Mémoire présenté le :

pour l'obtention du Diplôme Universitaire d'actuariat de l'ISFA
et l'admission à l'Institut des Actuaires

Par : Zacharie Guibert

Titre Facing the surrender risk in the life insurance business

Confidentialité : ☐ NON **x OUI** (Durée : ☐ 1 an **x 2 ans**)

Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

Membre présents du jury de l'Institut des Actuaires signature

Membres présents du jury de l'ISFA

Entreprise :

Nom : NPG Wealth Management

Signature :

Directeur de mémoire en entreprise :

Nom : Michael HODGES

Signature :

Invité :

Nom :

Signature :

Autorisation de publication et de mise en ligne sur un site de diffusion de documents actuariels *(après expiration de l'éventuel délai de confidentialité)*

Signature du responsable entreprise

Signature du candidat

Secrétariat

Bibliothèque :

Abstract (version Française)

Mots-Clés: Solvabilité II, Taux de rachat, Risque de Rachat, CAA (Commissariat aux Assurances), CSSF (Commission de Surveillance du Secteur Financier), PVFP (Projected Value of Future Profits), SCR (Solvency Capital Requirement), ALM (Asset & Liability Management), fonction de hasard, modèle de Vasicek, modèle semi – parametrique, modèle de Cox à hasards proportionels, Analyse de survie, modèle parametrique, estimateur de Kaplan-Meier, comportement de l'assuré, taux de marché, taux de rachat dynamique

Directement lié aux réserves, stratégie ALM et résultats finaux de l'assureur, le rachat est un des risques les plus élevés auquel l'assureur doit faire face. C'est pourquoi la modélisation de ce risque, de la manière la plus proche possible de la réalité, de façon à l'anticiper et le comprendre, est essentielle. En effet, un scenario haussier ou baissier de ce taux de rachat a des conséquences directes sur les flux futurs, actif et passif, de l'assureur.

Ce mémoire vise à comprendre et determiner les facteurs de rachat qui poussent les assurés détenteurs d'un contrat en unités de compte au sein d'une compagnie d'assurance vie luxembourgeoise à racheter, de manière totale ou partielle, leur police d'assurance – vie. Une analyse de survie sera realisé sur un portefeuille donné afin de déterminer l'impact de ces rachats sur la duration globale du portefeuille. Cela debouchera sur une étude approfondie de l'influence de chaque covariable sur le risque de rachat, ainsi que les forts débouchés aussi bien marketing que « risk – management » d'une telle analyse. Enfin, nous etudierons quels sont les facteurs de rachats en fonction des catégories socio – professionnelles des assurés, et les possibilités de modéliser ce phénomène via un taux de rachat dynamique.

Dans une première partie, nous traiterons de l'aspect réglementaire et prudentiel sur la place financière luxembourgeoise, tout en mentionnant ses spécificités. Nous présenterons également l'entreprise d'assurance – vie, NPG Wealth Management, au sein de laquelle ce mémoire a été réalisé.

Dans une seconde partie, nous adapterons un modèle de survie semi – paramétrique à nos données de rachat, et ainsi étudier les principaux pilotes de la décision de rachat de la part de l'assuré.

Enfin, nous nous intéresserons à l'impact des différents facteurs de rachats au sein des différentes catégories socio – profesionelles des assurés de l'échantillon, et à la construction d'un modèle de prédiciton du taux de rachat. Ce modèle sera basé sur un taux de rachat dynamique, qui aura pour but d'illustrer l'influence de l'évolution des taux de marché sur le taux de rachat observé au sein du portefeuille d'assurance – vie étudié.

Abstract (English version)

Key words : Solvency II, Lapse rate, Surrender risk, CAA (Commissariat aux Assurances), CSSF (Commission de Surveillance du Secteur Financier),PVFP (Projected Value of Future Profits), SCR (Solvency Capital Requirement), ALM (Asset & Liability Management), hazard function, Vasicek model, Semi – parametrical model, Cox PH model (Proportional Hazards), Survival, parametrical model, Kaplan-Meier estimate, policyholder behaviour, market rates, dynamic lapse rate, contract value

Directly connected to the insurer reserves, ALM strategy and performance final results, the surrender act is among the major risks a life insurer faces. In the actual prudential – risk approach regulatory framework, modeling this risk as proper as possible in order to anticipate them makes sense: a bullish or bearish interest rates scenario will have a direct consequence in terms of asset and liability management and stock of reserves. The better insurers would be able to model the surrender rates on their portfolios, the better they would be able to anticipate their own financial cashflows & liabilities.

This paper aims to determine which factors incite the policyholder to surrender or not, and in which proportions, within a life insurance portfolio. A survival analysis will be done in order to assess the impact of surrenders on the global portfolio duration. The sensitivity of each covariate on the surrender risk will be studied deeply afterwards. This will lead us at the end to draft a dynamic surrender model and describe a risk policyholder profile based on thepolicyholder job occupation.

In the first place, after a presentation of the regulatory framework in Luxembourg and NPG Wealth Management, we will get interested, through a survival analysis, to the variables which trigger the surrender decision. We will fit, to the portfolio data a semi – parametrical model – the Cox proportional hazards model, and determine the influent covariates. Hence, we will be able to determine how policyholders surrender their life insurance policy, and in which conditions.

In the second place, we will see that, in function of the job occupation, policyholders do not react equally in front of the decision to surrender. This study will indicate the type of profile which is the less susceptible to surrender, hence a decrease of the surrender risk for the life insurer

Finally, we will see the large impact of the evolution of financial market rates on the decision to surrender. In this sense, we will build a dynamic surrender model, in order to assess a qualitative prediction of what the surrenders would be, on a duration basis, over the next 100 months.

Thanks

Before starting festivities, I truly want to thank here all the persons who helped me to produce this paper.

First and foremost, I warmly thank Michael Hodges, my actuarial director, for all the actuarial and technical advices he gave me every day since I joined NPG Wealth Management.
Thanks to his help, I had the chance to experience a large overview of the life – insurance business from an actuarial and risk – management perspective.

I would also like to thank Beatrice Dessauger, Gabriel Robet and Gregory Campbell, for their advices and suggestions all over the redaction of this paper. I am also grateful to all NPGWM members I have been led to work with, for giving me a complete oversight of the company and its different departments.

In a second time, I adress my thanks to the ISFA faculty, and especially to Pierre Therond, my supervisor on this project, for all his advices, and Frederic Planchet, for the very same reasons.

I also sincerely and friendly thank Anne Eyraud – Loisel for her availability for any question I had and her advices all along my schooling.

Confidentiality

All the information contained within this paper is strictly confidential, and is meant to be broadcast for only two purposes:

- Internally, for the Actuarial, Risk – management, and the Sales & Marketing teams
- To be supported in front of the jury of the French Institute of Actuaries “Institut des Actuaires” and the “Institut de Science Financière et d’Assurance » (ISFA)

Any other use is strictly forbidden.

Table of Contents

Introduction

The life insurance industry in Luxembourg is profitable business, with more than €20 billion of reserves as of 2013. Thanks to unique investment vehicles in Europe, this business attracts life-insurance policyholders from all around the globe. The life insurance business benefits from an advantageous legal framework, which largely explains its popularity. Conversely, the insurer faces some risks all along the lifetime of a life –insurance policy. On one side, the market risks (bonds, equities, currency, real estate, counterparty ...); on the other side, human risks (mortality, surrender ...).
In the actual regulatory framework, insurers have to anticipate and know their risks, in order to be able to honor their commitments to the policyholders, remain solvent. Among these risks, the surrender risk is one of the biggest the life insurer has to face.

Several things might induce people to surrender their own life insurance portfolio. The first one, more financial, is depending on the gap between the benchmark market rate and the credited rate on the insurance product via a double S-curve. The second one is more “human”, depending on macro-economic variables. The “human” aspect of the structural rate (e.g. the investors' irrationality) makes this rate difficult to model, predict and assess.

Others surrendering factors are harder to assess, mainly because of a lack of data: For instance, since the 2008 sub primes crisis, more and more governments are hunting down tax evasion and tax heavens. The recent declarations (in March 2013) of Luxembourg to think about more transparency frightened some investors, who surrendered their portfolios in order to not be caught by their countries' authorities.

Modeling the conjectural surrender is not completely related to the world financial situation. In 2009-2010, life insurance investors did not surrender massively their portfolios, while bankruptcies and saving plans headlined. Conversely, it has been observed that investors become more attentive as soon as their portfolio's performances are compromised. Besides, hypothesis regarding surrender rates can have a huge impact on insurance company's results if they're not correct: Anti-selection, randomness, rate risk (when the insurer has to borrow money to reimburse the surrender value to the investor) … are among the surrender risks the insurer has to model and anticipate.

After the sub-primes crisis, European authorities set up a new regulatory framework, Solvency II. This new directive compelled life insurers to have a very clear idea of the risks they are facing and how sensitive they are towards them. The lapse risk is complex, by its dependence to both market risk (equity, bonds, counterparty ...) and human (death, age, job occupation ...), which makes its modelling essential for the life insurer. With this manner, by apprehending the policyholder behaviour, the life insurer will be not only able to model and predict the lapse rate he can expect. He will also be able to direct a marketing strategy to target some policyholders with the lowest surrender-risk profiles.

1. The Luxembourg Framework

1.1. The life insurance industry in Luxembourg

1.1.1. Overview

For several years now there has been a noticeable rise of interest in the life insurance products offered in Luxembourg, often considered as profitable investment vehicles. In an attractive framework, more and more people across Europe are choosing these products as a mean of placing their assets, structuring their wealth, and planning transmission of their wealth. This interest is backed up by the figures. The Luxembourg insurance supervisory authority, confirms it: After the 2011 drop for total premiums, the Luxembourg life insurance business succeeded to bounce, unlike others countries in the European Union in the same industry. As a consequence, the Luxembourg life insurance business represents 62% of the total insurance premiums in 2012 in the Grand-Duché.

The total amount of premiums for Luxembourg life insurers soared by 43.09%, which contrasts with the average around 0 predicted by the EIOPA for the Eurozone. With more than 20.5 million premiums in 2012 (22 million in 2012, and 14.5 million in 2011), the Luxembourg life insurance business in Luxembourg keeps being attractive aboard: the percentile of foreign investors grows up every year (Premiums coming from the French market grew up by 111.09% in 2012). The Luxembourg market does not escape to this rule, with an increase of premiums of 32.29%.[1]

Following the financial crisis, the wealthiest people have been looking for ways to protect their assets.

The interest in these products is first and foremost linked to the regulatory provisions put in place to regulate the business. At the heart of these provisions is the system known as the “security triangle” which guarantees an optimal security to the policyholder. The cornerstone of this system relies on the legal obligation that all the assets representing the client’s savings are deposited within a depositary bank approved by Luxembourg’s national control authority. The whole mechanism is regulated by an agreement between the insurance company, the depositary bank and the CAA. Indeed, the regulation of life insurance products in Luxembourg offers a framework that is both safe and flexible, a unique situation in Europe: the policyholder benefits from the “super-benefits” system: all the underlying assets the life insurance product will be, in the case of a very unlikely problem with the insurance company, swiftly returned to the policyholder.

This regulatory framework has been created in 1991, but it has taken on a new meaning since the financial crisis of 2008. Moreover, the life insurance contract proved its status of very popular vehicle these last few years, as being particularly suitable: by choosing these products, policyholders can optimize the structuring of their wealth and prepare their succession in an optimal tax framework and all this while keeping the same bank as well as the same asset manager – both CAA compliant. The asset manager will, for the largest accounts, be able to create a dedicated fund in which the registered assets will be managed.

Besides the guarantee of security, life insurance in Luxembourg offers other attractions. To begin with there is the flexibility offered by these policies with a view to making the most of the policyholder’s investments. Luxembourg offers great flexibility for life insurance-linked investment policies as well as a large number of eligible assets, regulated or otherwise (unlike France, for instance, where a life insurance contract one can only invest in conventional funds, as OPCVM). Based on the invested capital and the extent of the policyholder's wealth, investments can also be exploited through bonds, stocks, international funds or even structured products such as hedge funds or private equity funds.

On top of all that, Luxembourg is unanimously recognized for its financial expertise. Here, investors can find the right solutions for getting the most out of their investments in the way they want. Life insurance products allow them to benefit from specially adapted solutions.

In tax terms, life insurance is a neutral product with the advantage to avoid frictions between Luxembourg and the country of residence of the policyholder. It is not about tax avoidance here, no one in other European countries is pointing the finger at Luxembourg in this regard (until recently, with the tax evasion issue in France, subject then raised in the European parliament). Life insurance products are offered with the greatest respect for tax law, as established by the country of residence of the policyholder, and in many other countries, life insurance products are recognized and allow holders to benefit from tax exemptions. On top of this, the fact that the European savings directive, regulating tax on savings, has not yet been extended to life insurance products constitutes another attraction.

1.1.2. Luxembourg specificities

The "Commissariat aux Assurances" (CAA) is the legal and official entity to monitor the whole insurance industry in Luxembourg. The CAA has a range of missions, as, for instance, the ability to deliver the required accreditation to an insurance company (life, non-life, reinsurance, brokers …) or a prudential and frequent monitoring …The entity has to assist to international meetings in order to develop mutual standards inside the EU (as Solvency II), and submit deadlines to insurance companies to deliver prudential and EU frameworks reports and figures.

It is a public institution under the authority of the Minister of Treasury and Budget, endowed with legal personality separated the state and enjoying financial independence. However, even if the CAA has a legal entity separated from the state, the government still keeps some power inside the institution. Besides nominating and repealing its members, it gives its approvals regarding the accounts, the budgets and the accreditations. Whenever prudential reasons require collaborations, the CAA works with others supervisory authorities, national or foreign.

The CAA is composed of two distinct authorities:

- The Directive board is the highest executive authority of the CAA. It is composed of a chairman (In 2013, Mr. Victor ROD) and two other members, named by the government. The Management is responsible of all matters not specifically reserved –according to the law- to the Minister or the Council.

- The Council is a five government-members board, competent to fix the financial framework in which the Commissioner activity can be expanded. It is responsible to adopt the annual budget and accounts of the entity, and has also the right to give its opinion on the policy pursued by the management.

The CAA has the charge of all fees itneeds to operate. These fees are fully supported by taxes on companies and offices under the CAA supervision.

The monitoring is carried out in accordance with international and European standards.

- Before issuing an insurance license, the Commissioner instructed a particular case on the quality of the shareholders, professional integrity and morality of the officers, the future business plans sustainability and the invested capital adequacy.

- During the insurance activity, the Commission is regularly asking for and issuing on the activities of the companies, their assets, liabilities as well as their solvency margin. Most of the companies' statements must be certified by internal or external auditors, and/or by actuaries.

- The Office carries out regularly controls. It is entitled and allowed to be issued any useful document or information explaining the figures sent by the insurers. The Commissioner is also legally authorized to use a whole range of measures and/or sanctions to bring companies in failure or in difficulty to comply and reconcile with legal and regulatory statements. These measures may lead to a withdrawal of the accreditation in case of severe cases.

- After the cassation of the insurance operations (bankruptcy, failure, fraud …), the CAA remains able to supervise the conduct of the insurance business liquidation, in order to safeguard the interests of the policyholders.

1.1.3. Luxembourg insurance framework

Luxembourg established, in the 90's, a strict regulatory framework designed to provide an optimal protection for individual investors' savings and interests, using life insurance contracts as investments tools. The life insurance business is governed by prudential rules providing a safety reference model unique in Europe; thanks to a triple-protection system, detailed a few lines below. The insurance sector is monitored by the CAA.

Luxembourg-based life insurance policies are submitted to a triple protection thanks to the special and prudential Luxembourg regulatory framework:

- Quarterly CAA checks on balances between technical provisions and underlying (regulated) assets.

- Underlying (asset) securities are deposited within an approved bank in accordance with CAA's terms and conditions. The law stipulates that assets matching the insurer's liabilities must be deposited with a bank approved by the insurance industry regulatory authority, the Commissariat aux Assurances (CAA). Each life insurance company is required to sign a depositary agreement with a custodian bank and have this agreement approved by the

CAA. A difference is made between regulated assets and others assets belonging to the company (unregulated assets). If a life insurance company cannot face its engagements, the CAA can freeze its accounts to protect policyholders' savings.

In case of a company's default, Luxembourgish policyholders are first rank creditors for regulated assets: The law of 6th December 1991, as modified, grants subscribers to a Luxembourg life insurance contract the status of first ranking creditors on all assets in the technical reserves. This privilege, known as the "super privilege", takes precedence over all other creditors, whoever they are, granting contract holders priority in the recovery of credit related. This rule does not work in France, where the provided protection to a life insurance policyholder is capped to an amount of €70,000. Naturally, this high – level policyholder protection is one the advantages making the Luxembourg competitive for the life insurance business.

This mechanism, known as the "triangle of security", ensures that assets matching the insurer's liabilities are clearly separated from the company's other assets and lodged in a separate bank account. Client assets are thus legally separated from those of the insurance company's shareholders and creditors (meaning creditors are not allowed to seize a dime of the savings). Furthermore, the custodian bank itself is required to segregate assets and protect the interests of subscribers to a life assurance contract.

Other interests lure investors to bring their savings in the Grand Duché:

- Tax efficiency: all interests, dividends and capitals gains earned in a Luxembourgish life insurance contract are reinvested free of taxes (e.g. subject to the application of the international tax treaties).
- Competitive investment options: The CAA set up some investment restrictions to protect investors but also to provide them a wide range of investment solutions (linked to the amount invested and the investible wealth of each investor).
- Luxembourg products are various, with a high-technical level and a high-savings component.

Another part makes theLuxembourg financial centre attractive for the life insurance business : its particular asset management rules: A life insurance policyholder can invest, starting to a minimum of€ 250,000, in a wide range of assets which cannot be found somewhere else in Europe under the life insurance framework : securities, bonds, hedge funds, real estate, unlisted property … More than 90% of life insurance vehicles are unit – linked solutions within the Luxembourg life insurance market.
The 10% remaining are, for most of them, supporting the euro. This type of investment, eventually denominated in other currencies (USD, GBP, NOK, DKK ...) interest expatriate policyholders who do not wish to undergo exchange risk with regard to their country of residence mainly. Besides, these investments are less profitable because of the cost of reinsurance and the responsibility of the insurer towards the currency.

Capital management can be done with in-house mutual funds, owned by the insurer (generally a few funds offering hundreds of different fund managers) and / or within one or more funds managed by a private banker or asset manager CAA – compliant. The insured may split the management of its contract with several bankers or asset managers. The triangle does not guarantee safety to subscribers for recovering all their assets (excluding cash) in the presence of an insurer that does not market risk – e.g. offer only internal funds (dedicated or group).

In other cases (for example, in the presence of a Euro), there is pooling of assets and liabilities. On a starting point of€2.5 million, theLuxembourg life insurance framework – established in the Circular Letter (No. 08 / 1) of the CAA's January 2, 2008 – allow a high flexibility and investment, including without limitation issuer and / or asset class.

1.2. The Life insurance regulatory Framework: Solvency II

1.2.1. Solvency II

The solvency II directive is a new regulatory framework for the European insurance industry, that adopts a more risk-based approach (market consistent values), and implements a non-zero failure regime: The insurance company must be able to honor its own engagements with a probability of 99.5% (=0.5% probability of failure). [2]

1.2.2. Solvency II goals

Weaknesses, in the Solvency I directive were observed during the crisis. Even if AIG is not concerned by the European directive, its failure during the 2008 financial crisis made Europeans regulators realize that they needed a more advanced risk-thorough directive, in order to prevent European insurers, life and non-life, from bankrupting.

Luxembourg also experienced directly the impact of the 2008 financial crisis and the necessity to migrate their insurance company to the Solvency II framework as soon as possible: Excell Life International was a life insurance and Luxembourg-based company. In 2010 and 2011, the Luxembourg insurance regulator noticed irregularities in the accounts, a lot of non-respects to Luxembourg legal insurance framework, and the marketing of life-insurance policies fused to non-compliant investments, as for instance, in the Lehman Brothers fund "Orelius Golden Invest". The 17 February 2012, the insurance agreement was removed to this company, which was placed in liquidation. The "Commissariat aux Assurances" published a note on its website to summarize the weaknesses and irregularities listed [Annex 1].

While the former directive was aimed at revising and refreshing the solvency regime in use, the essence of the Directive is to require insurers to provide transparency over their risk and the levels of capital held to cover that risk. Insurers are required to demonstrate that they have fully defined, assessed, governed, quality tested and (where necessary) remediated the data that is material to Solvency II.

Solvency II has also created new requirements for the provision of asset data in the form of new data fields, new data coding conventions, greater granularity of data and increased frequency of reporting.

The Solvency Capital Requirement, developed in Solvency II, aims to:

- Reduce the risk than an insurer would be unable to meet claims
- Reduce losses suffered by policyholders in the event that a firm is unable to meet all claims fully
- Provide early warning to supervisors so that they can intervene promptly if capital falls below the required level
- Promote confidence in the financial stability of the insurance sector

One of the main goals of the directive is to contribute to the objectives of the European Union Financial Services Action Plan (EU FSAP), by encouraging the insurance sector to work and use with a single license/method throughout member countries. Indeed, the introduction of a unified legal framework for prudential regulation will help to maximize harmonization through the Eurozone, and be consistent with the principles used in banking supervision.

One of the main differences with the former directive, Solvency I, is the market consistent approach: this new approach is based on economic principles that measures assets and liabilities in order to align the insurers' risks with the capital they detain to safeguard policyholders' savings. Similar to the reasoning behind Basel II for the banking sector, and due to the weaknesses the 2008 financial crisis highlighted, the directive aims to modernize insurance standards and improve risk management techniques: establishment of a new set of capital requirements, valuation techniques, governance and reporting standards, and harmonization of the regulation all across the EU.

Finally, new capital requirements have been designed for providing a better reflexing about the insurer's individual risk, and giving, for small insurance companies, a formula to determinate their Solvency Capital Requirement. This is likely to lead to a supervisory need for companies to show greater competency in risk assessment, and an easier way to audit companies, thanks to a more unified approach for evaluating technical provisions.

To sum it all up, Solvency II intends to provide:

- An alignment of economic and regulatory capital
- Freedom for companies to choose their own risk profile and match it with the appropriate level of capital
- An active and market consistent capital management to have a risk prudential approach
- Encouraging improvements by identifying risks and their matching mitigation.
- Streamlining the way that insurance groups supervise and recognize their economic reality

1.2.3. Solvency II organization

As mentioned before, Solvency II is structurally speaking similar to Basel II regulation. Both are based on three pillars including quantitative and qualitative requirements, market discipline, and specific figures (capital, risk, supervision and disclosure).However, Basel II applies separate models for investment credit and operational risks, while Solvency II focuses on a risk-based portfolio analysis, considering dependencies between risk categories [3]. Besides, Basel II concentrates on the asset side while Solvency II assessment of capital adequacy applies economic principles on the total balance sheet (both assets and liabilities).

- **Pillar 1** deals with all the quantitative requirements. It ensures firms are capitalized enough, with an adequately risk-based capital. All valuations are done here with a prudential and market consistent methodology. Companies are free to use either the standard formula detailed in the directive (EIOPA), or an internal model approach. The use of internal models is subject to stringent standards and company needs.

- **Pillar 2** imposes high level standards regarding risk management and governance. It also gives to supervisors a greater power to challenge their firms on risk management issues (ORSA: Own Risk and Solvency Assessment). As a consequence, every insurance company has to undertake its own forward-looking assessment of its own risks, e.g. capital requirement plus adequacy of capital resources.

- **Pillar 3** insists on greater levels of transparency regarding both supervisors and the public. On a quarterly or annually basis, firms have to provide a report regarding both about their solvency and financial conditions. This ensures that a firm's overall financial position is better represented and included more up-to-date information.

1.2.4. Solvency II in Luxembourg

Originally, Solvency II wassupposed to be implemented in2012 but the complexity of the directive and the need to achieve Europe-wide consensus between regulators has seen the implementation date pushed back not once but twice, and is now expected to be January 2015.

