proportions.* Journal of Applied Statistics, 31(7), 799-815.

Finkelstein, A., McGarry, K., & Sufi, A. (2005). *Dynamic inefficiencies in insurance markets: evidence from long-term care insurance* (No. w11039). National Bureau of Economic Research.

Fortius d.o.o (n.d.). *The history of insurance.*

He, D. (2011). *Is there dynamic adverse selection in the life insurance market?.* Economics Letters, 112(1), 113-115.

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

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

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

Lombardi, L., Paich, M., & Rao, A. (2012). *Behavioral Simulations: Using agent-based modeling to understand policyholder behaviors.*

McCullagh, P., & Nelder, J. A. (1989). Generalized Linear Models, no. 37 in Monograph on Statistics and Applied Probability.

Michorius, C. Z. (2011). *Modeling lapse rates. Investigating the variables that drive lapse rates.* University of Twente.

Mulholland, B. S., & Finke, M. S. (2014). *Does Cognitive Ability Impact Life Insurance Policy Lapsation?.*

Munich Re(2000). *Unit-linked insurance. A general report.*

Ortiz, L. (2016). *Basic Concepts in Risk Management*, lecture notes distributed in Advanced Risk Quantification, Universitat de Barcelona.

Penn State (n.d., a). *Detecting Multicollinearity Using Variance Inflation Factors*, lecture notes distributed in STAT 501. Regression Methods. The Pennsylvania State University.

Penn State (n.d., b). *Introduction to Generalized Linear Models*, lecture notes distributed in STAT 504. Analysis of Discrete Data. The Pennsylvania State University.

Sirak, A. S. (2015). *Income and unemployment effects on life insurance lapse*. Working Paper.

Van der Veen, H., Trench, M., Cropsey, J (2010). *Definition of an Insurance Contract*.

Vandenabeele, T (2014). *Solvency II in a nutshell.* Milliman