In Luxembourg, insurers are more in advance on this subject than other countries, mainly for two reasons:

- The small size of the country: rules are easier and quicker to apply; The CAA and the government both work to implement EU directives promptly and adapt local legislation to support and develop cross-border life insurance business

- A hard-to-please regulator, the CAA, which kept another deadline imposed by the European Parliament, in 2012 as a limit for companies for the upgrade

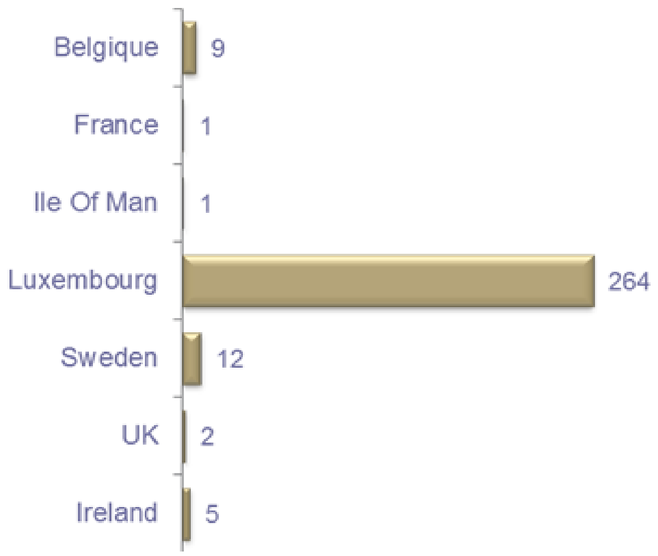
However, some parts still need to be thought by the insurers: for instance, providing data for each asset held on a security-by-security basis is very complicated to do, and, in July 2014, most of Luxembourgish insurers use the block of business basis, easier to set up.

Since year-end 2009, the CAA asks to every insurance company based in Luxembourg to produce and send to them a list of reporting files to ensure transparency and solvency in the Grand-Duché. They send quarterly excel spreadsheets specific to every insurance company they have to fill with their key figures: CAA Database, CAA Declaration 9, CAA stress-tests figures on guaranteed-rate portfolios, Annex D … [Annex 2].

1.3. NPGWM, New PEL Group Wealth Management

NPG Wealth Management, formerly PanEuro Life, is a life insurance holding,founded in 1991, and bought in 2006 by J.C. Flowers (found in 2001 by James Christopher Flowers, a former Goldman Sachs partner[4]), a leading private equity investment firm focused on investments in the financial services sector, which owns since 99.5% of the insurance holding.

All the insurance companies of the group (Private Estate Life, Altraplan Luxembourg, Altraplan Bermuda, Vestalife, Augura Life) are held by the Luxembourg-based holding, NPG Wealth Management. The firm offers wealth management solutions, private placement and unit-linked life insurance products. Operations are based in Luxembourg, Ireland, Gibraltar and Bermuda. At the 31 December 2013, NPG Wealth Management managed 6 billion assets under management. NPGWM workers, at the same date, are split as presented below:



NPGWM mostly targets wealthy clients all over Europe, reached through intermediaries as brokers, investment advisors or private banks. It offers a diversified panel of life insurance investment supports, separated in three main classes: internal funds, external funds and dedicated funds. Regarding internal funds, two categories of investment support exist: unit-linked investment or guaranteed-rate ones.

- Guaranteed rate solutions: Premiums are invested and managed by the company (the insurance company acts as a fund house); the funds are currency-labeled, and represent 10% of the total PEL contracts.
- Unit-linked investments (90% of PEL contracts are unit-linked): Usually composed of bonds, equities, and sometimes private estates values, unit-linked accounts follow the evolution of financial markets; it can provide higher gains, but also losses, according to the prevailing market conditions and the supported undertakings risks.

1.3.1. NPG Wealth Management funds

1.3.1.1. Internal funds

These funds are developed and managed, as guaranteed rate portfolios, by the insurance company herself. They can be compared, from a functional point of view, as undertakings for collective investment in transferable securities (UCITS). Collective internal funds are available to all the insurer customers.

NPGWM has 45 internal funds, for a global Assets Under Management" value of € 587 million, e.g. averagely speaking € 13 million per fund. More than 99% of internal funds belongs to PEL.

PEL has, in its internal funds both investment supports, on one hand, portfolios with a guaranteed rate, and on the other hand, unit-linked portfolios. On the € 586 million of internal funds, € 162 million belong to PEL guaranteed-rate funds (8 funds, in Euros except three in Dollars).

1.3.1.2. External funds

It concerns mutual funds and funds managed by fund houses. Luxembourg-based life insurance contracts particularly enjoy, unlike other countries in Europe, a wide range of funds. These fund houses are subject to an approval procedure and prudential monitoring from the Luxembourg supervisory authority, the CAA and the ASSF.

NPG WM has 382 external funds only composed of mutual funds, for instance JP Morgan Japan Equity.The global Assets Under Management value is equal to € 2,3 billion, e.g. averagely speaking € 6 million/fund. On PEL's side, there are 247 external funds, for a global AUM value equal to €2 billion.

52 fund houses are represented; the top three is composed of Carmignac Patrimoine, Dexia Money Market, and Carmignac Securities). It is usually very hard to negotiate with fund houses regarding their funds commissions. NPGWM chose to be, for internal fund an important client of the Carmignac fund house in order to be able to negotiate these: In total, 16 Carmignac funds are registered in PEL external funds, for a global assets under management value of € 725 million, e.g. 37%. For no dedicated funds, the CAA forbids investing more than 2.5% of the portfolio in private equity funds, hedge funds … For convenient reasons, NPGWM chose to invest only in mutual funds.

1.3.1.3. Structured products

In order to diversify its investment solutions, and to propose alternative products rather than bonds and equities, banks developed financial products, built with a bond and an option (call, put … on the relative index)

In NPGWM, structured products represent 975 funds, for a global AUM value of €618 million.In PEL, structured products represent 63 funds for a global amount of € 386 million.

The main structure designer is SociétéGénérale Investment Services, which designs more than 90% of broadcast structured products. However, Société Générale does not emit all of them, PEL counts 15 structured products broadcasters. The two broadcasters completing the podium are Barclays and Mediobanca.

1.3.1.4. Dedicated funds

With a premium of min €250,000, the investor can have access to a dedicated internal fund. The more money the investor brings, the more investment options he has. The policyholder has the liberty to choose the funds in its portfolio and fund houses.

The value of NPG Dedicated funds is € 1.3 billion, and around €400 million for PEL. The following table summarizes the financial supports the policyholder is able to invest in according to its savings.

Investment category	Investment	Personal wealth	Investment rules
A	➢ € 250,000	➢ € 250,000	Unlisted securities forbidden
B	➢ € 250,000	➢ € 500,000	All except assets with discretionary management
C	➢ € 250,000	➢ € 2,500,000	All assets and discretionary management
D	➢ € 2,500,000	➢ € 2,500,000	Every financial asset existing

1.3.2. NPG wealth Management Counterparties and Incomes

As detailed above, NPGWM life insurance products cover various asset classes. In counterparty, these products are structured as single premium policies or regular premiums ones. Policies can include death cover, a range of options (partial or full surrenders in case of specific events, top-ups, switches …) to adapt the contract to the client investment will, needs and/or strategy. As we saw, dedicated funds policyholders define themselves their own investment policy and assets classes.

Regarding the diversity of investment supports offered by the Luxembourg place, we easily understand that unit-linked contracts are more popular than guaranteed rate ones. In NPGWM, the proportion is 90%-10%. Indeed, investors are looking, besides safety, more freedom regarding their investments wills. However, they can find guaranteed-rate investments in their native country. Life-insurance contracts do not have a predefined maturity, the contracts ending with the policyholder's death. A surrender option exist for the majority of life-insurance policies: Clients can withdraw their assets in cash, fully or partially.

NPGWM income comes from various fees, from contract management to distribution, services and so on :

- Fees and charges are determined according to the policyholder fund value
- Fund house pay to the different companies rebates coming from assets and funds commissions
- Commissions are paid to brokers for their clients portfolio's management and for every new client they bring back
- Charges are received by the company in case of death, surrender, or any event defined in advance in the contract

1.3.3. The Camelea portfolio

The Camelea porfolio is one portfolio of Private Estate Life. It represents the PEL biggest portolio. At the end of December 2013, this portfolio is characterised by:

Portfolio Date	31 December 2013
Portfolio in EUR	1,260,434,666
Number of Clients	8,495
Number of contracts	9,735
Average client Age	56
% Belgium portfolio	99.60%

All the survival analysis and further studies presented below will be on this portolio.

1.4. NPGWM Solvency II outputs

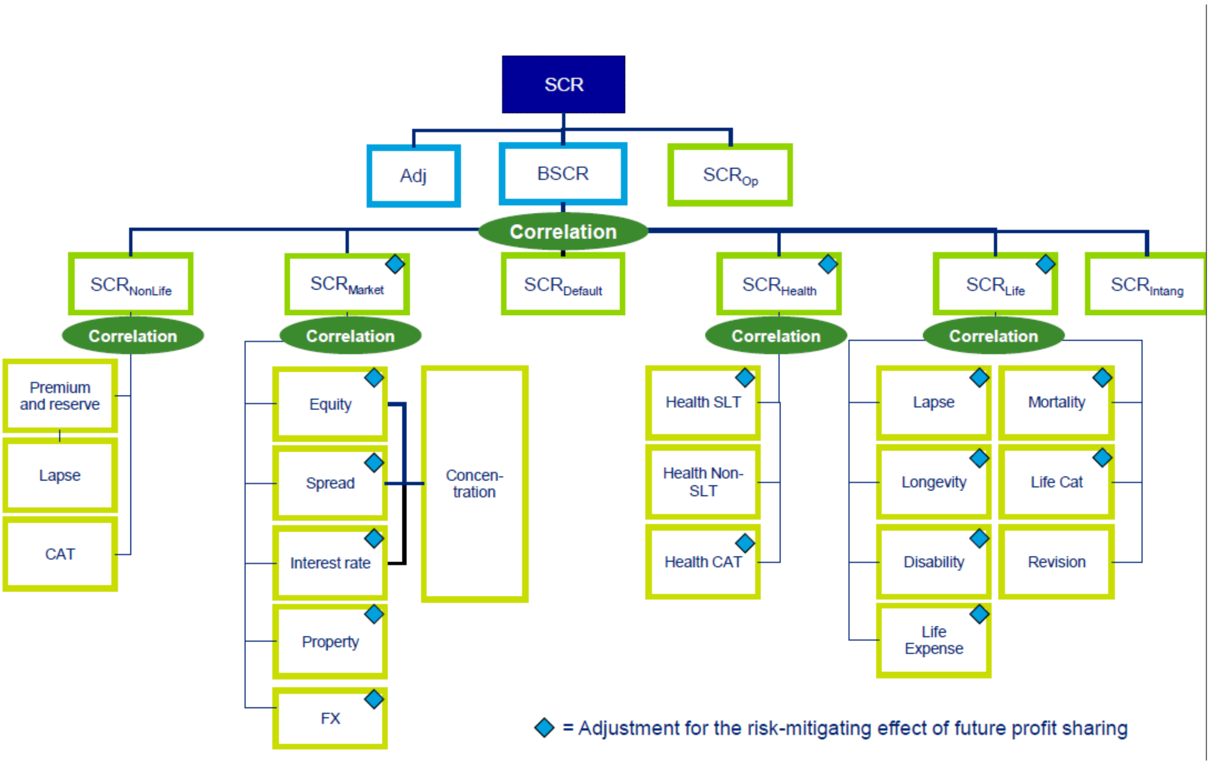
1.4.1. Solvency Capital Requirement

According to the article 101of the Solvency II directive, the SCR shall correspond to the value at Risk of the basic own funds of an insurance or reinsurance undertaking, subject to a minimum legal level of 99.5% over a one-year period: The citation 64 of the same directive goes deeper: The SCR should be determined as the economic capital to be held by insurance undertakings in order to ensure that those undertakings will still be in a position with a probability of at least 99.5%, to meet with their obligations to policy holders and beneficiaries over the following 12 months. This amount of capital's requirement is defined on Pillar I, and can be interpreted as the level of capital allowing insurers to absorb significant losses while giving guarantees to policyholders those payments will be honor as planned.

The SCR can be determined with two methods: equivalent or modular

- Equivalent approach: instantaneous shocks all made at the same time
- Modular: calculation per type of risks; for each type of risk, an SCR of each one is determined.

A portion of the risk is absorbed through both future discretionary bonuses and risk diversification effect. The aggregation of the individual SCR module is done by a correlation matrix provided by the EIOPA. [5]



Source: Deloitte

The final SCR as defined in the SII framework can be developed as follows:

$$SCR = BSCR + SCRoperational\ (-Adjustments)$$

- **BSCR**: Basic SCR; composed of 5 risks modules: non-life underwriting, life underwriting, health underwriting, market and counterparty risks.
- **SCRoperational**: the charge of capital for operational risk

BSCR Calculation:

$$BSCR = \sqrt{Corr(i,j) \times SCRi\ \times SCRj}$$

Where:

- $Corr(i,j)$denotes the entries of the correlation matrix (e.g. correlation parameters).
- $SCRi$ (Resp j) the SCR for the risk i (resp. j), with i and j run over all of the component risks.

SCR operational:

It represents the risk of a change in value caused by the fact that actual losses, incurred for inadequate or failed internal processes, people and systems, or from external events (including legal risk), differ from the expected losses.
It excluded however strategic decisions, reputation … risks.

1.4.2. Solvency Margin

The insurance business requires, each year, to determinate the risk margin, which corresponds to the cost of immobilization of an amount equal to the solvency capital requirement: The value of technical provisions shall correspond to the current amount insurance and reinsurance undertakings would have to pay if they had to transfer their insurance and reinsurance obligations immediately to another insurance or reinsurance undertaking. It effectively means that if an insurer was, as a result of a shock, to use up all its free surplus and capital, then it would still have sufficient assets to safely wind-up and transfer its obligations to a third party.

Twelve pages of the EIOPA report [5 bis]are describing the definition of this technical item and the general methodology for the risk margin calculation. It also gives the cost of capital rate to apply in the risk margin calculations, the level of granularity and the simplifications made and applied to the risk margin.

The risk margin should be calculated per line of business. A straight forward way to determine the margin per line of business is as follows. The calculation of the risk margin is based on a SCR projection scenario in the time.

$$RM = COC \times \sum_{t=1}^{n} SCR_t \times \frac{1}{(1+r_{t+1})^{t+1}}$$

Where:

- $SCR_t$is the SCR for year t and the corresponding undertaking
- r_{t+1} is the risk free rate for the maturity t+1 (no illiquidity premium included)
- $COC$is the cost of capital: Because the investors demand a certain return higher than the risk free rate on all capital, the company is making a cost by holding the extra amount of capital. Under the SII directive, this Cost of Capital is defined as equal to 6%.

However, the final risk margin must be free of market risk: indeed, according to QIS 5 requirements [6], the risk margin has been designed to guarantee that sufficient technical provisions are available even in case of a stressed scenario.

We determine the 2013 market risk tanks to the shock, considering

$$\boldsymbol{SCR\ market\ risk\ =\ Variation\ [\ PVFP(before\ and\ after\ equity\ shock)\ ]}$$

The risk margin has to be calculated with:

- amounts net of reinsurance
- a projection of obligations until extinction
- an appropriate allocation for each line of business
- an allowance for diversification between lines of business

1.4.3. CAA Excel spreadsheet

The insurance industry Luxembourg supervisor, the CAA imposed a migration to the Solvency II framework starting year-end 2011. Since then, annually, insurance companies have to report their Solvency II figures to CAA experts. To do so, the CAA sends to every insurance company, a secured reporting sheet (Annex 3), where insurers have to write down their shocks figures.

Nine tabs constitute the reporting sheet:

- **Validations**: Eventual messages if incoherencies noticed
- **Questionnaire**: The CAA asks methodology and assumptions questions that the company has to respond in a standard way (yes no, in which proportion). For instance, the CAA asks if simplifications have been made on a specific shock
- **Données**: Technical provisions analysis (only concerns guaranteed-rate portfolios for PEL)
- **BilanSolvabilité 2**: The Basis for the Solvency II balance sheet is the signed accounts. The CAA specifies the specific adjustments, which explanations can be found in Annex.
- **SCR**: it represents the solvency capital required to ensure that the insurance company will be able to meet its obligations over the next 12 months with a probability of at least 99.5%
  The CAA Excel spread sheets already include correlation matrixes between the different risks in order to model correctly the diversification effect and to consider the correlation between all the risks. The correlation matrixes in use are the ones defined in the EIOPA report. Insurers are free to change the correlation parameters if they have internal models or calculations to justify the changes. Here, we kept the original CAA correlation values.
- To calculate the SCR, we report the variation values coming from the shocks in the CAA spread sheet.
  The market risk ("Risque de marché"), is calculated as the sum of the interest rate risk, the equity risk, the real estate risk (we do not have real estate investments in NPGWM), the spread risk, the currency risk, and the concentration risk.
- **MCR**: The minimum capital requirement represents the threshold below which the supervisor, the CAA, would intervene. The MCR is intended to correspond to an 85% probability of adequacy over a one year period and is bounded between 25% and 45% of the SCR. As a consequence, the Best Estimate in use to calculate the MCR has to be free of insurance.Best Estimate values free of reinsurance are used to determine the MCR and the under risk capital and the importance/impact of the profit sharing.
- **Fondspropres**
- **Best Estimate**: Best Estimate value including reinsurance, Risk margin, reinsurance participation …

1.4.4. PEL Lapse risk in the SCR estimation

On the 2013 Solvency II reporting, the PEL Best Estimate for was equal to €3,039 million. The PEL SCR 2013 for the lapse risk is equal to €45 million, on a total SII 2013 SCR of €135 million. The lapse risk consequently represents more than 33% of the total SCR. This highlights the importance for the life insurer to model and estimate lapse rates. With the equity risk, this is the biggest risk the life insurer carries.

1.4.5. Duration

Once the solvency II calculation is done, we are able to propose a precise estimation of each portfolio/ company duration: The duration represents the change in the value of a fixed income security that will result from a 1% change in interest rates. Duration is stated in years. For example, a 5-year-durationmeans the bond will decrease in value by 5% if interest rates rise by 1% and increase in value by 5% if interest rates fall by 1%. Duration is a weighted measure of the length of time the bond will pay out. Unlike maturity, duration takes into account interest payments that occur throughout the course of holding the bond. Basically, duration is a weighted average of the maturity of all the income streams from a bond or portfolio of bonds.

As an example, for the Private Estate Life company, as of end of December 2012.

31/07/201313:33M:\SII\2012 Solvency II CAA deadline of the 12th July 2013____________\2012 Whole duration calculation for sheet G2\20130712_PEL and ATP LUX Duration annexe G2___________FINALSummary

PEL		
Source of information :		
Actuarial report version 2 for PEL	**Reserves**	Duration
Guaranteed portfolio	192 292 536	3,94
Unit linked portfolio	3 019 095 006	8,07
Total	3 211 387 542	7,82

However, this duration is highly affected by partial and full surrenders within the different NPGWM investment products. Indeed,a surrender affects the reserves, and consequently has a direct impact the duration.

1.5. The surrender risk

1.5.1. Overview of the surrender risk

From the reserves and the duration raises a new problem: The lapse modeling. Indeed, the lapse rate is directly connected to the insurer reserves and ALM strategy.

Indeed, a bullish or bearish interest rates scenario will have a direct consequence in terms of asset and liability management and in stock of reserves. These consequences may even become critical in case of large surrender (bullish scenario) or no surrender at all (bearish scenario), complying the insurer to pay a guaranteed rate as planned in its guarantees higher than its own assets' yield. Thus, The CAA request, regarding the transmission of the stress-tests results on guaranteed rate portfolios makes completely sense: the insurer and the regulator have to know the risk they can face and their ability to face it. And this goes through an anticipation of the lapse rate and an interest rate variation.

Lapses impact the duration of the portfolio in a significant way. Their surrender modeling is crucial and strategic for the insurer.

The surrender phenomenon is very important in the life insurance business: The better insurers would be able to model the surrender rates on their portfolios; the better they would be able to anticipate their own financial flows/ liabilities (costs – management underlies abetter asset-liability management) and satisfy their clients' requirements.

Several things might induce people to surrender their own life insurance portfolio. The first one, more financial, is depending on the gap between the benchmark market rate and the credited rate on the insurance product via a double S-curve:

- if the gap stays between two boundaries, deterministic surrenders are not modified
- If the gap is beyond these two limits, surrenders increase or decrease till a min or max

The second one is more "human", depending on macro-economic variables. Thanks to a lot of macro–economic data (stock exchange market rates, unemployment ...), these surrenders can be determined statistically. However, others risks, belonging to this category cannot be determined like this, for lack of data: For instance, since the 2008 sub primes crisis, more and more governments are hunting down tax evasion and tax heavens. The recent declarations (in March 2013) of Luxembourg to think about more transparency frightened some investors, who surrendered their portfolios in order to not be caught by their countries' authorities.

Neththeless, modeling the conjectural surrender is not completely related to the world financial situation. In 2009-2010, life insurance investors did not surrender massively their portfolios, while bankruptcies and saving plans headlined. Conversely,it has been observed that investors become more attentive as soon as their portfolio's performances are compromised (lower than another company for instance).

Surrenders focus the insurer's attention and researches since the 90's, for two reasons; a good understanding of investor's global behavior, and more generally, surrenders and explanatory factors allows to:

- Adapt new contracts' clauses and characteristics which purpose would be clients' retention (keep clients longer with stricter surrender conditions) for instance.
- Improve strategies regarding the Asset-Liability Management, EV calculations …

Hypothesis regarding surrender rates can have a huge impact on insurance company's results if they're not correct: Anti-selection, randomness, rate risk (when the insurer has to borrow money to reimburse the surrender value to the investor) … are among the surrender risks the insurer has to model and anticipate.

Three rules (IFRS, Solvency II, and the MCEV method (CFO Forum)) integrate the surrender risk in an international level, each one proposing methods to assess it:

- **IFRS**: (IFRS2 in particular), require to evaluate the insurance company's liabilities, and include the cost caused by lapses of options and guarantees.
- **Solvency II**: introduces a split by risk in the calculation of the solvency capital requirement, and a new design for the assessment of reserves. The surrender risk is the main center of this new European regulation [5]
- **MCEV**: the reinforcement of the reference benchmark measure for the valuation of an insurance company emphasizes the cost of options and guarantees, and consequently, surrenders [7]

The Luxembourgish position as a tax heaven (different legislation than the rest of Europe), and the shelter of various international funds makes the study of surrenders particularly interesting here. Indeed, surrenders in Luxembourg are not a reflection of the Luxembourgish surrenders, unlike, for instance, France, Belgium, Germany …, because, as mentioned before, 80% of the funds life insurance companies manage comes from abroad. Surrenders should consequently vary according the investors citizenship (increasing the last few months in Scandinavian and Belgium portfolios because of a stricter law regarding life – insurance savings abroad).

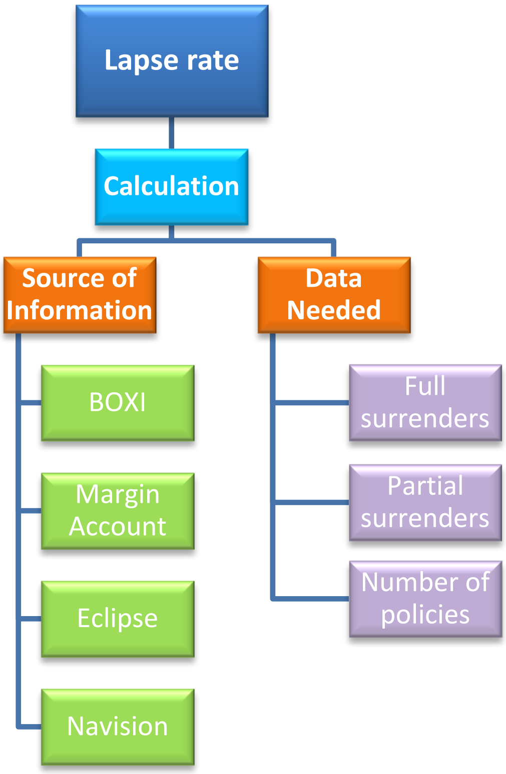
For all these reasons, it is essential for the life insurer to know what factors influence the policyholders' decision to surrender, in which proportions ... Beyond a ALM strategy, by apprehending the policyholder behaviour, the life insurer will be not only able to model and predict the lapse rate he can expect. He will also be able to direct a marketing strategy to target some policyholders with the lowest surrender-risk profiles.

1.5.2. NPGWM calculation lapses for Unit-Linked contracts

1.5.2.1. Required data

In NPGWM, lapses are calculated with a deterministic way. As for 2010, the gross amount of claims paid for PEL adds up to more than EUR 500 million in comparison with a total assets of more than EUR 3 300 Million, so that we need to consider lapses as an important item in the company. From that perspective we need to use a consistent method to calculate a lapse rate; moreover this lapse rate assumption will be used for different major matters such as the calculation of the Embedded value and of the Best Estimate though the projections of cash flows.

To proceed the lapses calculation, we need the following data:



The data comes from technical accounting or data extracts from BOXI, the data providing software in use within NPGWM. Eclipse and Navision corresponds to some companies not modeled into BOXI, and handled per technical accounting teams (for instance the 'dedicated funds' team).

1.5.2.2. NPGWM deterministic methodology

Considering the year N, we have to calculate an annual lapse rate for this year. We can consider that we have to calculate this rate for one contract, one product, one portfolio or one firm. In all cases, we can consider that the methodology remains the same.

We will calculate this lapse rate based on a three year history. We considered indeed that 3 years was the best choice. Less would give a greater influence on special events;more would not be consistent with the fact that we will use that rate for future years. Let's define a few notations:

- Reserve BOY_i is the reserve at the beginning of year i in €
- Reserve EOY_i is the reserve at the end of year i in €

For each year, we are performing an average of the reserve and then calculating a lapse rate. The final output is the lapse rate which is going to be used for the future years

We first have to calculate the average reserves of year i:

$$Average\ reserve\ i\ =\ Av\ i = \frac{Reserve\ BOY_i\ \ +\ Reserve\ EOY_i}{2}$$

$$Lapse\ for\ year\ i = \frac{Total\ amount\ of\ surrenders\ for\ year\ i\ in\ €}{Average\ reserve\ i}$$

- **For all companies the lapse rate** for the three year period is the following

$$Lapse\ rate = \frac{\left(\frac{1}{3}\right) * \sum_{i=1}^{3} Total\ surrender\ for\ year\ i}{\left(\frac{1}{4}\right) * \left[Reserve\ BOY_1 + \sum_{i=1}^{3} Reserve\ EOY_i\right]}$$

- **Which is the same than the following formula :**

$$Lapse\ rate = \frac{\sum_{i=1}^{3} Lapse\ for\ year\ i}{3}$$

We get, as a final extract:

Contract Sub Product Name (actual)	2010 Yearly Lapse rate	2011 Yearly Lapse rate	2012 Yearly Lapse rate	AVERAGE THREE YEAR LAPSE FOR MOSES	portfolio Ye 2009	Portfolio Ye 2010	Portfolio Ye 2011
53A RP	-6,87%	-2,47%	-7,74%	-5,69%	7.633.996	7.546.302	7.824.511
53A SP	-5,71%	-7,78%	-0,15%	-4,55%	1.118.771	1.099.771	1.064.686
ADIAMERIS		-1,25%	-2,62%	-1,93%		0	19.144.355
BBL RENTE	0,00%	-35,96%	-190,28%	-75,41%	30.589	31.504	22.608
BBL TOP LUX	-7,45%	-21,69%	-11,44%	-13,53%	7.730.493	7.552.039	3.736.548
BIL PRIVILEGE	-17,59%	-18,26%	-12,20%	-16,02%	35.682.201	31.147.076	25.188.037
Camelea	-7,51%	-5,68%	-10,66%	-7,95%	96.613.816	355.924.595	851.638.176
DUCAT	-7,75%	-12,01%	-16,10%	-11,95%	275.814.139	268.407.139	237.727.591
DUCAT+	-11,71%	-16,43%	-40,14%	-22,76%	4.892.252	5.103.947	4.143.793
DUCAT NEW GENERATION	-7,26%	-27,39%	-20,72%	-18,46%	743.135	715.431	453.408
E01 (ICA SPAIN - FRONT)	-25,00%	-13,47%	-8,77%	-15,75%	39.356.724	34.206.900	28.555.261
E01 (ICA SPAIN - SPREAD)	-14,06%	0,00%	0,00%	-4,69%	2.133.207	2.134.048	1.934.684
ELITE INVEST	-26,06%	-10,77%	-20,35%	-19,06%	5.324.960	3.069.908	2.541.819
ESTONIA	-23,15%	-27,87%	-19,97%	-23,66%	1.100.537	875.813	841.566
GIP-GIB1-GIB2 RP	-34,71%	-90,30%	-8,94%	-44,65%	195.580	159.947	61.673

This calculation is done of course for all products in every company; for consistency purposes, the same methodology is used for all different companies. This lapse rate calculation is assumed to be calculated with an expert judgment analyzing the impact of the lapse rate coming from the evolution of the number of policies and reserves.

1.5.2.3. NPGWM calculation for guaranteed-rate funds

What we call "Minimum guaranteed-rate funds" are funds providing to the policyholders an annual guaranteed-rate of return on an 8-years basis. "Resetting Guaranteed-Rate funds" are funds providing to the policyholders an annual rate of return on a yearly basis. The guaranteed rate is rest/updated the 31st January of each year. I won't detail here the setting of the guaranteed-rate for such funds.

The monitoring of the surrenders rates for the guaranteed rate funds (MGR/RGR) occurs on a quarterly basis. At the moment, there are 4 Minimum Guaranteed Rate Funds and 2 Reset Guaranteed Rate Funds.

67	MGR	EUR 2.75%
69	MGR	EUR 1.50%
77	MGR	USD 1.00%
78	MGR	EUR 1.00%
79	MGR	EUR 0.25%
73	RGR	USD
76	RGR	EUR

The result of this study will provide PEL with an experienced surrender rate that will be used in the stress tests calculation.

The calculation is also done on a 3 year-basis as presented before.

1.5.2.4. Opening

Having the surrender rate calculated on a yearly basis, based on the surrender results of the past three years, returns convenient results for the current year Solvency II modelling.

However, with this methodology, we are not able to make any precice prediction. Indeed, a Lagrange interpolation or fitting a trend on the available yearly surrender value would be not precise and relevant enough to make any sigificative predicton on a monthly basis. Besides, this type of calculation does not bring any information on the policyholder behaviour based on his characteristics and external events, such as the evolution of financial markets and unemployment.

The importance in the final 2013 SCR of the 2013 Lapse SCR, presented in section 1.4.4, shows us the importance to anticipate and understand this risk.

2. Survival analysis on the surrender rate

2.1. Introduction

The concept of the survival analysis is to study and model the failure time, which is, in our case, the surrender time.

2.1.1. Survival time

The survival time is defined as "a length of time that is measured from time origin to the time the event of interest occured"[8].

To determine survival time precisely, there are three requirements. A time origin must be unambigously defined, a scale for measuring the passage of time must be agreed uopon, and finally the definition of event must be entirely clear.
The difficulty of such a study dwells into the fact that some policyholders experienced the event while some did not at the end of the study, which makes their actual survival curves unknown. This is where intervenes the censoring effect

2.1.2. Censored Event

Censoring is defined as "the time when we have some information about individual survival time, but we do not know the survival time exactly". [9]

Three types of censoring exist [10]: right censoring, left censoring, and interval censoring.

Right censoring is said to occur if if the event occurs after the survival time.The censoring timeis the time beyond the studied subject cannot be observed. The observed survival time starts at time 0 and continues until the event X or a censoring time C, which ever comes first.

The observed data is resumed with (T, δ), where $T = \min(X, C)$ is the follow – up time, and $\delta = I_{X \le C}$ is an indicator for status at the end of follow – up time.

$$\delta = I_{X \le C} = \begin{matrix} 0 \; if \; X > C \; (observed \; censoring) \\ 1 \; if \; X \le C \; (observed \; failure) \end{matrix}$$

This case of censoring is particulary used when no events occured before the end of the study.

Censoring can also occur if we observe the presence of a condition, without especially knowing where the condition began – left censoring.

The interval censoring is the case of an individual known for having experienced an event within an interval of time, but without knowing though the acual survival time.

2.1.3. Survival time distribution

Let T represent survival time, and regard it as a random variable with a cumulative distribution function $P(t) = \Pr(T \leq t)$, and a probability density function $p(t) = \frac{dP(t)}{dt}$.The more optimistic survival function $S(t)$ is the complement of the distribution function:

$$S(t) = \Pr(T > t) = 1 - P(t)$$

The distribution of survival times can also be represented by the hazard function, which assesses the instantaneous risk of demise at any time t, conditional on survival to that time:

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

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

Models for survival data usually employ the hazard function or the log hazard. For example, assuming a constant hazard, $h(t) = \mu$, implies an exponential distribution of survival times , with the density function $p(t) = \mu e^{-\mu t}$. Other common hazard models include $h(t) = \exp(\mu + \rho t)$, leading to the Gompertz distribution of survival times.

2.2. Cox regression model

The not parametric method does not control covariates andrequires categorical predicators [11]. When there are several prognostic variables, multivariates approaches should be used. However, a multiple linear regression or a logisitic one cannot be used here, because they cannotdeal with censored observations. Another method is needed to model survival data with the presence of censoring. One very popular model in survival data is the Cox proportional hazards model, introduced by Cox in 1972[12].

Modeling the surrender risk with a semi-parametrical model: Survival analysis examines and models the time for events to occur. The prototypical such event is death, from which the name “survival analysis”.
This makes this analysis especially adapted to the surrender case. Instead of modeling the death (0 for survival, 1 for death), we model the surrender event (0 if the policyholder remains in the portfolio, 1 if he surrenders).

The survival analysis focuses on the distribution of survival times. Although there are well known methods for estimating unconditional survival distributions, most interesting survival modeling examines the relationship between survival and one or more predicators, usually termed covariates in the survival analysis literature.

Survival analysis typically examines the relationship of the survival distribution of covariates.

We are seeking, with the Cox model, to achieve to [13]

- Incorporate continuous covariates into our survival analysis
- Analyze the effect of covariates on survival (and not only the presence)

A natural first guess for a survival regression model would have been $h(t, x) = \beta_0 + \beta_1 x$

There is in this case no error term, as the randmoness is implicit to the survival process. Here, the notation in use is $h(t, x)$ the hazard function for an individual whose "independent" variable has the value x, while β_0 is abaseline hazard function (for the time being assumed constant in time t) for individuals with $x = 0$.

However, this is a bad model. The range of $\beta_0 + \beta_1 x$ may extend below zero for certain values of β_1 or x, but the range of $h(t, x)$ must be $[0, \infty[$.
By chance, a similar problem has arisen and been solved in generalized linear modeling. There, the predicators are incorporated into different distributions for the dependent variable. For a Poisson model, the mean must be positive, and the exponential function is used as the canonical link function between covariates and mean. Thus, we can suit by exponentiating the covariate terms :
$h(t, x) = \exp(\beta_0 + \beta_1 x) = h_0 \exp(\beta_1 x) > 0$

In case of more than one predictor: $h(t, x) = h_0 \exp(\beta^T x) > 0$

For a cohort with identical predictors x, the above form implies that lifetimes are exponential distributed, which we know to be unrealistic.

This examination entails the specification of a linear-like model for the log hazard. For example, a parametric model based on the exponential distribution may be written as a multiplicative model for the hazard $h_i(t) = \exp(\alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik})$

In this scenario, i is a suscript for observation, and the x are the covariates. The constant α in this model represents a kind of log baseline hazard – considering $h_i(t) = e^{\alpha}$ when all the x are equal to zero.

2.2.1. The Cox model

2.2.1.1. Overview

A Cox model is s statistical technique for exploring a relationship between the survival of a patient (in our case, the surrendering event) and several explanatory variables. [14]

As for the survival analysis, it aims to study the time between the entry to the study (subscription time of a Camelea life-insurance policy) and the subsequent event (the surrender).

The Cox model has the pros to provide an estimate of the surrender effect on survival after adjustment on the other explanatory variables. On top on that, it offers the possibility to estimate the hazard (or risk) of surrender for a policyholder, considering its prognostic variables.

The Cox model is based on a modeling approach to the analysis of the survival data. Its purpose is to simultaneously explore the effects of several variables and survival. The model allows isolating one effect from the rest of the covariates. The model can also be used to determine the covariates which influence the policyholder's survival within the Camelea portfolio.

From a set of observed survival times in a sample of policyholders, we can estimate the proportion of the population who would remain in the portfolio a given length of time under the same circumstances (fixed covariates). This is the Kaplan-Meier method, which is used for producing the survival function.

The regression method introduced by Cox is used to investigate several variables at any time t – and is also known as the proportional hazard regression analysis.

The procedure models and/ or regresses the survival times – the hazard function, on the explanatory variables.

In order to be valid, the Cox model must be fitted before coming to a conclusion. The final model, coming from a Cox regression analysis, will yield an equation for the surrender risk as a function of several explanatory variables.

Interpreting the Cox model involves examining the coefficients for each explanatory variable [15]:

- A positive regression coefficient for an explanatory covariate means that the higher the risk is, the worst the prognostic will be – e.g. a higher surrender rate
- A negative regression coefficient implies a lower surrender rate for policyholders with higher values of that variable

2.2.1.2. The hazard function

The Cox model allows defining a hazard function based of several variables. A hazard function is defined as the probability that an individual will experience an event (in our case, a policyholder facing the surrender of his policy) within a small time interval, given that the individual was alive at the beginning of the time interval. It can therefore be interpreted as the surrender risk at time t.

The hazard function, generally noted $h(t)$ can be estimated as such:

$$h(t) = \frac{\begin{array}{c} number\ of\ policyholders\ experiencing\ a\ surrender \\ in\ the\ interval\ beginning\ at\ t \end{array}}{\begin{array}{c} (number\ of\ policyholders\ remaining\ in\ the\ portfolio \\ without\ surrendering\ it)\ \times (interval\ width) \end{array}}$$

2.2.1.3. The regression function

The regression is a way to describe the relationship between the different variables. Let's illustrate this with an example:

We have two variables, X and Y:

- X, the age of the policyholders
- Y, the respective amount of savings

Performing a regression of Y on X comes to investigate on the relationship between the dependent variable Y, based on the explanatory variable X.

When more than one explanatory variable need to be included in the regression model, the method is known as multiple regression (for instance, including the variable W as the sex of each policyholder).

The Cox method is based on a multiple regression, except that the dependent variable Y is the hazard function at a given time t. If we have several explanatory variables of interest X (for example, for our problem, the sex, the age, the level of savings), then we can express the hazard or surrender risk at time t as

$$h(t) = h_0(t) \times \exp(\beta_{age} age + \beta_{sex} sex + \beta_{savings} savings)$$

The quantity $h_0(t)$ is the baseline – underlying hazard – function and corresponds to the probability of surrendering when all the explanatory variables are set to zero. The baseline hazard function is analogous to the intercept in ordinary regression (due to $e^0 = 1$).

The regression coefficients $\beta_{age}, \beta_{sex}, \beta_{savings}$ give the proportional changes that can be expected in the hazard, related to changes in the explanatory covariates. These coefficients are estimated by the likelihood statistical methodology (see 2.2.2.4.).

The assumption of a constant relationship between the dependent variable and the explanatory ones is called proportional hazards. It means that hazard functions for any two random policyholders at any point in time are proportional. In other words, if a policyholder has one risk to surrender at some initial point that is twice as high as that another policyholder, then at all later times, the surrender risk remains twice as high. This main model assumption of proportional hazards must be tested to validate the model.

2.2.2. The Cox proportional hazards model

2.2.2.1. Overview

Let T be a nonnegative variable representing the failure time of an individual in the populatiopn. The distribution of failure time, T, can be represented in the usual manner in terms of density or distribution functions as well as in more specialized ways such as the hazard function. Specifically, the hazard function at time t among individuals with a covariate z is defined as:

$$h(t|z) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

which represents the risk of failure at any time t, given that individual has not failed prior to t. Indeed, $h(t|z)$ provides a convenient starting point for modeling the relationship of hazard functions among different covariates z.

One such modelassumes that covariates affect the hazard functions in a multiplicative manner based on

$$h(t|z) = h_0(t)e^{\beta z}$$

where β is a row vector of p unknown parameters and $h_0(t)$ is an arbitrary baseline hazard function. The factor $e^{\beta z}$ describes the risk of failure for an individual with regression variable z related to the factor $e^{\beta z}$ at a standard value $z = 0$ [16]. From a ratio of hazard functions corresponding to any two z-values not dependent on t is coming the name of “proportional hazards”.

Let’s consider the following generalization:

$$h(t,x) = h_0(t,\alpha)\exp(\beta^T x)$$

where α are some parameters influencing the baseline hazard function. The hazard ratio is decomposed into a product of two items

- $h_0(t,\alpha)$, a term that depends on time but not the covariates
- $\exp(\beta^T x)$, a term that depends on the covariates but not time

This is the COX PH model, for Cox Proportional Hazards. The specificity and the beauty of this model, as observed by Cox, is that if you use a model of this form, and you are interested in the effects of the covariates on survival, then you do not need to specify the form of $h_0(t,\alpha)$. Even though, we can still estimate β. The COX PHM is thus semi-parametrical, as some assumptions are made on $\exp(\beta^T x)$, but no form is pre-specified for the baseline hazard $h_0(t,\alpha)$.

We are talking in this model about proportional hazards for the following reason: Consider two individuals with covariates x_1 and x_2. The ratio of their hazards at time t is

$$\frac{h(t,x_1)}{h(t,x_2)} = \frac{h_0(t,\alpha)\exp(\beta_1 x)}{h_0(t,\alpha)\exp(\beta_2 x)}$$

$$\frac{h(t,x_1)}{h(t,x_2)} = \exp\{(\beta_1 - \beta_2)\}$$

This concludes that $h(t,x_1) \propto h(t,x_2)$, e.g. hazards are proportional each other and do not depend on time. In particular, the hazard for the individual with the covariate x_1 is $\exp\{(\beta_1 - \beta_2)\}$ times that of the individual with the covariate x_2. The term $\exp\{(\beta_1 - \beta_2)x\}$ is called hazard ratio comparing x_1 to x_2.

If $\beta = 0$, then the hazard ratio for the covariate is equal to $e^0 = 1$.This means that the very covariate has no impact on the survival. Thus, we can use the notion of hazard ratios to test ifcovariates influence survival. The hazard ratio also indicates how much more likely one individual is to surrender at any particular point of time. If the hazard ratio comparing men to women was 2, this would mean that, at any time, men are twice as likely to surrender than women.

However, there may be some interactions between covariates and time, in which case hazards are no longer proportional anymore. Similarly, there is no reason to expect the log of the hazard function to be linear with the covariates. At the beginning, the assumption of proportional hazards will be assumed and appropriate, and then verified.

2.2.2.2. Survival function

The survival fonction is expressed as such: $S(t) = \exp(-\int_0^t h(\tau)d\tau)$.
In our case here, our estimate of the hazard function is a discrete approximation to a continuous function. As the estimation of the baseline function that we will see just after, we will use the estimate $\hat{h}_0(t_i)$ the estimate of the baseline function to express the estimate of the survival one.

$$\hat{h}_0(t_i) = \frac{d_i}{\sum_{j\epsilon R(t_i)} \exp(\hat{\beta}^T x_j)}$$

With :

- d_i the number of surrenders at time t
- $R(t_i)$ the set of individuals that could surrender at time t

With $\hat{h}_0(t_i)$, we can estimate $\int_{t_{i-1}}^{t_i} h_0(\tau)d\tau$, and follows our estimate of the baseline survival function $\hat{S}_0(t_i) = exp\left[-\sum_{j\leq i} \hat{h}_0(t_j)\right]$.

Breslow (1972) provided an estimate for $\hat{h}_0(t)$, which is obtained by maximising $h_0(t)$ in which the parameters β are substituted by the maximum partial likelihood estimators $\hat{\beta}$. The estimator of the baseline survival function $S_0(t)$ is given by

$$\hat{S}_{O,B}(t) = \prod_{i!t_i<t} \left(1 - \frac{d_i}{\sum_{j\epsilon R(t_i)} exp\{\beta z_j\}}\right)$$

the estimated survival function,$\hat{S}(t)$, as an illustration of the time until the first surrender event. The dashed lines are showing a point-wise 95% confidence envelope around the survival function

2.2.2.3. Cumulative hazard

The Cox PH model is a semi-parametric method of estimation. We do specify a model for the effect of the covariates, but anything specifically modelled on the baseline hazard function side. The Kaplan Meir estimator used through the R package to estimate the survival function does not required either to specify a model for the survival function [17] Thus, considering both hazard and survival functions are untimely linked, we can adapt the Kaplan-Meier method to estimate the baseline hazard function.

The estimate for the baseline hazard function at the time $t_i$of the ith event is:

$$\hat{h}_0(t_i) = \frac{d_i}{\sum_{j\epsilon R(t_i)} \exp(\hat{\beta}^T x_j)}$$

With :

- d_i the number of surrenders at time t
- $R(t_i)$ the set of individuals that could surrender at time t

2.2.2.4. *Partial likelihood estimate*

By fitting the Cox proportional hazards, we wish to evaluate $h_0(t)$ and β. One approach is to attempt to maximise the likelihood function for the observed data simultaneously with respect to $h_0(t)$ and β. Cox proposed an approach in which the partial likelihood function, not depending on $h_0(t)$ is obtained for β. The partial likelihood is a technique developed to make inferences on regresssion parameters, within the presence of regressison parameters ($h_0(t)$ in the Cox PH model). Based on the Cox proportional hazards model, the partial likelihood function is expressed as follows[18]

Let $t_1, t_2, \dots, t_n$ be the observed survival time for n individuals.Let the ordered surrender time of p individuals be $t_{(1)} < t_{(2)} < \dots < t_{(p)}$, and $R(t_{(j)})$ the risk set of individuals who are investing in the Camelea portfolio and uncensored at the time just prior to $t_{(j)}$. The conditional probability that the ith individual surrenders at $t_{(j)}$ given that one individual from the risk set on $R(t_{(j)})$ surrenders at $t_{(j)}$ is

$$\nabla = P\left(policyholder\ i\ surrenders\ at\ t_{(j)} \middle| \ one\ surrender\ from\ the\ risk\ set\ R\left(t_{(j)}\right) at\ t_{(j)}\right)$$

$$\nabla = \frac{P(policyholder\ i\ surrenders\ at t_{(j)})}{\sum_{k \in R(t_{(j)})} P(policyholder\ k\ surrenders\ at\ t_{(k)})}$$

Taking the expression to the limit, with $\Delta t \to 0$, we get

$$\nabla = \frac{h_i(t_{(j)})}{\sum_{k \in R(t_{(j)})} h_k(t_{(j)})} = \frac{\exp(\beta' x_i\ (t_{(j)}))}{\sum_{k \in R(t_{(j)})} \exp(\beta' x_k\ (t_{(j)}))}$$

Hence a partial likelihood function for the Cox PH model given by

$$\mathrm{L}(\beta) = \prod_{j=1}^{p} \frac{\exp(\beta' x_i\ (t_{(j)}))}{\sum_{k \in R(t_{(j)})} \exp(\beta' x_k\ (t_{(j)}))}$$

In which $x_i(t_{(j)})$is the vector of covariate values for the policyholder i surrendering at time $t_{(j)}$. The general method of partial likelihood was discussed by Cox [link]

However, this likelihood function is only for uncensored policyholders. Let $t_1, t_2, \dots, t_n$ be the observed survival time for n individuals, and δ_i be the event indicator, which is 0 if the ith survival time is censored (e.g. no surrender), 1 otherwise.

This way, the log likelihood function presented above can be expressed as follows:

$$\mathrm{L}(\beta) = \prod_{i=1}^{n} \left[\frac{\exp(\beta' x_i\,(t_i))}{\sum_{k \in R(t_i)} \exp(\beta' x_k\,(t_i))} \right]^{\delta_i}$$

Where $R(t_i)$ is the risk set at t_i.

2.2.3. Model validation

2.2.3.1. Proportional hazards assumption validation

The main assumption of the Cox proportional hazards model is precisely proportional hazards. Proportional hazards means that the hazard function of one individual is proportional to the hazard function of a second individual, e.g. a hazard ratio constant over time. There are several methods for verifying that a model satisfies the assumption of proportionality [13] [19].

2.2.3.2. Graphical method

The Cox PH model survival function is obtained by the relationship between hazard function and survival function.

$$S(t,x) = S_0(t)^{\exp(\sum_{i=1}^{p} \beta_i x_i)}$$

Where $x = (x_1, x_2, \dots, x_p)'$ is the values of the vector of explanatory variables for a particular individual. Taking the logarith twice of this expression leads to

$$\ln(-\ln S(t,x)) = \sum_{i=1}^{p} \beta_i x_i + \ln(-\ln(S_0(t)))$$

Then the difference in log-log curves corresponding to two different individuals with variables $x_1 = (x_{11}, x_{12}, \dots, x_{1p})$ and $x_2 = (x_{21}, x_{22}, \dots, x_{2p})$ is given by

$$\ln(-\ln S(t,x_1)) - \ln(-\ln S(t,x_2)) = \sum_{i=1}^{p} \beta_i (x_{1i} - x_{2i})$$

expression which does not depend on the time t. This relationship is very helpful inasmuch as it helps identifying situations where it may have proportional hazards. By plotting estimated $\ln(-\ln(survival))$ versus the survival time, parrallel curves should be observed in case of proportional hazards.

However, this method doesnot work well for continuous or categorical predictors having many levels,the graph becoming in this very case, cluttered. In addition, the curves are sparse when there are a few time points and it may be difficult to tell how close to paralle is close enough.

Furthermore, looking at the Kaplan-Meier curves and $\ln(-\ln(survival))$ is not enough to be certain of proportionality since they are univariate analysis and do not show whether hazards will be still proportional when a model includes many other predictors. But they support the argument. Some other statistical methods should be used for checking more precisely the proportionality.

2.2.3.3. Adding time –dependent covariates in the Cox model

The aim is to create time-dependent variables by creating interactions between predictors and a function of survival time, in order to inlude them in the model at the end [20].

Let's call a predictor of interest X_j. We can create a time dependent covariate $X_j(t) = X_j \times f(t)$, with $f(t)$ a function of time, for instance t or $\ln(t)$. The model assessing the proportional hazards assumption for X_j adjusted with the other covariates is

$$h\big(t, x(t)\big) = h_0(t) \exp\big(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_j x_j + \cdots + \beta_p x_p + \delta x_j \times f(t)\big)$$

where $x(t) = \big(x_1, x_2, \ldots x_p, x_j(t)\big)'$ is the values of the vector of explanatory variables for a particular individual. The null hypothesis to check proportionality is the condition $\delta = 0$. The statistic test can be carried out using either Wald or likelihood tests.

In the Wald test, the test is $W = \left(\frac{\hat{\delta}}{se(\hat{\delta})}\right)^2$

The likelihood ratio test calculates the likelihood under the null hypothesis, L_0, and the likelihood under an alternative hypothesis, L_a. The likelihood ratio statistic LR is then

$$LR = -2\ln\left(\frac{L_0}{L_a}\right) = -2(l_0 - l_a)$$

where l_0 and l_a are log-likelihood under two hypothesis respectively.

Both statistics have a χ^2distribution with one degree of freedom under the null hypothesis. If thetime – dependent covariate is significant (e.g. null hypothesis rejected), then the predictor is not proportional.

Note: Similarly, the PH assumption for several predictors can be assessed simultaneously

2.2.3.4. Tests based on the Schoenfeld residuals

Another statistical test for checking the proportional hazard assumption is based on the Schoenfeld residuals. If the PH assumption holds for a due covariate, then the Schoenfeld residuals for that covariate will not be related to survival time. So this test is successful by finding the correlation between the Schoenfeld residuals for a particular covariate and the ranking of survival times. The null hypothesis is that correlation between the Schoenfeld residuals and the ranked survival times is zero. The rejection fo the null hypothesis can be summarized by the violation of the PH assumption.

2.2.4. Cox proportional hazards diagnostics

2.2.4.1. Schoenfeld residuals

Once the model has been fitted, the adequacy of the fitted model needs to be assessed. The model checking procedures below are based on residuals.

Unlike Cox Snell residuals and deviance results, the Schoenfels residuals are covariate-wise residuals. They were at the origin called partial residuals because the Schoenfeld residuals for the ith individual on the jth explanatory variable X_j is an estimate of the ith component of the first derivative of the logarith and the partial likelihood function with respect to β_j. Taking the logarith of the partial likelihood drives to

$$\frac{\partial \ln L(\beta)}{\partial \beta_j} = \sum_{i=1}^{p} \delta_i (x_{ij} - a_{ij})$$

Where x_{ij} is the value of the jth explanatory variable $j = 1,2,\dots,p$ for the ith individual, and

$$a_{ji} = \frac{\sum_{l\epsilon R(t_i)} x_{jl} \exp(\beta' x_l)}{\sum_{l\epsilon R(t_i)} \exp(\beta' x_l)}$$

The Schoenfled residual for ith individual on X_j is given by $r_{p_{ji}} = \delta_i \{x_{ji} - a_{ji}\}$. The Schoenfeld residuals sum to zero.

2.2.4.2. *Model Schoenfeld residuals*

In order to validate the model and its assumptions, we need to plot the Schoenfeld residuals for each covariate in order to check the proportional hazards assumption over time. Indeed, this assumption is the main one for keeping the Cox PH model valid and draw conclusions; The Cox PH model is valid only if the surrrender effect is not a function of time.

The Schoenfeld test allows to validate the hypothesis. The mathematical formulation of the surrender risk of a policyholder i in the Cox model can be expressed as

$$h_i(t) = h_0(t).\, e^{(X_i(t), \beta(t))}$$

- If$\beta(t) = \beta$, the risks are proportional, and the surrender risk can be expressed as detailed in the Cox model.
- If β is not constant, the impact of one or several covariates changes over time

For the covariate j, plotting the function $\beta_j(t)$ is a way to analyze the risk variations over time. If the proportional hazards assumption is validated, residuals have in theory an aspect completely randomize, and the avergae evolution of the covariate over time is a horizontal line.

2.2.5. Solving the Proportional hazards assumption

Let's suppose now that statistic tests or other diagnostic techniques gave strong evidence of non proportionality for one or more covariates. In order to face and solve this problem, two solutions can be proposed [20]:

- Stratified Cox model
- Cox regression model with time-dependent variables

2.2.5.1. Stratified Cox model

2.2.5.1.1. Principle

An alternative for dealing with non proportional hazards is to stratify over the covariates not satisfying the proportional hazards assumption. In essence, stratification involves fitting a model that has a different baseline hazard in each stratum.
The advantage of this model is that it doesn't involve worrying about a (subjective) functional form assumed for the time interaction.However, the disadvantages are:

- The baseline hazards are estimated with in strata only, meaning that there is more uncertainty in their estimates as information is not pooled over strata as in the Cox extended model
- By stratifying over the covariate x, we lose ability to quantify its effect
- Continuous covariates have to be arbitrarily categories

Nonetheless, stratification is a solution for solving the issue of PHA violation.

2.2.5.1.2. Formulation

The stratifed Cox model stratifies the predictors not satisfying the proportional hazards assumption. Data is stratified into subgroups and the model is applied for each stratum. The model is given by $h_{ig}(t) = h_{0g}(t)\exp(\beta' x_{ig})$, where g represents the stratum.

It is remarquable in this situation that the hazards are not proportional because the baseline hazard can be different between strata. The coeffcients β are assumed to be the same for each stratum g. The partial likelihood function is here the product of the partial likelihood in each stratum. A drawback of this approach is that we cannot identify the effect of this stratified predictor. This technique is most useful when the covariate with non-proportionality is categorical and not of direct interest.

2.2.5.2. Cox regression model with time-dependent covariates

2.2.5.2.1. Principle

Until now, we have assumed that the values of all the covariates did not change over the period of observations. However, the values of covariates can change over time t. Such a covariate is called a time-dependent covariate. The second method to consider is to model non-proportionality by time-dependent covariates [21]. The violation of the proportional hazards assumption is equivalent to interactions between covariates and time. That is, the PH model assumes that the effect of each covariate is the same in all points in time. If the effect of a covariate varies over time, the proportional hazards assumption is violated for this very covariate.

2.2.5.2.2. Formulation

In order to model a time-dependent effect, an option is creating a time – dependent covariate $X(t)$, where $\beta X(t) = \beta X \times f(t)$. $f(t)$ is a function of time, such as $t, \ln t$, …The choice of time-dependent covariates may be based on theoretical considerations and strong clinical evidences.

The Cox regression model, with both time dependent covariates $X_j(t)$ and time – independent predictors X_i can be rewritten as

$$h(t|x(t)) = h_0(t)\exp\left[\sum_{i=1}^{p_1}\beta_i x_i + \sum_{j=1}^{p_2}\alpha_j x_j(t)\right]$$

The hazard ratio at time t for the two individuals with two different covariates x and x^* is given by

$$\widehat{HR}(t) = \exp\left[\sum_{i=1}^{p_1}\hat{\beta}_i(x_i^* - x_i) + \sum_{j=1}^{p_2}\hat{\alpha}_j(x_j^*(t) - x_j(t))\right]$$

In the hazard ratio formula, the coefficient $\hat{\alpha}_j$ is not time – dependent. $\hat{\alpha}_j$ represents the over all effect of $X_j(t)$. Conversely, the hazard ratio is a function of time t. This means that the hazards of an event occuring at time t is no longer proportional, and the model is no longer a proportional hazard model.

In addition to considering time-dependent variables for analyzing a time – dependent variable not satisfying the PH assumption, there are variables that are inherently defined as time – dependent variables.
Indeed, time – dependent variables have been lately classified as internal and external.

- An internal time – dependent variable is defined as a covariate which changes over time, based on the characteristics or behavior of the policyholder (age, job occupation ...)

- An external time-dependent variable is defined as a covariate whose value changes because of characteristics external to the policyholders (stock exchange market rates, unemployment ...)

2.2.5.3. *Fitting a Cox regression model with time – dependent covariates*

The coxph function handles time – dependent covariates by requiring that each time period for an individual appear as a separate observation – that is, as a separate row or record, in the data set. The extended Cox model reflects an interaction between the covariates and time, e.g. a change in the effect of covariates. The model can also be extended to reflect dynamic changes in the covariates [20].

By dynamically changing covariates, we mean here covariates that actually change with time, rather than just their effect changing with time. To incorporate such time – varying covariates, we need to use the extended Cox model. For instance, the hazard function for a model with one constant covariate x_1 and one time – varying covariate $x_2(t)$ can be written:

$$h(t, x_1, x_2(t)) = h_0(t)\exp\{\beta_1 x_1 + \beta_2 x_2(t)\}$$

The approach to deal with time – varying data is splitting individual at risk at the time of any change in any individual's covariate: The aim is to segment, for every policyholder, his policy lifetime within regular time intervals, indicating for each interval all the covariates' parameters. The first thing to do is here to create a new data set, with start and end times at periodic (here, monthly) intervals and a single covariate indicating the surrender status each month. There will be recorded per policyholder per month spent as an investor within the portfolio. Each record will also be containing the relevant covariates selected in the previous Cox model. The covariate Event is here as an indicator variable, mentioning if the policyholder was surrendering during the studied interval of time, or censored.

- In case of surrendering during the studied monthly period, we study a new policyholder profile, starting to zero and studying each month of his policy lifetime.
- In case of no surrender during the studied monthly period –indicator censored, all the policyholder covariates will be replaced for the next month, identical to all respects expect with a new value of surrender status, and market rate

2.3. Parametric model

The Cox PH model described earlier is the most common way for analyzing prognostic factors on survival data. This is probably due to the fact that this model allows to estimate parameters without assuming any distribution on the survival time.
However, when the proportional hazards method is challenged and/or not acceptable, these models are not suitable anymore.
This section aims to present both parametric proportional hazards and accelerated failure time models [22].

2.3.1. Parametric proportional hazards model

The parametric proportional hazards model is the parametric version of the Cox PH model, and is consequently expressed with a similar form. The hazard function at time t for a policyholder characterized by a set of p covariates $(x_1, x_2, \dots, x_p)$ can be expressed as such:

$$h(t|x) = h_0(t)\exp\left(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p\right) = h_0(t)\exp(\beta' x)$$

The essential difference between these two kinds of models is that the baseline hazard function is assumed to be a specific distribution when a fully parametric PH model fits to the data, whereas the Cox PH one has no such constraint. These coefficients are estimated by full likelihood – while partial in the Cox PH model. Otherwise, these two models are similar, hazards ratios have the same interpratation and proportionality of hazards is still assumed. These models are usually suitable in the case of exponential , Weibull, or Gompertz models.

2.3.1.1. Weibull PH model

Let's assume that the survival time T follows a Weibull law, $T \sim Weibull(\lambda, p)$, with a probability function $f(t) = \lambda p t^{p-1}\exp(-\lambda t^p)$, with $\lambda > 0$ and $p > 0$.

The hazard function is given by $h(t) = \lambda p t^{p-1}$

p is called the shape parameter;

- If $p > 1$, the hazard increases
- If $p = 1$, the hazard is constant (exponential model scenario)
- If $p < 1$, the hazard decreases

As particularities, the Weibull function has its$\ln(-\ln S(t))$ function linear with the time logarithm; indeed,

$$S(t) = \exp(\lambda t^p)$$

$$-\ln\big(S(t)\big) = \lambda t^p$$

$$\ln(-\ln(S(t))) = \ln(\lambda) + pln\,(t)$$

This property allows a graphical evaluation of the appropriateness of a Weibull model by plotting

$\ln(-\ln \hat{S}(t)))$ vs $\ln(t)$, where $\hat{S}(t)$ is the Kaplan-Meier estimate.

We consider here, in addition of all presented before, a survival time function. This “time to event” function is the time until a policyholder surrenders , or not.

The parameter λ consequently needs to be reparametrized as such: $\lambda = \exp(\beta_0 + \beta_1 Surr)$, with $Surr = 1$if the policyholder surrendered his policy, 0 otherwise.

The hazard ratio “surrendering versus not surrendering” can be written as:

$$HR = \frac{\exp(\beta_0 + \beta_1 Surr) pt^{p-1}}{\exp(\beta_0)\, pt^{p-1}} = \exp(\beta_1)$$

This expression indicates that the proportional hazards assumption is satisfied.

However, this expression depends on p having the same value in case of surrender or not (otherwise, the time would not cancel out).

More generally, under the Weibull PH model, the hazard function of a particular patient with covariates $(x_1, x_2, \dots, x_p)$ is given by

$$h(t|x) = \lambda p(t)^{p-1} \exp\big(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p\big) = \lambda p(t)^{p-1} \exp(\beta' x)$$

The survival timeof this patient has the Weibull distribution with scale parameter $\lambda \exp(\beta' x)$ and shape parameter p. Therefore, the Weibull family with fixed $p$has the proportional hazards property. This highlights that the effect of the explanatory variables within the model alters the scale parameter of the distribution, while the shape parameter remains constant.

The survival function can be expressed as such

$$S(t|x) = \exp\{-\exp(\beta' x)\lambda t^p\}$$

After a transformation of the survival function based on the property just seen earlier

$$\ln\{-\ln S(t)\} = \ln \lambda + p \ln t$$

Its plot should return approximately a straight line if the Weibull distribution is reasonable.The intercept and slope of the line will be rough estimates of $\ln \lambda$ and p respectfully. If the two lines for two groups are essential parralel, this means that the proportional hazards model is valid.

Another approach to assess the suitability of a parametric model is to estimate the hazard function using the non – parametric method. If the hazard function increased or decreased monotically with increasing survival time, a Weibull distribution might be considered.

2.3.1.2. Exponential PH model

The exponential model is a special case of the Weibull one, with $p = 1$. The hazard function under this model is assumed constant over time. Both survival and hazard functions are written as

$$S(t) = \exp(-\lambda t) \ \ and \ h(t) = \lambda$$

Under the exponential PH model, the hazard function can consequently be given by

$$h(t|x) = \lambda \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p) = \lambda \exp(\beta' x)$$

In the line, the Weibull density function, expressed as $f(t) = \exp(-\lambda t^p)\lambda p t^{p-1}$, becomes, for $p = 1$:

$$f(t) = \lambda \exp(-\lambda t)$$

Its plot should return approximately a straight line if the exponential distribution is reasonable: if the straight line has a slope nearly one, and goes through the origin, the exponential distribution can be assumed. Besides, if the hazard function is reasonably constant over time, this would suggest an exponential distribution.

2.3.1.3. Gompertz PH model

Both survival and hazard functions of a Gompertz distribution are given by

$$S(t) = \exp\ \left(\frac{\lambda}{\theta}(1 - e^{\theta t})\right)$$

$$h(t) = \lambda \exp(\theta t)$$

for $\lambda > 0$ and $0 \leq t < \infty$.

In this case, the Gompertz distribution, $\ln(h(t))$ is linear with t. The parameter θ determines the shape of the hazard function. When $\theta = 0$, the survival time has an exponential distribution. And like the Weibull hazard function, the Gompertz hazard increases or decreases monotically.

As the two previous cases, the hazard function for a particular patient is expressed as

$$h(t|x) = \lambda \exp\,(\theta t) \exp\left(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p\right) = \lambda \exp(\theta t) \exp(\beta' x)$$

Altough it is straight forward to see that the Gompertz distribution has the proportional hazards property, this model is rarely used in practice. Moreover, most computer softwares for fitting exponential and Weibull models use a different form, the Accelerated Failure Time (AFT) model.

2.3.2. Accelerated Failure Time model

Although parametric PH models are very applicable to analyze survival data, there are relatively a few probability distribution for the survival time which can be used with these models. To face this issue, the AFT model is an interesting alternative to the proportional hazards model for the analysis of survival time data. This is measuring the direct effect of the explanatory variables on the survival time instead of hazard – as done in the PH model. It allows then an easier interpratation of the results, because the parameters measure the effect of the corresponding covariate on the average survival time. On a similar way to the PH model, the AFT model describes the relationship between survival probabilities and a set of covariates.

For convenience reasons, γ is to be refered to the regression coefficients and θ to the accelerating ones.The usual distributions specified for ε in the accelerated model are: Weibull (if shape=1, exponential), log normal and log logisitic.The corresponding distributions for W are : minimum extreme value, normal, logistic [23].

Under an AFT model, the covariate effect are assumed to be constant and multiplicative on the time scale – e.g. the covariate impacts on survival by a constant factor (accelerator factor).
Based on the relationship between between both survival and hazard functions, the hazard function for a policyholder with covariates $(x_1, x_2, \dots, x_p)$ is

$$h(t|x) = \left(\frac{1}{\omega(x)}\right) h_0\left(\frac{t}{\omega(x)}\right)$$

The associated log – linear form, in respect to survival time is

$$\ln(T_i) = \mu + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \cdots + \alpha_p x_{pi} + \sigma \varepsilon_i$$

- μ is the intercept
- σ is the scale parameter
- ε_i a random variable

This expression assumes to have a particular distribution. It is this type of form for the AFT model whichis adopted by most software packages for the AFT model.

The survival function of $T_i$can be expressed by the survival function of ε_i.

$$S_i(t) = P(T_i \geq t)$$

$$S_i(t) = P(\ln(T_i) \geq \ln(t))$$

$$S_i(t) = P(\mu + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \cdots + \alpha_p x_{pi} + \sigma\varepsilon_i \geq \ln t$$

$$S_i(t) = P\left(\varepsilon_i \geq \frac{\ln t - \mu - \alpha X}{\sigma}\right)$$

$$S_i(t) = S_{\varepsilon_i}\left(\frac{\ln t - \mu - \alpha X}{\sigma}\right)$$

For each distribution of ε_i, there is a matching distribution of T. Generally, AFT models are named for the distribution of T rather than the distribution of ε_i or $\ln T$. The distributions of ε_i and the associated distributions of T_i are summarized below

T	**ln (T)**
Exponential	Extreme value
Weibull	Extreme value
Log-logisitc	Logistic
Lognormal	Normal
Log-gamma	Gamma

The effect size for the AFT model is the time ratio. This time ratio, comparing two levels of covariates x_i ($x_i = 1$ versus $x_i = 0$), after controling all the other covariates is $\exp(\alpha_i)$. This expression is the intrepretation of the estimated ratio of the expected survival times for two groups.

A time ratio above 1 for the covariate implies that this covariate prolongs the time to event, while a time ratio below 1 indicates that an earlier event is more likely. Therefore, the AFT model can be intreprated in terms of the speed of progression of the event/ censor within the data. The effect of a covariate in an AFT model is to change the scale, and not the location of the baseline distribution of survival times.

2.3.2.1. Estimation of AFT model

AFT models are fitted using the maximum likelihood method. The likelihood of the N observed survival times $t_1, t_2, \dots, t_N$ is given by

$$L(\alpha, \mu, \sigma) = \prod_{i=1}^{N} \{f_i(t_i)\}^{\delta_i} \{S_i(t_i)\}^{1-\delta_i}$$

Where $f_i(t_i)$ and $S_i(t_i)$ are respectfully the density and survival functionsfor the ith individual at t_i. δ_i is the event indicator for the ith observation.

The log likelyhood function can consequently be expressed as

$$\ln L\,(\alpha,\mu,\sigma) = \sum\nolimits_{i=1}^{N} \{-\delta_i \ln(\sigma t_i + \delta_i \ln f_{\varepsilon_i}(z_i) + (1-\delta_i)\ln S_{\varepsilon_i}(z_i))\}$$

Where $z_i = \frac{\ln t_i - \mu - \alpha_1 x_{1i} - \alpha_2 x_{2i} - \cdots - \alpha_p x_{pi}}{\sigma}$. The maximum likelihood estimates the the unknown parameters $\mu, \sigma, \alpha_1, \alpha_2, \dots, \alpha_p$, by maximising this function using the Newton Raphson procedure – same methodology in use for the partial likelihood in the Cox regression model.

2.3.2.2. *Weibull AFT model*

Let's assume the survival time T has $W(\lambda,\gamma)$ distribution with a scale parameter λ and shape paramater γ. The hazard function under the FAT model for the ith policyholder is

$$h_i(t) = \left[\frac{1}{\omega_i(x)}\right] h_0 \left[\frac{t}{\omega_i(x)}\right]$$

$$h_i(t) = \left[\frac{1}{\omega_i(x)}\right] \lambda\gamma \left(\frac{t}{\omega_i(x)}\right)^{\gamma-1}$$

$$h_i(t) = \left(\frac{1}{\omega_i(x)}\right)^{\gamma} \lambda\gamma(t)^{\gamma-1}$$

Where $\omega_i(x) = \exp(\alpha_1 x_{1i} + \alpha_2 x_{2i} + \cdots + \alpha_p x_{pi}$ for a policyholder i with p explanatory covariates. So the survival time for the ith policholder is $W\left(\lambda\left(\frac{1}{\omega_i(x)}\right)^{\gamma}, \gamma\right)$. The Weibull distribution consequently has the Accelerated Failure Time property.

If T_i has a Weibull distribution, then ε_i has an extreme value distribution (Gumbel distribution). The survival function of Gumbel distribution is given by $S_{\varepsilon_i}(t) = \exp(-\exp t))$.

The AFT representation of the survival function of the Weibull model is given by

$$S_i(t) = \exp\left[-\exp\left(\frac{\ln t - \mu - \alpha_1 x_{1i} - \alpha_2 x_{2i} - \cdots - \alpha_p x_{pi}}{\sigma}\right)\right]$$

$$S_i(t) = \exp\left[-\exp\left(\frac{-\mu - \alpha_1 x_{1i} - \alpha_2 x_{2i} - \cdots - \alpha_p x_{pi}}{\sigma}\right) t^{1/\sigma}\right]$$

From this expression, the proportional hazards representation of the Weibull model is given by $S_i(t) = \exp[-\exp(\beta_1 x_{1i} + \cdots + \beta_p x_{pi})\,\lambda t^{\gamma}]$

Using the two last formulas, the AFT parameters λ, γ, β_j in the PH model can be expressed by the parameters μ, σ, α_j

$$\lambda = \exp\left(-\frac{\mu}{\sigma}\right), \gamma = \frac{1}{\sigma}, \beta_j = -\frac{\alpha_j}{\sigma}$$

As a reminder, both hazard and survival functions are expressed as

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d \ln S(t)}{dt}$$

$$S(t) = exp\left[-\int_0^t h(u)du\right]$$

Hence the AFT representation of the Weibull hazard function:

$$h_i(t) = \frac{1}{\sigma} t^{\frac{1}{\sigma}-1} exp\left(\frac{-\mu - \alpha_1 x_{1i} - \alpha_2 x_{2i} - \cdots - \alpha_p x_{pi}}{\sigma}\right)$$

The approximate variance of a function of two parameters $\theta_1$et $\theta_2$is given by

$$\left(\frac{\partial g}{\partial \hat{\theta}_1}\right)^2 V(\hat{\theta}_1) + \left(\frac{\partial g}{\partial \hat{\theta}_2}\right)^2 V(\hat{\theta}_2) + 2\left(\frac{\partial g}{\partial \hat{\theta}_1}\frac{\partial g}{\partial \hat{\theta}_2}\right) Cov(\hat{\theta}_1, \hat{\theta}_2)$$

The standard error of $\hat{\beta}_j$ is expressed as

$$V(\hat{\beta}_j) = \left(-\frac{1}{\hat{\sigma}}\right)^2 V(\hat{\alpha}_j) + \left(\frac{\hat{\alpha}_j}{\hat{\sigma}^2}\right)^2 V(\hat{\sigma}) + 2\left(-\frac{1}{\hat{\sigma}}\right)\left(\frac{\hat{\alpha}_j}{\hat{\sigma}^2}\right) Cov\,(\hat{\alpha}_j, \hat{\sigma})$$

2.3.2.3. Log-logistic AFT model

The Weibull hazard function has one limit, which is a monotonic function of time. However, the hazard can change of directions and vary over time in some situations.

Log-logistic survival and hazard function are given by

$S(t) = \frac{1}{1+e^{\theta}t^k}$ and $h(t) = \frac{e^{\theta}kt^{k-1}}{1+e^{\theta}t^k}$

Where θ and k, $k > 0$, are unknown parameters.

- When $k \leq 1$, the hazard rate decreases monotonically
- When$k > 1$, it increases from zero to a maximum and then decreases to zero.

Assuming the survival times have a log-logistic distribution with parameter θ and k, the hazard function under the AFT model for the ith policyholder is

$$h_i(t) = \left(\frac{1}{\omega_i}\right) h_0\left(\frac{t}{\omega_i}\right)$$

$$h_i(t) = \frac{e^{\theta} k \left(\frac{t}{\omega_i}\right)^{k-1}}{\omega_i\left(1 + e^{\theta}\left(\frac{t}{\omega_i}\right)^k\right)}$$

$$h_i(t) = \frac{e^{\theta - k\ln\omega_i}\, k\, t^k}{1 + e^{\theta - k\ln\omega_i} t^k}$$

Thus, the survival time for the ith policyholder has a log-logistic distribution with parameter $\theta - k\ln\omega_i$ and k. As a consequence, the log – logistic distribution has the Accelerated Failure Time property.

If the baseline survival function is $S_0(t) = \frac{1}{1+e^{\theta}t^k}$, with θ and k unknown parameters, the baseline odds of surviving beyond time t are given by

$$\frac{S_0(t)}{1 - S_0(t)} = e^{-\theta} t^{-k}$$

The survival time for the ith policyholder also has a log-logistic destribution, which is

$$S_i(t) = \frac{1}{1 + e^{\theta - k\ln\omega_i} t^k}$$

The odds for the ith policyholder surviving beyond time t is then

$$\frac{S_i(t)}{1 - S_i(t)} = \exp(\beta_1 x_{1i} + \cdots + \beta_p x_{pi}) \frac{S_0(t)}{1 - S_0(t)}$$

The logarith of the ith policyholder surviving beyond time t returns

$$\ln \frac{S_i(t)}{1 - S_i(t)} = \beta x_i - \theta - k\ln t$$

Where x_i is the censor variable, taking the value 1 in case of event, and 0 otherwise (censor). A plot of $\ln\left(\frac{S(t)}{1-S(t)}\right)$ versus $\ln t$ should be linear of the log – logistic distribution is appropriate.

If T_i has a log – logistic distribution, then ε_i has a logistic distribution. The survival function of a logistic distribution is

$$S_{\varepsilon_i}(t) = \frac{1}{1 + \exp t}$$

The expression of the survival function of a log – logistic model becomes

$$S_i(t) = \left[1 + t^{1/\sigma} exp\left(\frac{-\mu - \alpha_1 x_{1i} - \alpha_2 x_{2i} - \cdots - \alpha_p x_{pi}}{\sigma}\right)\right]^{-1}$$

The two expressions of $S_i(t)$ leads to the expressions of

$$\theta = -\frac{\mu}{\sigma}, k = \frac{1}{\sigma}$$

According to the relationship between survival and hazard function, the hazard function for the ith policyholder is

$$h_i(t) = \frac{1}{\sigma t}\left\{1 + t^{-\frac{1}{\sigma}} \exp\left(\frac{-\mu - \alpha_1 x_{1i} - \alpha_2 x_{2i} - \cdots - \alpha_p x_{pi}}{\sigma}\right)\right\}^{-1}$$

2.3.2.4. Log – normal AFT model

The assumption of having survival times following a log – normal distribution leads to an expression of the baseline survival and hazards functionswritten below:

$$S_0(t) = 1 - \emptyset\left(\frac{\ln t - \mu}{\sigma}\right), h_0(t) = \frac{\varphi\left(\frac{\ln t}{\sigma}\right)}{1 - \emptyset\left(\frac{\ln t}{\sigma}\right)\sigma t}$$

Where μ and σ are parameters, $\varphi(x)$ is the probability density function and $\emptyset(x)$ is the cumulative density function of the standard normal distribution. The survival function fot the ith policyholder is

$$S_i(t) = S_0\left(\frac{t}{\omega_i}\right) = 1 - \emptyset\left(\frac{\ln t - \alpha' x_i - \mu}{\sigma}\right)$$

The log – survival time for the ith policyholder has normal $(\mu + \alpha' x_i, \sigma)$. The log – normal distribution has the Accelerated Failure Time property.

It comes, in a two –groups study, with x_i the censor covariate (1 in case of event, 0 otherwise – censor), the relation

$$\emptyset^{-1}\left(1 - S(t)\right) = \frac{1}{\sigma}(\ln t - \alpha' x_i - \mu)$$

The plot of $\emptyset^{-1}\left(1 - S(t)\right)$ versus $\ln t$ will be linear if the log – normal distribution is adequate.

2.3.2.5. Gamma AFT model

The (generalized) gamma model is described by a probability density function of three parameters, $\lambda > 0, \alpha > 0$ and $\gamma > 0$

$$\forall t > 0, f(t) = \frac{\alpha\lambda^{\alpha\gamma}}{G(\gamma)} t^{\alpha\gamma - 1} \exp[-(\lambda t)^{\alpha}]$$

Where γ is the shape parameter of the distribution. Both survival and hazard functions do not have a closed form for the generalized gamma distribution. Exponential, Weibull and log – normal models are all special cases of the gamma model. In case of

- $\gamma = \alpha = 1$, the generalized gamma distribution becomes the exponential distribution
- $\gamma = 1$, the generalized gamma distribution becomes the Weibull distribution
- $\gamma \to \infty$, the generalized gamma distribution becomes the log – normal distribution

This highlights the fact that the generalized gamma model can have a wide variety of shapes, except for any of the special cases.

2.3.3. Checks on models

2.3.3.1. Graphical

Graphical methods can be used in the first place to check if the paramteric distribution fits the observed data. As we saw before, we have:

- If the survival time follows an exponential distribution, the plot of $\ln[-\ln S(t)]$ versus $\ln t$ should yield a straight line with slope of 1
- If the plots are parallel but not straight, then the proportional hazards assumption is valid, but not the Weibull model. If the lines for two groups are straight but not parallel, the Weibull model assumption is supported, but both PH and AFT assumptions are violated.
- The log –logistic assumption can be checked by plotting $\ln\left(\frac{1-S(t)}{S(t)}\right)$ versus $\ln t$. If the distribution of survival function is log – logistic then the result plot should be a straight line.
- For the log – normal distribution, a plot of $\emptyset^{-1}\left(1-S(t)\right)$ versus $\ln t$ should be linear

However, all these plots are based on the underlying assumption that the data sample on where the model is fitted is drawn from an homogeneous population – implying then that no covariates are taken into account. This consequently makes this graphical checking method not very reliable in practice.

2.3.3.2. Quantile – quantile plot

One method for assessing the potential of an AFT model is to produce the quantile – quantile plot. For any value of p within the interval [0;100], the pth percentile is

$$t(p) = S^{-1}\left(\frac{100-p}{100}\right)$$

Let's consider $t_0(p)$ and $t_1(p)$ the pth percentiles estilmated from the survival functions of two groups of survival data. The percentiles for the two groups may be expressed as

$$t_0(p) = {S_0}^{-1}\left(\frac{100-p}{100}\right), t_0(p) = {S_1}^{-1}\left(\frac{100-p}{100}\right)$$

Where $S_0(t)$ et $S_1(t)$ are the survival functions for the two groups. Hence,

$$S_1[t_1(p)] = S_0[t_0(p)]$$

Under the AFT model,

$$S_1(t) = S_0\left(\frac{t}{\omega}\right)$$

So as a deduction

$$t_0(p) = \ \omega^{-1}t_1(p)$$

The percentiles of the survival distributions for the two groups can be estimated with the Kaplan – Meier estimator. A plot of percentiles of the Kaplan – Meier estimated survival function from one group against another should give an approximate straight line through the origin if the Accelerated Failure Time model is appropriated. The slope of this line will be an estimate of the acceleration factor ω^{-1}.

2.3.3.3. Statistical criterias

AFT models can be compared between them with statistical tests. Nested models, as Exponential, Weibull and log-normal – nested within the gamma model, can be compared using the likelihood ratio test.

Otherwise, for nested and non – nested models, the Akaike Information Criterion (AIC) can be used. This statistical criteria is defined as

$$AIC = -2l + 2(k + c)$$

Where l is the log – likelihood, k the number of covariates within the model, and c the number of model specific parameters. The lowest the AIC value is, the better the model is.

However, one difficulty remains by using the AIC; there are no other statistical equivalent tests to compare the AIC with. This makes the choice between two models with close AIC values difficult.

2.4. Fit a survival model

Based on what has been presented on this chapter, we decide here to fit a Cox PH model deeply on our data set [Annex 4, R code]. The reason of this decision is the convenience of this model, e.g. not making any assumption on the distribution times distribution, particularly because of the high number of failures (surrenders).

Indeed, the main drawback of a parameterical model is its potential for arbitrary decisions regarding the nature of the baseline hazard rate. On the other hand, the relationship between covariates and the hazard rate in the Cox model can be estimated without having to make any assumptions about the nature and shape of the baseline hazard rate.

In this sense, the less assumptions we are making on the data, the better. That why the semi – parametrical model is studied deeper here. However, in order to compare both models, we will make a short parametric study at the end and compare the two survival times, in order to validate our choice.

2.4.1. Data

2.4.1.1. Data presentation

The file “Data.csv” gathers all the surrender data of the Camelea portfolio. More exactly, we are talking here of 9 159 policyholders, allowed to surrender their life insurance policies at any time. There were 8,495 policyholders as of 31th of December, 2013. The 664 remaining to reach 9,159 are policyholders who fully surrendered their life insurance policy before end of December 2013.

This data has been, on one part, generated by BOXI, a business object interface in use within NPGWM. A second part comes from Bloomberg

It is based on this data set that all the results from the various studies which will be detailed below will come from.

2.4.1.2. BOXI Data

- **Duration**: The duration corresponds to the time the policyholder remains in the portfolio until the 31th of December, 2013.
- **Event**: censor indicator; it is equal to 1 if the policyholder surrenders at least once his portfolio – partially and/or fully, 0 otherwise – right censored data
- **Still**: Indicates if the policyholder surrendered his policy partially or fully. If the covariate is equal to 1, it’s a full surrender and the policyholder is no longer within the Camelea portfolio. If the covariate is equal to 0, then the policyholder is still in the Camelea portfolio as of 31th of December, 2013; he could have surrendered his policy, but only partially.
- **Savings**: Amount of savings of each policyholder, on a €100,000 basis
- **Age**: the policyholder age as of the time he fully surrendered his policy, and as of the 31th of December 2013 otherwise.
- **Gender**: 0 for male, 1 for female
- **Job**: the job occupation for each policyholder

Job occupation	Job occupation code
UNEMPLOYED	1
STUDENT	2
STATE EMPLOYEE (PUBLIC SECTOR)	3
SELF-EMPLOYED/SHOPKEEPER (LEGAL ENTITY)	4
SELF-EMPLOYED/SHOPKEEPER (INDIVIDUAL)	5
RETIRED	6
PRIVATE/INDEPENDENT PRACTICE	7
OTHER	8
EXECUTIVE	9
EMPLOYEE (PRIVATE SECTOR)	10
COMPANY DIRECTOR	11

- **Gender:**0 for male, 1 for female
- **Risk:** The NPGWM compliance department set up some items to define a low (covariate coded 0) or high (covariate coded 1) risk policyholder profile.
  Are defined as high risk profile
    - Individuals who hold or have been entrusted with prominent public functions: Heads of State, heads of government, ministers, ambassadors, members of supreme courts, political parties' responsible persons ...
    - Immediate family members: spouse, husband, partners, children, parents...
    - Known associates: any natural person who is known to have joint beneficial ownership of legal entities or legal arrangements, or any other close business relations, with a person holding or entrusted with prominent public functions

  Additionally, enhanced due diligence measures are taken for each new policyholder. NPGWM mustestablish the source of wealth and funds involved in the business relationship or transaction with great care and details.

2.4.1.3. Bloomberg data

- **Unemployment:**This covariate is the difference of Belgium unemployment rate between the time the policyholder suscribes his life insurance policy, and the time he surrenders. If the policyholder did not surrender, the covariate is the difference between the suscription time and and Belgium unemployment rate as of 31th of December, 2013. The purpose of this index is to notice if policyholders are surrendering because they are facing unemployment
- **SX5T**: This covariate is the difference of the European MSCI equity market rates between two dates, the suscription time and the surrender time (31/12/2013 value if the policyholder did not surrender before this date). I chose this index, which I think is quite well representative of the equity market all around the world. The purpose of taking an european equity index is to see if policyholders are surrendering for investing in more profitable markets or not.

The data is summarized and consequently presented as such on the R software

```
R Console
> summary(dat)
     Policy        Duration           Event            Still           Portfolio
 Min.   :   1   Min.   :0.00274   Min.   :0.0000   Min.   :0.0000   Min.   : 0.0031
 1st Qu.:2290   1st Qu.:1.00000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.: 0.2298
 Median :4580   Median :1.76712   Median :0.0000   Median :0.0000   Median : 0.5595
 Mean   :4580   Mean   :1.91989   Mean   :0.3731   Mean   :0.1856   Mean   : 1.3827
 3rd Qu.:6870   3rd Qu.:2.78356   3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.: 1.4724
 Max.   :9159   Max.   :5.06575   Max.   :1.0000   Max.   :1.0000   Max.   :94.4390

   Surrender          DeltaU            Unpmt           DeltaS            SX5T
 Min.   :0.0000   Min.   :-1.3000   Min.   :7.100   Min.   :-13.917   Min.   :27.72
 1st Qu.:0.0000   1st Qu.: 0.0000   1st Qu.:7.400   1st Qu.:  2.868   1st Qu.:42.22
 Median :0.0000   Median : 0.3000   Median :8.000   Median :  8.850   Median :45.21
 Mean   :0.2596   Mean   : 0.3267   Mean   :7.883   Mean   :  7.642   Mean   :44.57
 3rd Qu.:0.4349   3rd Qu.: 0.8000   3rd Qu.:8.400   3rd Qu.: 12.278   3rd Qu.:47.09
 Max.   :1.0000   Max.   : 1.4000   Max.   :8.500   Max.   : 28.534   Max.   :56.30

      Age             Risk              Job             Gender
 Min.   : 6.00   Min.   :0.00000   Min.   : 1.000   Min.   :0.0000
 1st Qu.:46.00   1st Qu.:0.00000   1st Qu.: 5.000   1st Qu.:0.0000
 Median :57.00   Median :0.00000   Median : 6.000   Median :0.0000
 Mean   :56.55   Mean   :0.02238   Mean   : 6.395   Mean   :0.2796
 3rd Qu.:67.00   3rd Qu.:0.00000   3rd Qu.: 9.000   3rd Qu.:1.0000
 Max.   :96.00   Max.   :1.00000   Max.   :11.000   Max.   :1.0000
                 NA's   :1
>
```

```
R Console
> dat[200:205,1:14]
    Policy   Duration Event Still Portfolio Surrender DeltaU Unpmt DeltaS    SX5T Age Risk Job Gender
200    200 0.09041096     0     0 0.5027238 0.0000000      0   8.4 0.3505 55.9032  59    0   7      0
201    201 0.09041096     0     0 0.7578059 0.0000000      0   8.4 0.3505 55.9032  79    0   1      1
202    202 0.09041096     0     0 2.5121156 0.0000000      0   8.4 0.3505 55.9032  30    0  10      1
203    210 0.09315068     1     0 9.2911198 0.1076297      0   8.4 3.6380 45.7201  74    0   6      0
204    203 0.09315068     0     0 0.1005767 0.0000000      0   8.4 0.5336 55.7201  78    0   6      1
205    204 0.09315068     0     0 0.1329600 0.0000000      0   8.4 0.5336 55.7201  37    0  10      0
>
```

2.4.1.4. Correlation between covariates

The first step to do with this data set is studying the dependence between covariates. Having a quick look on the correlation gives us a first idea of how covariates are linked to each other

R Console

```
> cor(dat[,4:14])
                 Still    Portfolio    Surrender      DeltaU        Unpmt      DeltaS
Still      1.000000000 -0.008939408  0.852941754 -0.20408765  0.007392209 -0.38028082
Portfolio -0.008939408  1.000000000 -0.003876404 -0.04304885  0.027755254 -0.03007817
Surrender  0.852941754 -0.003876404  1.000000000 -0.31992796  0.010591803 -0.52682738
DeltaU    -0.204087646 -0.043048855 -0.319927962  1.00000000 -0.776832690  0.57834692
Unpmt      0.007392209  0.027755254  0.010591803 -0.77683269  1.000000000 -0.22085870
DeltaS    -0.380280816 -0.030078166 -0.526827380  0.57834692 -0.220858699  1.00000000
SX5T      -0.066182949 -0.026113995 -0.105033516 -0.20113263  0.346959459 -0.52076528
Age        0.080542419  0.030240863  0.103174415 -0.11233402  0.055785316 -0.10680439
Risk                NA           NA           NA          NA           NA          NA
Job       -0.028533789 -0.040187239 -0.043725842  0.04984010 -0.042634256  0.03898898
Gender    -0.284914465  0.009104108 -0.226413009  0.01052709  0.026694690  0.05395001
                 SX5T         Age Risk         Job       Gender
Still     -0.06618295  0.08054242   NA -0.02853379 -0.284914465
Portfolio -0.02611399  0.03024086   NA -0.04018724  0.009104108
Surrender -0.10503352  0.10317442   NA -0.04372584 -0.226413009
DeltaU    -0.20113263 -0.11233402   NA  0.04984010  0.010527092
Unpmt      0.34695946  0.05578532   NA -0.04263426  0.026694690
DeltaS    -0.52076528 -0.10680439   NA  0.03898898  0.053950010
SX5T       1.00000000 -0.02352910   NA  0.00190440  0.036631195
Age       -0.02352910  1.00000000   NA -0.11324816 -0.032905080
Risk               NA          NA    1          NA           NA
Job        0.00190440 -0.11324816   NA  1.00000000 -0.082351084
Gender     0.03663120 -0.03290508   NA -0.08235108  1.000000000
> 
```

In statistics, dependence is any statistical relationship between two random variables or two sets of data. Correlation refers to any of a broad class of statistical relationships involving dependence. Correlations are useful because they can indicate a predictive relationship that can be exploited in practice. Formally, dependence refers to any situation in which random variables do not satisfy a mathematical condition of probabilistic independence.

Among others, some observations from the output:

- Market and unemployment rates are strongly linked
- There is a link between the proportion to surrender, market rates and gender
- Minor correlations between covariates otherwise

2.4.2. Cox Proportional hazards model

In R, the Cox PHM can fit the data with the adequate packages. This requires a formula object whose form is $Surv(\,)\sim covariates$, named *coxph*, from the *survival* package.

2.4.2.1. Adjustement of a Cox PH Model

The code for adjusting a Cox model with R is as follows:

$Call$:

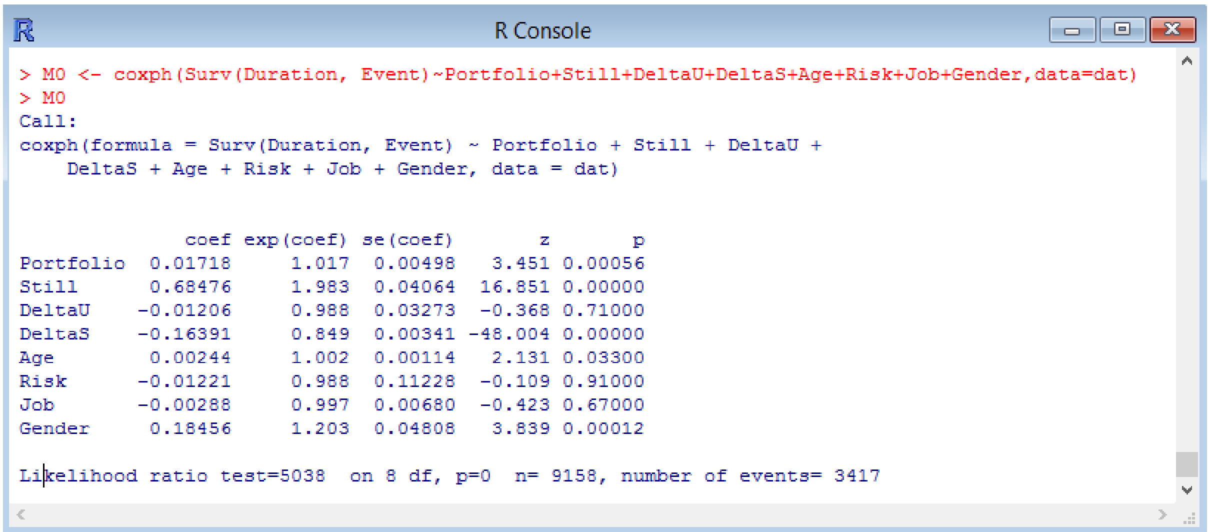
$coxph\ (formula =$

$Surv(Duration, Event) \sim Portfolio + Still + SX5T + DeltaU + DeltaS + Age + Risk + Job + Gender$

The adjusted model is as follows:

$$h(duration) = h_0(duration)exp^{\beta_1 Portfolio+ \beta_2 Still+\beta_3 DeltaU+\beta_4 DeltaS+\beta_5 Age+\beta_6 Risk+\beta_7 Job+\beta_8 Gender}$$

The R output gives:



```
R Console
> M0 <- coxph(Surv(Duration, Event)~Portfolio+Still+DeltaU+DeltaS+Age+Risk+Job+Gender,data=dat)
> M0
Call:
coxph(formula = Surv(Duration, Event) ~ Portfolio + Still + DeltaU +
    DeltaS + Age + Risk + Job + Gender, data = dat)


              coef exp(coef) se(coef)       z       p
Portfolio  0.01718     1.017  0.00498   3.451 0.00056
Still      0.68476     1.983  0.04064  16.851 0.00000
DeltaU    -0.01206     0.988  0.03273  -0.368 0.71000
DeltaS    -0.16391     0.849  0.00341 -48.004 0.00000
Age        0.00244     1.002  0.00114   2.131 0.03300
Risk      -0.01221     0.988  0.11228  -0.109 0.91000
Job       -0.00288     0.997  0.00680  -0.423 0.67000
Gender     0.18456     1.203  0.04808   3.839 0.00012

Likelihood ratio test=5038  on 8 df, p=0  n= 9158, number of events= 3417
```

The R output on the Cox model summarizes, for all $j = 1, \dots, 7$:

- $coef_j = \widehat{\beta_j}$ the parameter for each covariate
- $\exp(coef_j) = e^{\widehat{\beta_j}}$
- The main hypothesis to test is $H_0 : \beta_j = 0, e.g. \beta_1 = \dots = \beta_7 = 0$,with $z_j = \frac{\sqrt{n}\widehat{\beta_j}}{\sqrt{\widehat{V}(\widehat{\beta_j})}}$ the Wald statistical value
- Finally, $p_j = P(U > z_j)$, the p-value for each covariate – with $U \sim \aleph(0,1)$
- The likelihood ratio is the statistical value of the maximum likelihood test
- df is the abbreviation od degree of freedom, which corresponds here to the number of covariates
- p is the p-value of the global test, n the total number of individuals, and the number of events corresponds here to the total number of surrenders within the data sample

```
                coef exp(coef) se(coef)       z Pr(>|z|)
Portfolio  0.017183  1.017332 0.004979   3.451 0.000558 ***
Still      0.684763  1.983302 0.040636  16.851  < 2e-16 ***
DeltaU    -0.012056  0.988016 0.032733  -0.368 0.712643
DeltaS    -0.163907  0.848821 0.003414 -48.004  < 2e-16 ***
Risk      -0.012209  0.987865 0.112282  -0.109 0.913414
Job       -0.002878  0.997126 0.006802  -0.423 0.672188
Age        0.002439  1.002442 0.001144   2.131 0.033078 *
Gender     0.184557  1.202685 0.048078   3.839 0.000124 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

          exp(coef) exp(-coef) lower .95 upper .95
Portfolio    1.0173     0.9830    1.0075    1.0273
Still        1.9833     0.5042    1.8315    2.1477
DeltaU       0.9880     1.0121    0.9266    1.0535
DeltaS       0.8488     1.1781    0.8432    0.8545
Risk         0.9879     1.0123    0.7927    1.2310
Job          0.9971     1.0029    0.9839    1.0105
Age          1.0024     0.9976    1.0002    1.0047
Gender       1.2027     0.8315    1.0945    1.3215

Concordance= 0.822  (se = 0.006 )
Rsquare= 0.423   (max possible= 0.998 )
Likelihood ratio test= 5038  on 8 df,   p=0
Wald test            = 4360  on 8 df,   p=0
Score (logrank) test = 5605  on 8 df,   p=0

>
```

The *summary* of the *coxph function* returns the values of three statistical tests (Likelihood ratio, Wald and Logrank) for the test of $H_0 : \beta_j = 0, e.g. \beta_1 = \ldots = \beta_7 = 0$, with the corresponding degrees of freedom (df = 7) and p values (p= ...) for the statistical maximal law under the test H_0.

2.4.2.2. Interpretation of the outputs

The three p-values calculated by R (Wald, Log – Rank, Likelihood) are all inferior to 5%: Consequently, it does exist at least one covariate which has an impact on the surrender rate. The adjustment of a Cox PH model with a 5% threshold is, as a conclusion, coherent!

2.4.2.3. Interpetation of the results on a covariate basis

2.4.2.3.1. Relevant covariates

Based on the summary results of the coxph test, we have to investigate on the covariates which impact significantly the surrender rate (p-value < 5%)).

The Wald test is testing each covariate while including all the others into the model. If the test is not significant, e.g. a p-value largely higher than 5%, it means that the test would not be significant either with a model built with this covariate only.

- Wald tests for the covariates Portfolio, Still, DeltaS and Gender are highly significant, and have a large impact on the surrender rate (p-value << 5%)
- Age has a marginal impact, much lower than the four previous covariates (p-value $p \leq 5\%$)
- Conversely, covariatesDeltaU, Risk and Job do not impact at all the surrender rate (p-value$p \gg 5\%$). These covariates have no impact on the surrender rate when the other covariates are included in the model.

The amount of savings (Portfolio), being for a long time within the portfolio (Still), the evolution of financial markets on the stock exchanges (DeltaS) and the sex of the policyholder (Gender) have a significant impact on the duration until surrendering. Conversely, the age (Age) has a marginal effect on the duration. The unemployment (DeltaU), the job occupation (Job) and the risk profile (Risk) do not impact significantly the duration when the other covariates are already within the model.

2.4.2.3.2. Covariates' interaction

The exponential of the coefficients measures the multiplicative effect of a one-unit increase of the covariate on the surrender rate, the other covariates remaining unchanged, constant.

R Console

```
          exp(coef) exp(-coef) lower .95 upper .95
Portfolio    1.0173     0.9830    1.0075    1.0273
Still        1.9833     0.5042    1.8315    2.1477
DeltaU       0.9880     1.0121    0.9266    1.0535
DeltaS       0.8488     1.1781    0.8432    0.8545
Risk         0.9879     1.0123    0.7927    1.2310
Job          0.9971     1.0029    0.9839    1.0105
Age          1.0024     0.9976    1.0002    1.0047
Gender       1.2027     0.8315    1.0945    1.3215
```

This output summarizes the exponential of the coefficient for each variabe, and the 95% confidence interval linked to it:

$$IC_j^{95\%} = [lower.95, upper.95] = \left[e^{\hat{\beta}_j - 1.96\sqrt{\hat{V}(\hat{\beta}_j)}};\ e^{\hat{\beta}_j + 1.96\sqrt{\hat{V}(\hat{\beta}_j)}}\right]$$

These confidence intervals are built on the R software based on the fact that

$$\frac{\widehat{\beta_j} - \beta_j}{\sqrt{\hat{V}(\widehat{\beta_j})}} \rightarrow \aleph(0,1).$$

2.4.2.3.3. Interpretation of the first results

- The richer the policyholder is, the soonest he surrenders: Assuming all other covariates remaining constant, wealth has a negative impact on the duration until surrendering. The marginal effect of an increase of 100,000€ of savings increases the annual lapse rate with an average factor of $e^{1.0173}$, e.g. by (101,7-100)% = 1.7%
- The older the policyholder is, the more he surrenders: Assuming all covariates remaining constant, age has a negative impact on the duration on the Camelea portfolio until surrendering. The marginal effect of an increase of one year old increases the annual lapse rate with an average factor of $e^{0.0024}$, e.g. by (100.2-100) = 0.2%
- The more volatile financial markets are, the less the policyholder surrenders: Assuming all covariates remaining constant, market movements have a positive impact on the duration until surrendering. The marginal effect of an increase of 100 basis points on European financial markets decreases the annual lapse rate with an average factor of $e^{-0.1639}$, e.g. by (100-84.9)% = 15.1%
- The gender has an impact on the annual lapse rate. Women are more likely to surrender their policies of (120.3 – 100)% = 20.3% that men. However, this result should be pondered. Men represent more than 66% of the Camelea portfolio
- The Still covariate returns a coherent result; if the covarate becomes 1 instead of 0, it means that the policyholder fully surrendered his life insurance policy. It consequently increases his surrender risk to 100%. Based on the Cox PH model results, this increase is shlightly lower 100%: (198.33 – 100)% =98.33% – the 1.77% difference coming from the model approximations.

2.4.2.4. Selection of significant covariates

Based on what has been presented before, what we have to do now is selecting the right significant covariates, by removing from the model the covariates for which the p-value is higher than 5%.

We get a second model, cleaned of all non-significant covariates:

R Console

```
> M0 <- coxph(Surv(Duration, Event)~Portfolio+Still+DeltaS+Age+Gender,data=dat)
> M0
Call:
coxph(formula = Surv(Duration, Event) ~ Portfolio + Still + DeltaS +
    Age + Gender, data = dat)


              coef exp(coef) se(coef)      z       p
Portfolio  0.01736     1.018  0.00496   3.50 0.00047
Still      0.68527     1.984  0.04052  16.91 0.00000
DeltaS    -0.16457     0.848  0.00295 -55.84 0.00000
Age        0.00252     1.003  0.00113   2.24 0.02500
Gender     0.18732     1.206  0.04784   3.92 0.00009

Likelihood ratio test=5040  on 5 df, p=0  n= 9159, number of events= 3417
>
```

```
R Console
> summary(M0)
Call:
coxph(formula = Surv(Duration, Event) ~ Portfolio + Still + DeltaS +
    Age + Gender, data = dat)

  n= 9159, number of events= 3417

               coef exp(coef)  se(coef)       z Pr(>|z|)
Portfolio  0.017363  1.017515  0.004963   3.499 0.000467 ***
Still      0.685270  1.984307  0.040520  16.912  < 2e-16 ***
DeltaS    -0.164574  0.848255  0.002947 -55.843  < 2e-16 ***
Age        0.002524  1.002527  0.001128   2.236 0.025326 *
Gender     0.187323  1.206017  0.047842   3.915 9.02e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

          exp(coef) exp(-coef) lower .95 upper .95
Portfolio    1.0175     0.9828    1.0077    1.0275
Still        1.9843     0.5040    1.8328    2.1483
DeltaS       0.8483     1.1789    0.8434    0.8532
Age          1.0025     0.9975    1.0003    1.0047
Gender       1.2060     0.8292    1.0981    1.3246

Concordance= 0.822  (se = 0.006 )
Rsquare= 0.423   (max possible= 0.998 )
Likelihood ratio test= 5040  on 5 df,   p=0
Wald test            = 4362  on 5 df,   p=0
Score (logrank) test = 5582  on 5 df,   p=0

> |
```

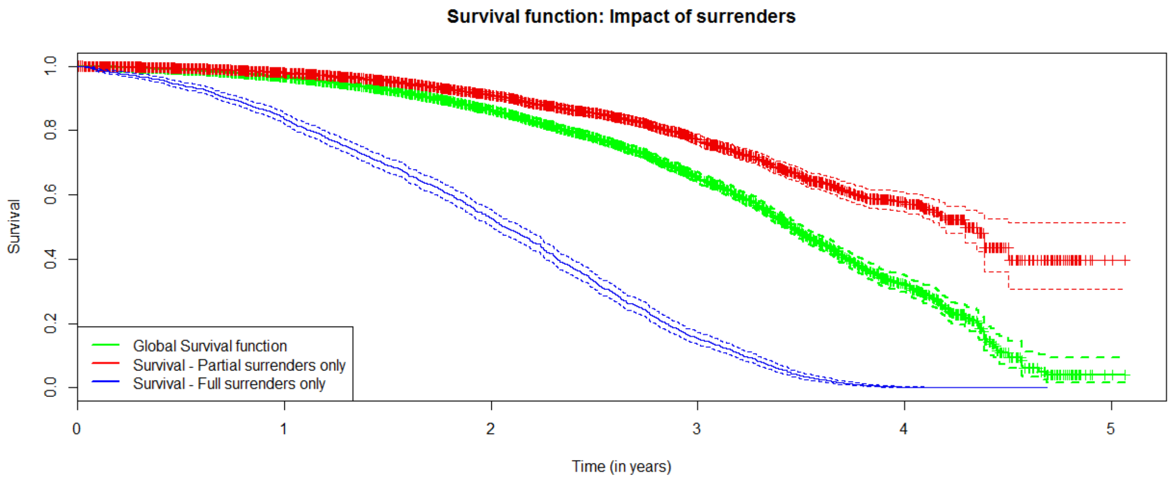
2.4.2.5. *Survival function*

Once all covariates are significant, we get interested on the survival function of the Cox regression on time until surrendering, based on all the valid (e.g. significant for the model) covariates.
Indeed, having a Cox model fit to the data is generally in the purpose to examine the estimated distribution of survival times.
To do so, we use the *R survfit* function, which estimates the survival one with taking by default the mean values of the covariates.

For a policyholder i, the survival function is as such:

$$\widehat{S_\iota}(t) = \hat{S}_0(t) exp^{\beta_1 Portfolio+ \beta_2 Still+\beta_3 DeltaU+\beta_4 DeltaS+\beta_5 Age+\beta_6 Risk+\beta_7 Job+\beta_8 Gender}$$

We can also get interested in the surival functions, considering the split between full and partial surrenders. Plotting the three curves leads to



- In green is the survival function, including both and partial surrenders
- In blue, the survival function for full surrenders
- In red, the survival function taking into account partial surrenders only, e.g. it represents the survival for the policyholders still in the Camelea portfolio

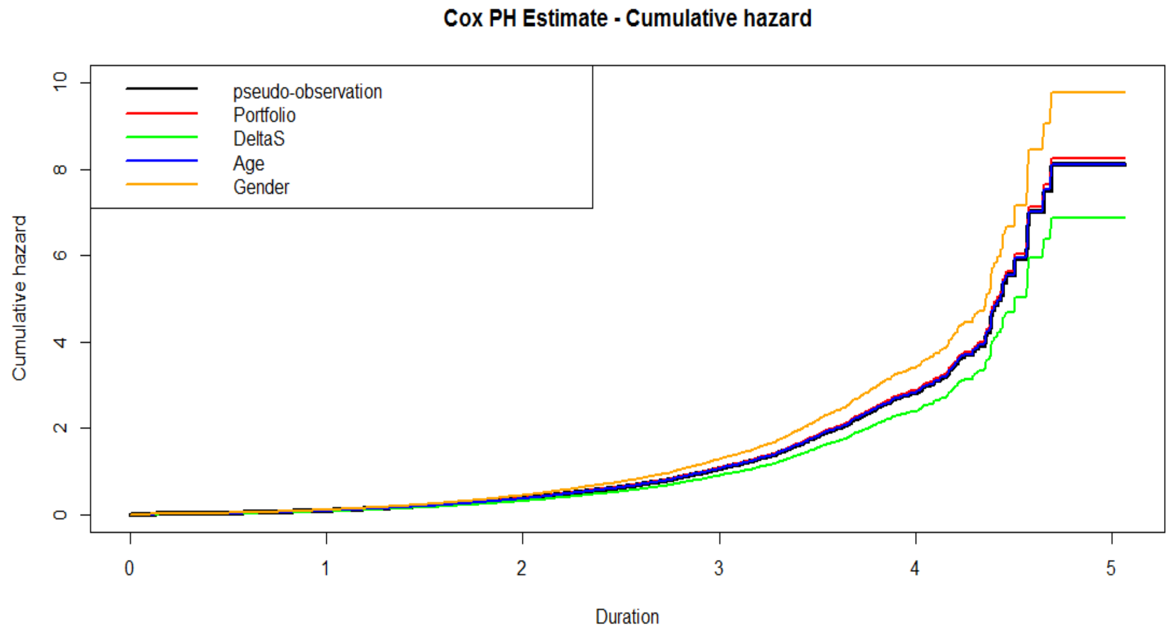
The estimation method used here is the Kaplan-Meier one. 95% confidence intervals is asymptotically normal.

The study of the survival function with R uses the Kaplan-Meier estimate as mentioned above. The plot of the Kaplan-Meier estimate of the survival function (confer graph below) is a step function in which the estimated survival probabilities are constant between adjacent surrender times and only decrease at each surrender.

2.4.2.6. Baseline hazard function

In R, the baseline hazard function gives us, for a policyholder i,

$$\widehat{H}_i(t) = \widehat{H}_0(t) exp^{\beta_1 Portfolio + \beta_2 Still + \beta_3 DeltaU + \beta_4 DeltaS + \beta_5 Age + \beta_6 Risk + \beta_7 Job + \beta_8 Gender}$$



In dark black is the estimate baseline hazard function, based on the observations [24]. The other curves are representing the baseline hazard function conditionally to one covariate.

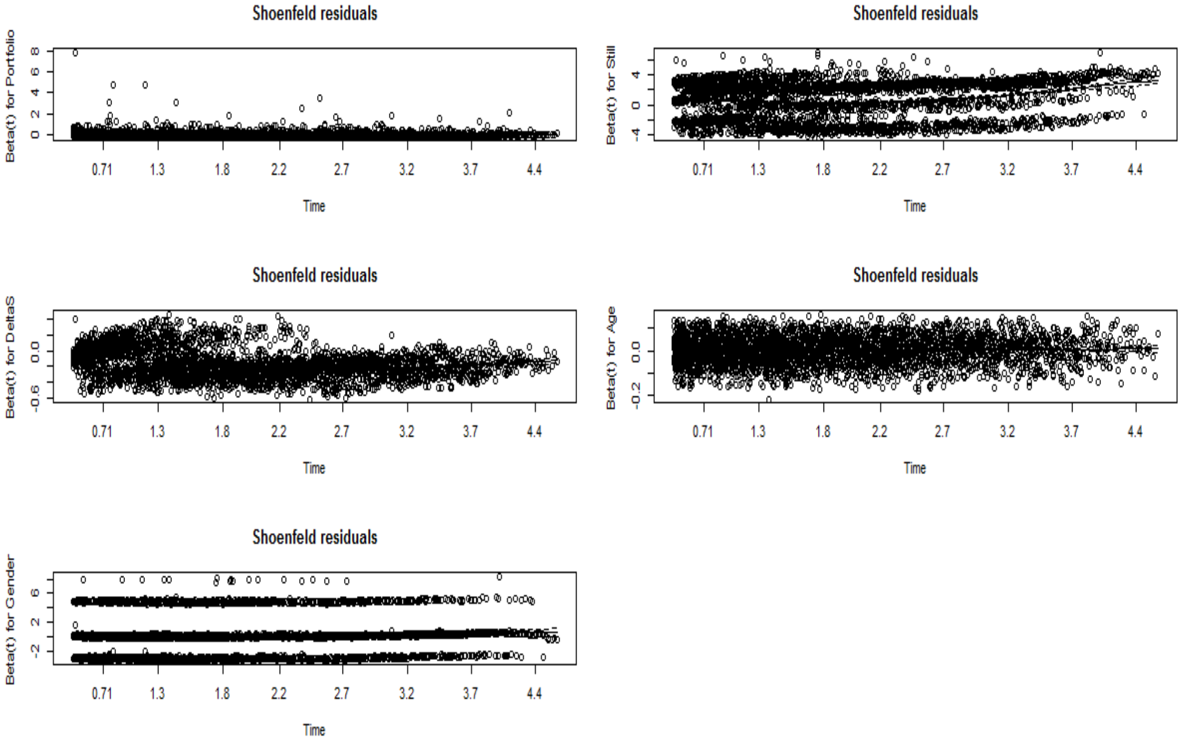
What we do observe here is in line with what we saw earlier.

- Portfolio (red) and Age (blue) do not impact a lot the surrender risk. Indeed, they are very close to the (black) baseline hazard line. The Portolio covariate increases slightly the risk, while the age decreases it
- The market rate covariate, DeltaS, impacts significantly and positively the surrender risk. Rising markets make the surrender risk becoming lower
- The gender covariate has a significant impact on the surrender risk. Women are more likely to surrender than men

2.4.2.7. *Model schoenfeld residuals*

Significant covariates for the Cox PH model have been all selected. What we have to do now is checking that the proportional hazards assumption is valid within the model we just built.

Let's look first the plot of Schoenfeld residuals. Every deviation from the horizontal is an indication of a time-dependent covariate.



We can estimate from these plots that both Still and DeltaS covariates are time –dependent. Indeed, we can observe for the two covariates a rising trend. A statistical Schoenfeld test should go in this way.
In practice, these results are coherent. Life insurance policies are considered as popular investment vehicles, generally tax – exempted, hence their popularity. In this way, the longer policyholders invest in their portfolios, the richer they are. Consequently, they surrender once they saved enough money in order to finance a project ... The time –dependence for market movements is not surprising either, inasmuch as financial market volatility and variations over the past few years.

The statisical proportionality test based on Schoenfled results leads to

R Console

```
> cox.zph(M0)
              rho   chisq        p
Portfolio -0.0332    3.52 6.06e-02
Still      0.1300   67.64 2.22e-16
DeltaS    -0.2696  306.60 0.00e+00
Age        0.0331    3.62 5.73e-02
Gender     0.0365    4.51 3.37e-02
GLOBAL         NA  506.08 0.00e+00
>
```

“rho” is the Pearson correlation coeffcient between Schoenfled residuals and time for each covariate.

The test comparing to zero the regression straight returns p-values varying from 0 to0.31 . the first conclusion to this test is:

- The portfolio covariate appears to be as not time –dependent (p-value p >5%), e.g. the condition of proportional hazards is respected
- Age and Gender covariates is marginally time-dependent, e.g. a p-value $p\sim$ 5%), e.g. they do not respect entirely the proportional hazards assumption
- deltaS and Still covariates are completely time –dependent.

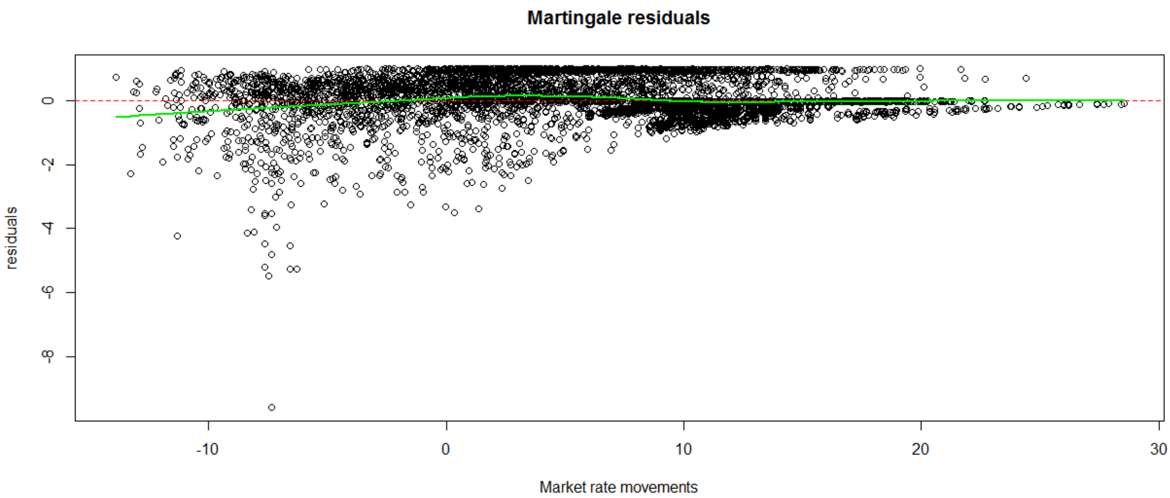
Mainly due to the DeltaS covariate, and marginally with the Portfolio and Still ones, the proportional hazards assumption of our Cox model is not verified. The global validation test of the PH assumption, $H_0: \beta(t) = \beta$ versus $H_1: \beta(t) \neq \beta$, conducts to reject it, because of time – dependent covariates.

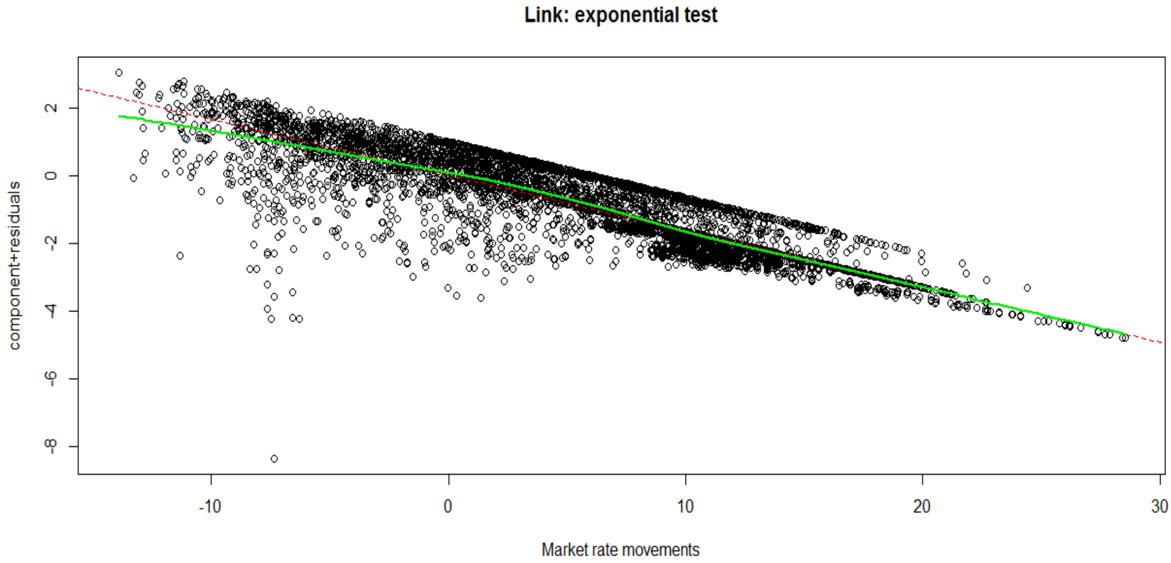
2.4.2.8. Martingale residuals

The Covariate linked to financial market is strongly time-dependent, meaning there is a link between the covariate with time. The martingale residuals will test the link form, allowing us to assess the type of dependence between the covariate and time.

We plot the martingale residuals versus the time-dependent covariates. We smooth the curves with local polynomes with only one degree of freedom to highlight any trend [25].
In case of an exponential link function, the rate logarithm will be a linear function of the covariates. Plotting the estimated link function with the martingale residuals versus the covariates will highlight the form of the link function. A fast growth of the curve will suggest a power transformation, with $p > 1$. On the opposite, a slow growth will suggest a logarithm/ square transformation ($p < 1$).

The plot in the first time of the martingale residuals, and in a second time, of the estimated link function + martingale residuals led to:



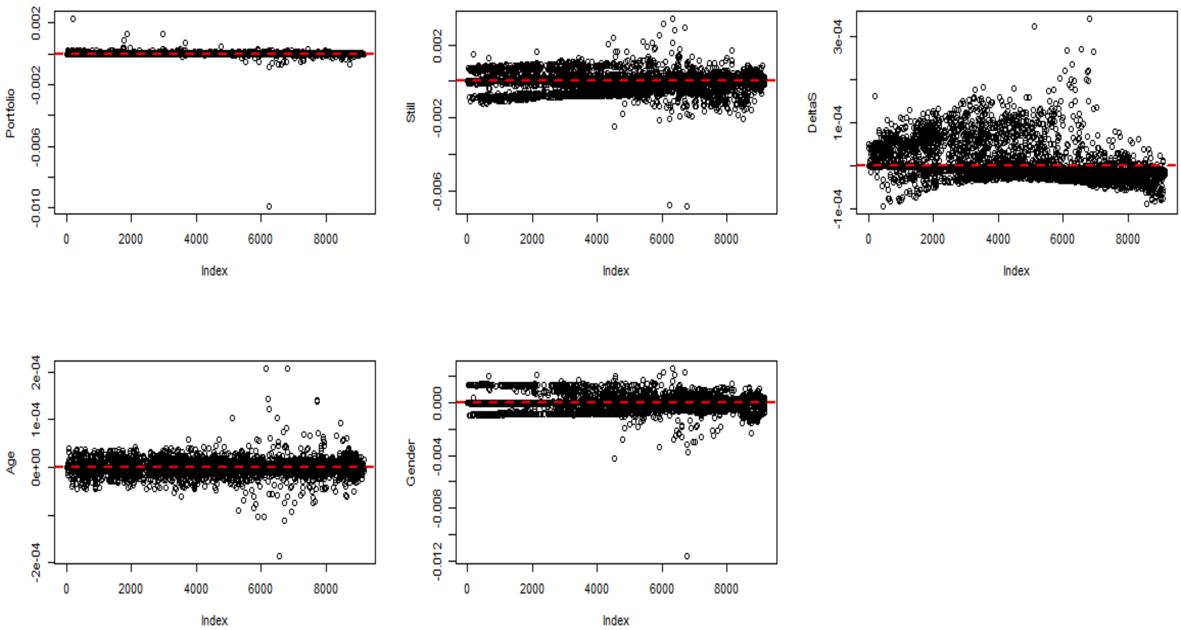


These two plots clearly indicate an exponential link between time and the MSCI european equity rate.

2.4.2.9. Influence of observations

We need to perform one last check : verifying that the model coeffcients are not defined only from a small number of observations but represent the entire population of the data set. Comparing orders of magnitude of the DfBeta residuals with the coefficients will indicate us the significance of the coefficients.

The plot of the DfBetas coefficients leads to



Respectively, coefficients for the Cox model covariates are: 0.01736, 0.6852, -0.1645, 0.00252, 0.18732. Comparing the orders of magnitude of both DfBetas observations with the last Cox model coefficients (confer section 2.4.2.4.) draws to the conclusion:
DfBeta residuals are small comparing to their respective coefficient values. Hence no anormally influent observations on the Cox model.

2.4.2.10. Time dependent covariates – fixing the issue

2.4.2.10.1. New data frame

Time –dependent covariates is not suitable for the Cox PH model. In order to solve the issue, we built, on a monthly basis, a new data set based on what we already presented before, which is presented as follows:

```
R Console
> summary(dat)
     Policy         Duration            start             stop        
 Min.   :   1   Min.   :0.00274   Min.   :0.000   Min.   :0.00274  
 1st Qu.:4567   1st Qu.:1.76438   1st Qu.:0.500   1st Qu.:0.58333  
 Median :6564   Median :2.65753   Median :1.083   Median :1.16667  
 Mean   :6130   Mean   :2.60676   Mean   :1.283   Mean   :1.36444  
 3rd Qu.:8028   3rd Qu.:3.49041   3rd Qu.:1.917   3rd Qu.:2.00000  
 Max.   :9159   Max.   :5.06575   Max.   :5.000   Max.   :5.06575  

     Event              Still              Portfolio            DeltaU      
 Min.   :0.00000   Min.   :0.000000   Min.   : 0.0031   Min.   :-1.30   
 1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.: 0.2315   1st Qu.: 0.00   
 Median :0.00000   Median :0.000000   Median : 0.5447   Median : 0.30   
 Mean   :0.01585   Mean   :0.007886   Mean   : 1.3436   Mean   : 0.33   
 3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.: 1.4301   3rd Qu.: 0.80   
 Max.   :1.00000   Max.   :1.000000   Max.   :94.4390   Max.   : 1.40   
                                                        NA's   :206426  
     DeltaS             Age            Gender      
 Min.   :-1.30    Min.   : 6.00   Min.   :0.0000  
 1st Qu.: 0.00    1st Qu.:47.00   1st Qu.:0.0000  
 Median : 0.30    Median :57.00   Median :0.0000  
 Mean   : 0.33    Mean   :57.05   Mean   :0.2697  
 3rd Qu.: 0.80    3rd Qu.:68.00   3rd Qu.:1.0000  
 Max.   : 1.40    Max.   :96.00   Max.   :1.0000  
 NA's   :206426                                   
> 
```

The surrender data has been separated on a monthly basis for each policyholder. At the surrender time – or at the end of the study, we indicate the value of the DeltaU (unemployment) and DeltaS (market rates).

R Console

```
> dat[200:205,1:11]
    Policy   Duration      start       stop Event Still Portfolio DeltaU DeltaS Age Gender
200    193 0.08767123 0.08333333 0.08767123     0     0 1.8963571    0.0    0.0  71      0
201    194 0.09041096 0.00000000 0.08333333     0     0 2.0133861     NA     NA  37      0
202    194 0.09041096 0.08333333 0.09041096     1     1 2.0133861    0.1    0.1  37      0
203    195 0.09041096 0.00000000 0.08333333     0     0 0.1001948     NA     NA  62      0
204    195 0.09041096 0.08333333 0.09041096     0     0 0.1001948    0.0    0.0  62      0
205    196 0.09041096 0.00000000 0.08333333     0     0 0.1224317     NA     NA  73      0
> 
```

We can see in this case, that the policy n° 194, has a total duration of 0.090 year within the Camelea portfolio (e.g. 33 days), which was fully surrendered (Event = 1), has a portfolio value equal to €201,338.61. The policyholder was at the surrender time, 37 years old, male. When he surrendered, unemployment and financial markets both increased of 0.1%.

2.4.2.10.2. Fitting Cox model

We can now fit the Cox model. Commands and outputs are

R Console

```
> M0
Call:
coxph(formula = Surv(start, stop, Event) ~ Portfolio + DeltaS +
    Age + Gender, data = dat)


             coef exp(coef) se(coef)      z       p
Portfolio  0.01428     1.014  0.00449   3.18 1.5e-03
DeltaS    -0.71797     0.488  0.02561 -28.04 0.0e+00
Age        0.00847     1.009  0.00120   7.09 1.4e-12
Gender    -0.31406     0.730  0.04307  -7.29 3.1e-13

Likelihood ratio test=996  on 4 df, p=0  n= 9159, number of events= 3417
   (206426 observations deleted due to missingness)
> 
```

```
R Console

              coef exp(coef)  se(coef)       z Pr(>|z|)
Portfolio  0.014281  1.014384  0.004489   3.182  0.00146 **
DeltaS    -0.717968  0.487742  0.025607 -28.038  < 2e-16 ***
Age        0.008469  1.008505  0.001195   7.086 1.38e-12 ***
Gender    -0.314063  0.730473  0.043071  -7.292 3.06e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

          exp(coef) exp(-coef) lower .95 upper .95
Portfolio    1.0144     0.9858    1.0055    1.0233
DeltaS       0.4877     2.0503    0.4639    0.5128
Age          1.0085     0.9916    1.0061    1.0109
Gender       0.7305     1.3690    0.6713    0.7948

Concordance= 0.633  (se = 0.006 )
Rsquare= 0.103   (max possible= 0.959 )
Likelihood ratio test= 996.1  on 4 df,   p=0
Wald test            = 992.9  on 4 df,   p=0
Score (logrank) test = 1093  on 4 df,   p=0

>
```

As the previous model, all the covariates are remaining significant in this model with this new data (p −value results all below 5%).

Before interpreting the results, let's validate both Schoenfeld and DfBeta residuals.

2.4.2.10.3. New Cox model residuals

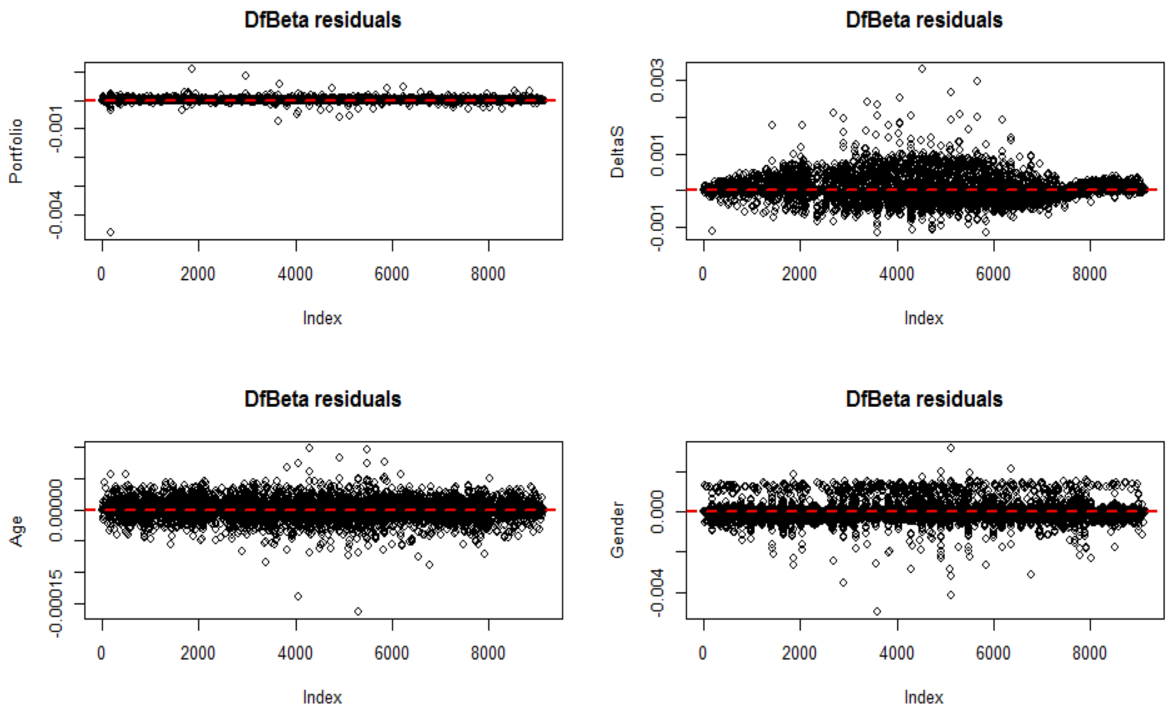
2.4.2.10.3.1. Schoenfeld residuals

```
R Console

> cox.zph(M0)
              rho   chisq     p
Portfolio  0.20418 112.347 0.000
DeltaS     0.00554   0.117 0.732
Age       -0.02247   1.845 0.174
Gender     0.01309   0.596 0.440
GLOBAL          NA 114.687 0.000
>
```

We observe now that, by splitting our data in monthly intervals, the Portfolio covariate became time – dependent. This observation makes sense. Policyholders investing in a life insurance policy can add some extra money to their first investment all over their policy lifetime (what is usually called in the life insurance business, a top-up). As much as a policyholder is allowed to surrender his policy, he's also allowed to reinvest in it. It is consequently normal to see a policyholder reinvests in his portfolio all over the policy lifetime. The longest the policyholder invests, the richer he becomes.

2.4.2.10.3.2. DfBetas residuals

Plotting the DfBetas residuals leads to



Respectively, Cox regression coefficients for this new CoxPH model are: 0.01428, -0.71797, 0.008477, -0.31406.

All DfBeta's are gathered around 0 homogeneously, with very low level of variations (around$10^{-3}, 10^{-4}$, e.g. much more lower than the fitted Cox regression model coefficients). We can then deduce than there are no abnormally influent observations on the model.

Even though the DfBeta residuals led to conclusive result, the Schoenfeld test failed. Considering the data has already been processed in order to solve time – dependence issues on the market rate covariate mainly, we must look towards another solution.

2.4.2.11. Fitting the Portfolio covariate

Looking on the portfolio data, we realize that, for each policyholder, there is a specific amount of savings, e.g. 9,159 different figures. We decide to recode the portfolio covariate into several classes in order to have a more significant model covariate, more representative of the level of wealth of each policyholder, as of the time of surrender or end of December 2013. The portfolio covariate is so recoded by quartile, as such

Portfolio	From	To	Observations
1	0,000	0,230	From 0 to the 1st quartile
2	0,230	0,559	From the 1st to the 2nd quartile
3	0,559	1,472	From the 2nd to the 3rd Quartile
4	1,472	94,439	From the 3rd quartile to the maximum value

Splitting the policyholders into four classes of wealth allows seeing how policyholders react in function of the level of wealth they are belonging to.

Once re-treated, the data is presented as follows:

R Console

```
> dat[200:205,1:11]
    Policy   Duration      start       stop Event Still Portfolio DeltaU DeltaS Age Gender
200    193 0.08767123 0.08333333 0.08767123     0     0         4    0.0    0.0  71      0
201    194 0.09041096 0.00000000 0.08333333     0     0         4     NA     NA  37      0
202    194 0.09041096 0.08333333 0.09041096     1     1         4    0.1    0.1  37      0
203    195 0.09041096 0.00000000 0.08333333     0     0         1     NA     NA  62      0
204    195 0.09041096 0.08333333 0.09041096     0     0         1    0.0    0.0  62      0
205    196 0.09041096 0.00000000 0.08333333     0     0         1     NA     NA  73      0
> |
```

The interpretation remains the same as the one in paragraph 2.4.2.10.1.The policyholder n°194 belongs to the highest portfolio class, 4, while the policyholder 195 belongs to the lowest one, 1.

2.4.2.11.1. Cox model validation

The new Cox model returns

R Console

```
> M0 <- coxph(Surv(start,stop, Event)~Portfolio+DeltaS+Age+Gender,data=dat)
> M0
Call:
coxph(formula = Surv(start, stop, Event) ~ Portfolio + DeltaS +
    Age + Gender, data = dat)


              coef exp(coef) se(coef)      z       p
Portfolio -0.07797     0.925   0.0157  -4.98 6.5e-07
DeltaS    -0.71884     0.487   0.0256 -28.09 0.0e+00
Age        0.00895     1.009   0.0012   7.47 8.3e-14
Gender    -0.31264     0.732   0.0431  -7.26 3.9e-13

Likelihood ratio test=1012  on 4 df, p=0  n= 9159, number of events= 3417
   (206426 observations deleted due to missingness)
> |
```

The Wald test on the covariates indicates that all of them are significant for the model. More detailed, the new Cox model has as characteristics

```
R Console

                coef exp(coef)  se(coef)       z Pr(>|z|)
Portfolio -0.077967  0.924995  0.015668  -4.976 6.48e-07 ***
DeltaS    -0.718836  0.487319  0.025589 -28.092  < 2e-16 ***
Age        0.008948  1.008988  0.001199   7.466 8.28e-14 ***
Gender    -0.312638  0.731515  0.043068  -7.259 3.89e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

          exp(coef) exp(-coef) lower .95 upper .95
Portfolio    0.9250     1.0811    0.8970    0.9538
DeltaS       0.4873     2.0520    0.4635    0.5124
Age          1.0090     0.9911    1.0066    1.0114
Gender       0.7315     1.3670    0.6723    0.7959

Concordance= 0.637  (se = 0.006 )
Rsquare= 0.105   (max possible= 0.959 )
Likelihood ratio test= 1012  on 4 df,   p=0
Wald test            = 1008  on 4 df,   p=0
Score (logrank) test = 1107  on 4 df,   p=0

> |
```

2.4.2.11.2. Model residuals

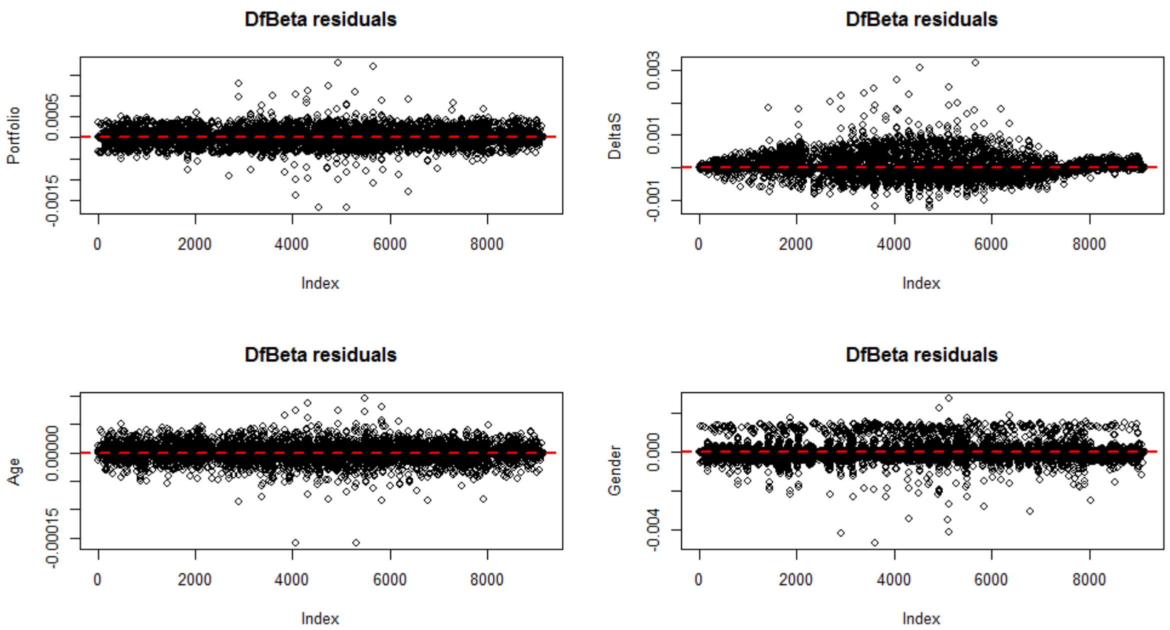
2.4.2.11.2.1. Schoenfeld residuals

```
R Console

> M0 <- coxph(Surv(start,stop, Event)~Portfolio+DeltaS+Age+Gender,data=dat)
> cox.zph(M0)
               rho  chisq     p
Portfolio  0.01274 0.6035 0.437
DeltaS    -0.00204 0.0158 0.900
Age       -0.01537 0.8604 0.354
Gender     0.01515 0.7985 0.372
GLOBAL          NA 2.1655 0.705
> |
```

All covariates are time –independent. The proportional hazards assumption is consequently validated, thanks to the separation on four different classes of wealth of the portfolio covariate.

2.4.2.11.2.2. DfBetas residuals

The study of the DfBeta residuals leads to



Respectively, Cox PH regression coefficients are equal to: -0.07979, -0.71884, 0.00895, -0.31264.Residuals all very close and centre homogeneously on 0. We can deduct no abnormally influent observations.

2.4.3. Final Cox Proportional Hazards Model

We managed to fit a Cox proportional hazards model on our surrender rate. After facing the violation of the main model assumption, the proportional hazards one, we succeeded fitting this model and respecting its assumptions.

The Wald test on all the covariates showed which covariates were significant in the first place, which allowed us to select these very ones. Tests on residuals (Schoenfeld, DfBetas and martingale) where definitely helpful to validate the Cox model assumptions

With the last data set, the final Cox model is now written as follows:

$$h(duration) = h_0(duration)\exp(\beta_1 Portfolio + \beta_2 DeltaS + \beta_3 Age + \beta_4 Gender)$$

In R, commands and outputs are

```
R Console
> M0 <- coxph(Surv(start,stop, Event)~Portfolio+DeltaS+Age+Gender,data=dat)
> M0
Call:
coxph(formula = Surv(start, stop, Event) ~ Portfolio + DeltaS +
    Age + Gender, data = dat)


              coef exp(coef) se(coef)      z       p
Portfolio -0.07797     0.925   0.0157  -4.98 6.5e-07
DeltaS    -0.71884     0.487   0.0256 -28.09 0.0e+00
Age        0.00895     1.009   0.0012   7.47 8.3e-14
Gender    -0.31264     0.732   0.0431  -7.26 3.9e-13

Likelihood ratio test=1012  on 4 df, p=0  n= 9159, number of events= 3417
```

Studying all covariates on a monthly basis allow us to get interested on the impact of each of them every month, until the policyholder decision to surrender.

Calibrating a Cox PH model on our data finally highlights the following policyholder behaviour:

The richer a policyholder is, the less he surrenders. The wealth of the policyholder affects his will to surrender. All covariates remaining constant, stepping from level 1 of savings to level 2 in a month reduces the annual surrender rate by$(100-92{,}5)\ \% = \ 7{,}5\%$.

The more financial markets grow, the less the policyholder surrenders. This result makes sense. As we saw earlier in section 1.3.1, the Camelea portfolio offers a large range of investment solutions and should as such reflect financial market movements – up to a factor. The more the market grows, the more profitable it is to invest within. All covariates remaining constant, the increase of 1% of the MSCI European equity market reduces the annual surrender rate by$(100-48{,}7)\ \% = \ 51{,}3\%$.

The older a policyholder is, the more he surrenders. Indeed, getting one more years old makes the policyholder surrendering $(100{,}9-100)\ \% = \ 0{,}9\%$ faster.

Men surrender $(100-73{,}2)\ \% = \ 26{,}8\%$ more than women – however, they are over represented in the Camelea portfolio. More than 72% of Camelea policyholders are men.

2.4.4. Parametric model

2.4.4.1. Parametric model selection

We are looking to fit a parametric model on the original data set. Seven models are tested:

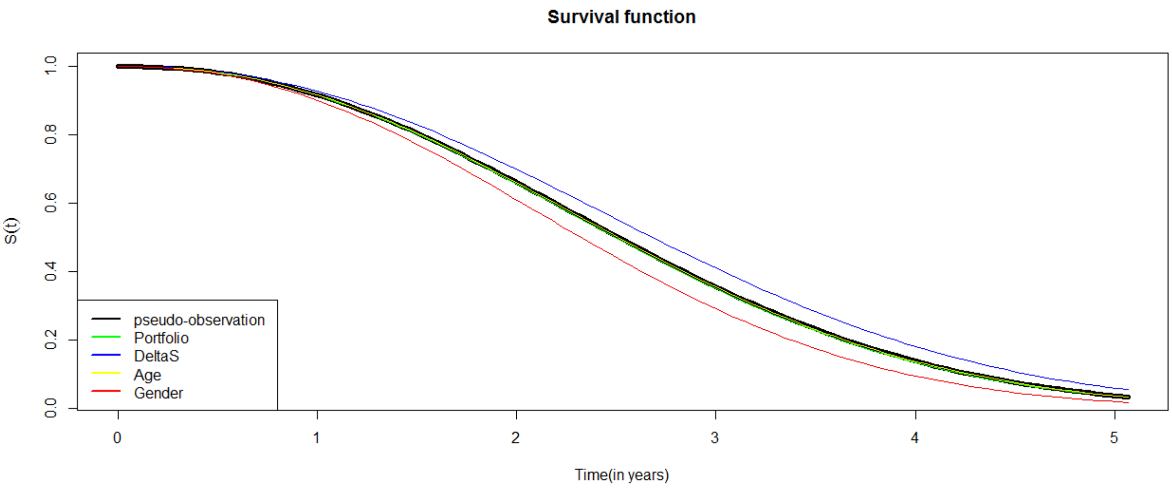
- Log – normal
- Weibull
- Log – logistic
- Gaussian
- Exponential
- t
- Extreme

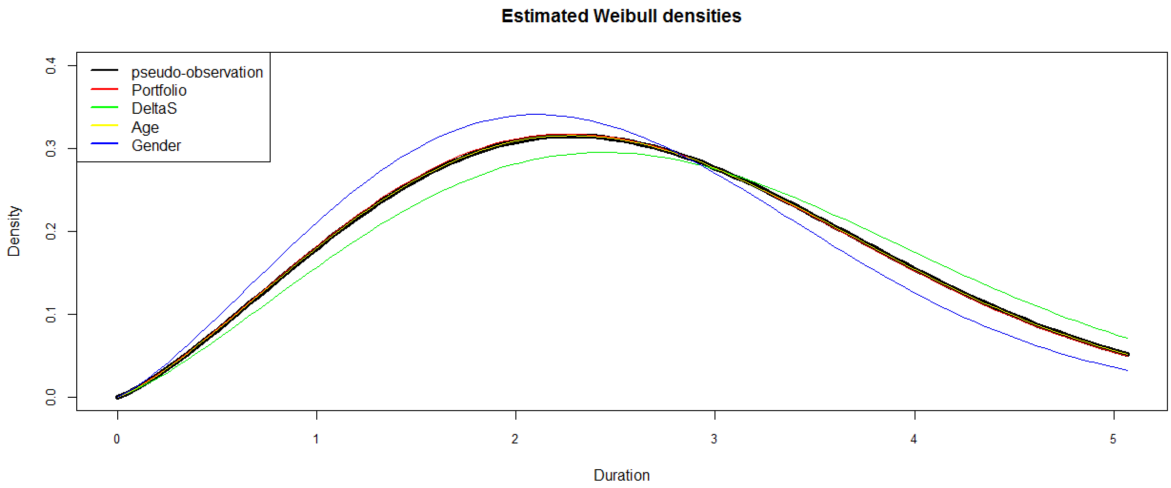
They are selected with the AIC criteria. The R command returns

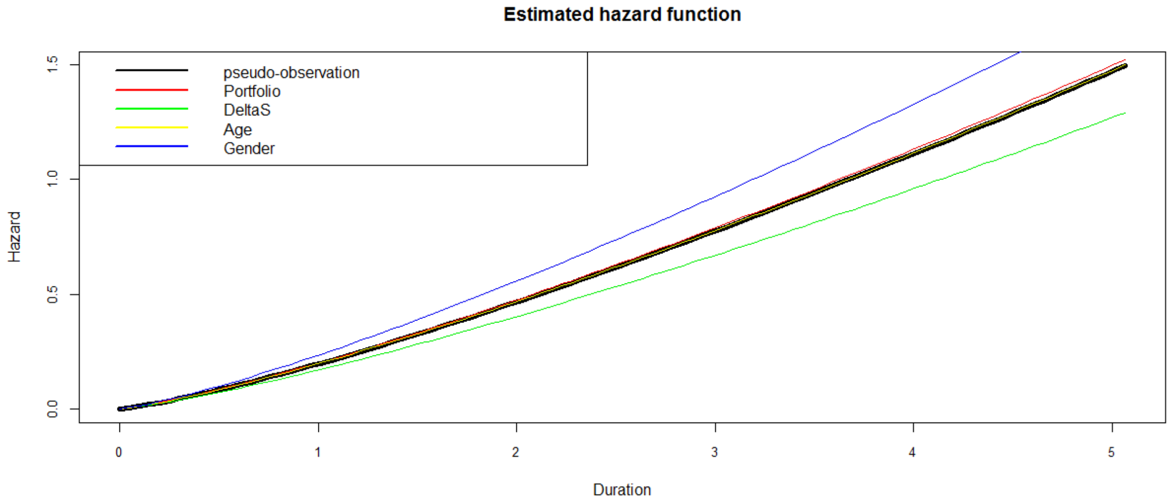
```
R Console
> anova(AFTO_1,AFTO_2,AFTO_3,AFTO_4,AFTO_5, AFTO_6, AFTO_7)
                                  Terms Resid. Df    -2*LL Test Df    Deviance      Pr(>Chi)
1 Portfolio + Still + DeltaS + Age + Gender      9152 13526.44        NA          NA            NA
2 Portfolio + Still + DeltaS + Age + Gender      9152 12007.34     =  0  1519.09498            NA
3 Portfolio + Still + DeltaS + Age + Gender      9152 12897.92     =  0  -890.58139            NA
4 Portfolio + Still + DeltaS + Age + Gender      9152 12929.85     =  0   -31.92676            NA
5 Portfolio + Still + DeltaS + Age + Gender      9153 14409.29     = -1 -1479.44515 9.881313e-324
6 Portfolio + Still + DeltaS + Age + Gender      9152 12895.72     =  1  1513.57574  0.000000e+00
7 Portfolio + Still + DeltaS + Age + Gender      9152 12322.45     =  0   573.26568            NA
```

From this test, we can deduce that, based on the AIC criteria, the Weibull distribution is the one which fits best the survival time [26].

2.4.4.2. Interest functions



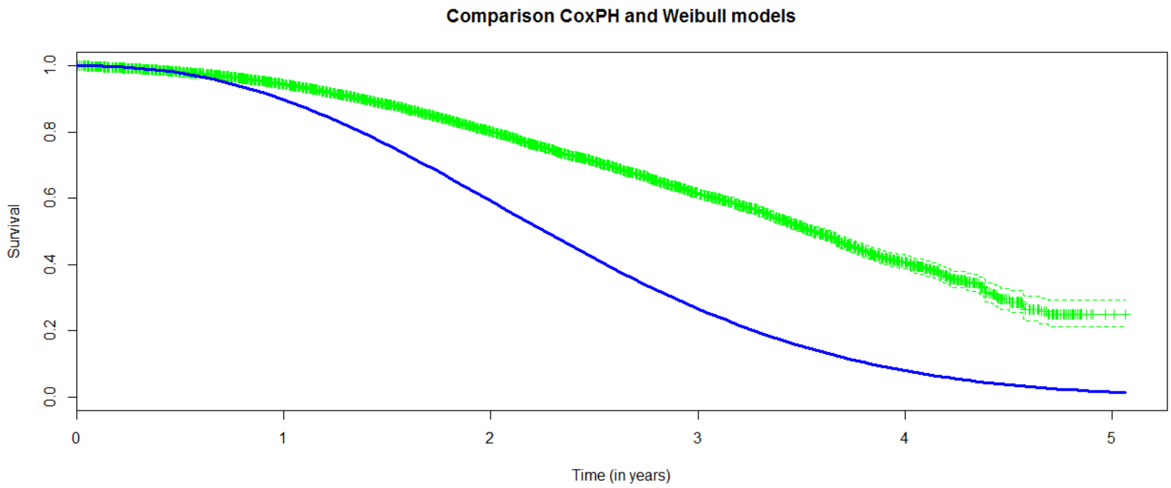




These three plots all conduct to the same conclusions than the Cox PH model. The portfolio and the age do not impact a lot the decision to surrender. Increasing financial markets impact positively and reduce the surrender risk significantly. On the other hand, being a women has a large impact on the surrender rate.

2.4.4.3. Model comparisons

The plot of both (Kaplan- Meier) Cox and Weibull models shows clearly that the Weibull model is making too much assumptions on the survival times distributions. Indeed,



Proportionally, the parametric survival function is making too many assumptions on the survival times distributions and decreases toofast to reflect the portfolio.
The Cox PH model seems so, in our case, much more appropriate.

2.4.5. Conclusion

The Cox PH model has the disadvantage of having its distribution of survival times unknown. Indeed, any assumption is made on the baseline hazard function $h_0(t)$, which makes this model non – parametric. Another disadvantage which might be cited here is that its survival function is less consistent with a theoretical one. The Cox PH survival function is typically a step function, got with the plot of the Kaplan – Meier estimate.

However, two essential advantages for the semi – parametrical model. The main one is that the model does not rely on distribution assumptions. The second is that the baseline hazard is not necessary for ratio use, hence no assumption of the distribution of survival times.

Even though a parametric model would have completely specify the hazard function $h(t)$ and the survival function $S(t)$ – without mentionning the fact of being more consitent with the theoretical function, the major issue here would have been making an assumption of the underlying distribution. This would have been problematic, inasmuch as a wrong fit of the survival times distribution would have made the final parametric model completely wrong.

The Cox PH model returns some interesting results, illustrating the covariates which will increase or decrease the policyholder motivation to surrender. As instance; a female policyholder in portolio class 2 of 70 has less chances to surrender than a male in portfolio class 3 of 60.

As we saw earlier, calibrating a semi – parametrical survival model highlighted the global behaviour of the policyholders within the Camelea portfolio. The next section will look deeper in these bahaviours, in order to define a sales and marketing strategy for lowering the surrender risk, by chosing the policyholders in function of several criterias.

3. Surrender factors

3.1. Influence of the Job occupation

All policyholders of the data sample are classified by job occupation. The interest is here to study the reaction of one class compared to another based on the global final Cox model we build.

Results are summarized below

MODEL	N	P	covariates	coef	exp(coef)	p	rho	p
GLOBAL	9159	3417	Portfolio	-0.07797	0.925	0	0.013	0.44
			DeltaS	-0.71884	0.487	0	-0.002	0.90
			Age	0.00895	1.009	0	-0.015	0.35
			Gender	-0.31264	0.732	0	0.015	0.37
UNEMPLOYED	524	229	DeltaS	-0.557	0.573	0	0.008	0.91
STUDENT	209	56	Gender	-1.184	0.306	0.0055	0.073	0.56
			DeltaS	-0.839	0.432	0.0027	0.007	0.95
STATE EMPLOYEE (PUBLIC SECTOR)	620	225	Portfolio	-0.139	0.870	0.035	0.016	0.80
			DeltaS	-1.12	0.326	0	-0.149	0.08
SELF-EMPLOYED/SHOPKEEPER (LEGAL ENTITY)	649	202	Portfolio	-0.102	0.903	0.01	0.123	0.08
SELF-EMPLOYED/SHOPKEEPER (INDIVIDUAL)	922	365	Portfolio	-0.105	0.900	0.03	0.051	0.32
			Gender	-0.307	0.736	0.029	-0.026	0.62
RETIRED	2806	1200	Delta S	-0.6122	0.542	0	0.005	0.87
			Age	0.0092	1.009	0.0006	0.015	0.59
			Gender	-0.2897	0.748	0	-0.001	0.97
PRIVATE/INDEPENDENT PRACTICE	616	187	DeltaS	-0.841	0.431	0	0.001	0.99
OTHER	314	100	DeltaS	-0.8779	0.416	0	0.030	0.76
			Age	0.0206	1.021	0.021	-0.105	0.24
EXECUTIVE	309	93	DeltaS	-0.9866	0.373	0	-0.019	0.84
			Age	0.0203	1.021	0.049	0.039	0.73
EMPLOYEE (PRIVATE SECTOR)	1968	670	Portfolio	-0.1426	0.867	0	0.043	0.25
			DeltaS	-0.7472	0.474	0	-0.014	0.71
			Age	0.0108	1.011	0.00073	0.033	0.40
			Gender	-0.3705	0.690	0	0.061	0.11
COMPANY DIRECTOR	222	90	DeltaS	-0.527	0.590	0	0.000	0.89

- From the global model, we can deduce:

- An increase from 1 class to another reduces the surrender risk by 7.5%
- An increase of 1% of financial markets has a large impact on the surrender risk, decreasing it by 52.3%.
- An increase of 1 year old on a policyholder increases by 0.9% the chances to surrender
- Being a women impacts positively the surrender rate by decreasing it of 26.8%.

- Unemployed persons have a large surrender rate. Indeed, 229 policyholders on a total of 524, e.g. 44% already surrendered at least once, fully or partially their life insurance policy. Their decision to surrender is mainly driven, in our model, the evolution of financial markets. However, the need of liquidities is hard to model in this case. Indeed, in practise, unemployed people are surrendering their life insurance policies in order to face unemployment and survive.
  It could be intresting in this very case to include the unemployment data to see the its impact on the surrender rate. Indeed, even though the unemployment rate was not significant on our global model basis, it would make sense it has a more significant impact here.

- 27% of students already surrendered once their life insurance policy. If they are on average as reactive to surrender or not as other policyholders within the global model in case of financial markets movements, men are surrendering a lot more than women. Indeed, women woud surrender until 70.4% less than men.

- State employees from the public sector are reactive to both financial markets movements and their level of wealth. The richer they are, the less they surrender (-23% of chances to surrender from one class of wealth to another)

- Self – employed and shopkeeper, legal entity or individual, are both sensitive in the same way to their level of wealth. The richer they are, the less the surrender (-10% of chances to surrender). For idividuals, the gender also has an impact. In this category, men are more likely to surrender by 27.4%

- Retired policyholders are the most numerous within Camelea. They are sensitive to the market rates, the age and the gender. The older the policyholder is, the more he surrenders (+0.9% for one year older). As usual in the study, men surrender 25.2% more than women.

- "Other" and excutive job occupations are very sensitive to the market risk. An increase of 1% of financial markets will reduce their will to surrender by respectively 59.4% and 67.3%. The age also drives their decision to surrender. Getting one year older increases their decision to surrender by 2.1%.

- Employees from the private sector are senitive to the four covariates of the global model. The richer they are the less the surrender (-13.3%). An increase of 1% of financial market reduces their will to surrender by 52.6%, while getting one year older increases their chances to surrender by 1.1%

- The decision to surrender for company directors is driven by the evolution of financial markets. An increase of 1% of those will reduce the surrrender risk by 41%.

Whatever the job occupation, most of policyholders manage their policies and their decision to surrender based on the financial markets movements. Increasing markets will make them confident and reinvest within their policies while decreasing markets will frighten them and make them surrender sooner or later.
Job occupations “executive”, “student”, “other”, “state employee (public sector)”, “employee (private sector)” are the most reactive towards this risk.

Gender has the biggest impact for “student” policyholders, where men surrender a lot more than women. It is the same conclusion, but in lower proportions, for retired policyholders and employees from the private sector.

Age has essentially an impact on the willingness to surrender on retired policyholders and employees from the private sector.

The level of wealth of the portfolio has an impact on the decision to surrender for shop-keeper, legal and individual, and employees from the private sector.

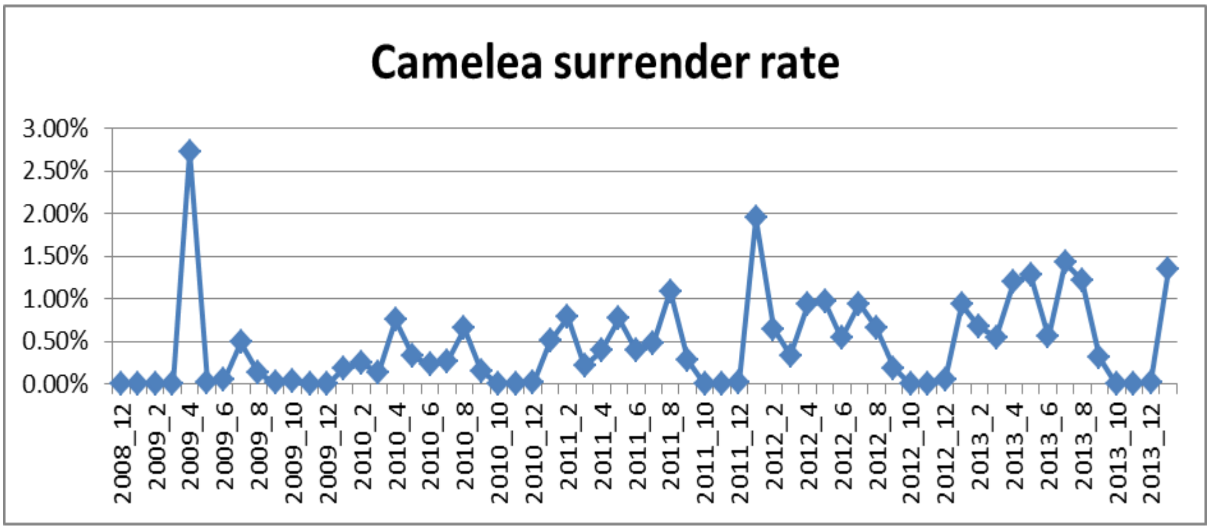
This highlights that the life insurer has to manage carefully the interest rate he is offering to its clients. Indeed, market yields impact a lot the policyholder’s decision to surrender or not his policy.
Conversely, it might worth to target wealthy policyholders, the study showing that the wealthier they are, the less they surrender. Age does not impact significantly, but marginally the surrender risk. It makes sense that the older the policyholder gets, the more he retired – especially after 60, the average age for retirement. We have the same conclusion than the one in 2.4.5. regarding the gender: Female policyholders surrender less than male policyholders.

3.2. Influence of financial markets

3.2.1. Camelea historical lapse rate

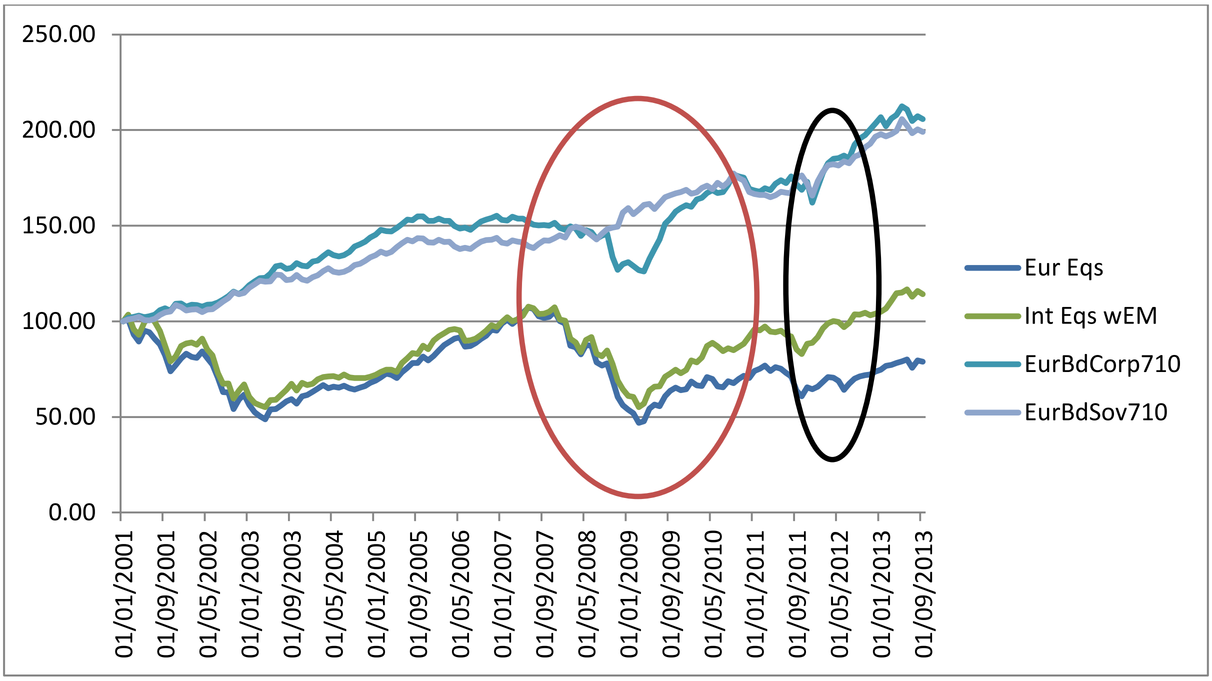
3.2.1.1. On a monthly basis

The Camelea lapse rate since December 2008 evolved as follows:



We can notice on this chart two significant high points, one in early 2009, and another one early 2012. Besides, we can also notice that, starting 2010, the monthly lapse rate osciallate around an average value of 0.50%.

Let's have a look now on the market returns at these specific periods of time: We chose here four MSCI indexes to represent the market (MSCI Equity Europe; MSCI Equity international with Emerging Markets; MSCI Corporate bonds 7-10 years; MSCI Sovereign bonds 7-10 years). In market values:



We can notice on the chart two main results

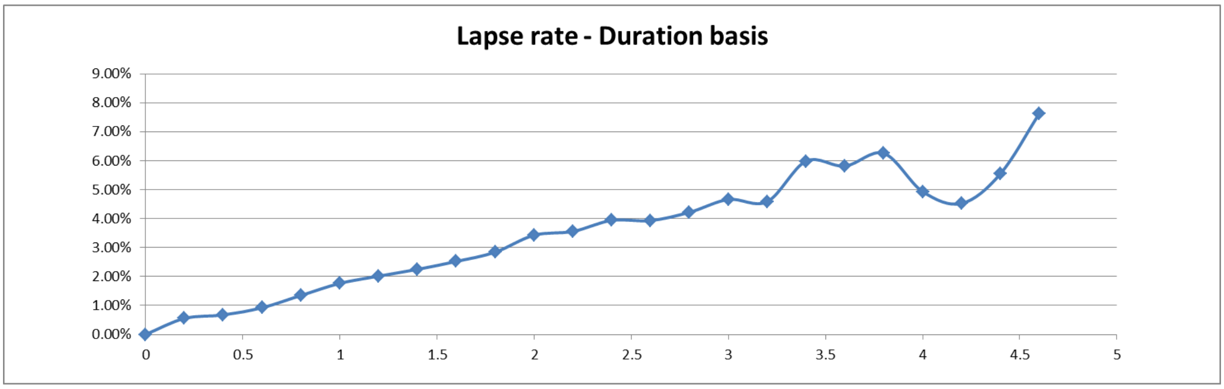
- The market drop resulting of both subprime (circled in red) and debt crisis (circled in black).
- The rising trend of bonds' investments

This additonal observation highlights what as been previsously shown while fitting the Cox PH model: Financial markets have a large influence on the lapse rate, and its evolution over time is one of the main driver of the surrender risk. In case of rising markets, policyholders are more motivated to remain investing. Conversely, falling markets make them frightened about the future and more prompt to surrender in order to protect their savings.

Modelling this risk with classical method, as time series or Cox-Ingersol Ross ones, seems not appropriate here. Indeed, we only have a data set on a monthly basis, which makes the number of observations to maximum 62 points of observation. There is not enough data to properly calibrate one of these models on such a few points[27] [28], without speaking of having a projection distored and biaised. Besides, the subprime crisis and the debt one completely changed the investor's behavior: the policyholder became much risk-averse, which makes him surrendering his policy faster than 2008, in case of a market rate decline higher than usual [29].

3.2.1.2. On a duration basis

The plot of theCamelea lapse rate on a duration basis leads to:



We observe a linear trend, which is interesting. Policyholders surrender in proportion to the time they remain within the Camelea investment product. After four years, the drop is mainly due to the few numbers of observations we have at our disposal. Indeed, between a 4 and 5 – years duration, there are only a few contracts within the Camelea portfolio, just launched on the life insurance market.
This observation is acceptable. Let's have two policyholders having a life insurance contracts for two purposes and investing in the same proportions each month. The first wants to buy a car, the second a house. Respectively, the first policyholder will have enough money to buy his car than the second, who still need to wait to buy his house.

Based on these two observations, the idea is now to build a surrender model which will depends on the market rates, and which will return an average lapse rate on a monthly basis for the next 100 months.

3.3. Modeling a dynamic lapse rate

We would like to model two portfolios, one offering a guatanteed – rate to its policyholders (confer section 1.5.2.3), and one unit – linked, e.g. with a non guatanteed rate – as the Camelea portfolio.

In order to model the conjectural lapse rate – e.g, as a reminder, the one depending of the market rates, we use the recommendations from August 2010, relatively to the QIS 5 of Solvency II [6]. This methodology is also precognised by the French supervisor ("l'ACP, Autorité de Contrôle Prudentiel").

3.3.1. Model assumptions

According to the preliminary studies, the model we have to build should:

- Return a lapse rate function depending on the dynamic lapse rate. This dynamic lapse rate is based the difference the market rate and the interest one served within the portoflio
- Integrate a dynamic lapse rate in the contract value, in order to have a dynamic value of the portfolio actualized with the lapse rates.

Based on historical evolution, we previously noticed on the lapse rates a returning-to-average effect (around 0.50%). We also mention in introduction the problem of modeling the investors' irrationality (more represented in the structural part), which a stochastic modeling allows to model, thanks to the hazard function.

Thus, all these reasons induce us to look once again on a stochastic model.

We model in this section the lapse rate with the Vasicek model: the non-zero probability to have negative lapse rate has here the main advantage to eventually, on a further study, model top-ups. The Vasicek model being easy to manipulate thanks to a Gaussian distribution and a short simulation time, we will also use it to model the market rate.

Under the Vasicek model [30] -proposed in 1977- the only factor is the lapse rate $r_t$modeled under the shape of a process of Ornstein-Uhlenbeck. Under the filtration(F_t), the probability P and the risk-neutral probability$Q$, follows:

$$dr_t = a(b - r_t)dt + \sigma dW_t$$

$$r_0 = r(0)$$

With

- r_t: lapse rate value at the dealing time t
- b: lapse rate long-term average
- a: returning speed average; the rate variation between t and$t$+Δt
- W_t: Q- brownian motion

Eventually, the Vasicek expression can be simplified as:

$$dr(t) = \mu dt + \sigma dW_t$$

In case of discrete distribution, we will have instead:

$$r(t+\Delta t) \approx r(t) + \Delta r(t+\Delta t) \approx r(t) + a\big(b - r(t)\big)\Delta t + \sigma \Delta W_t$$

A demonstration of the results which will follow can be found in the bibliography references [4] & [5] and will not be developed in this paper.

For $0 < s < t$, the expression of the lapse rate is:

$$r_t = r_s e^{-a(t-s)} + b(1 - e^{-a(t-s)}) + \sigma \int_s^t e^{-a(t-u)} dW_u$$

The lapse rates trajectories oscillate between a long-term average, with a volatility equaled to $\frac{\sigma^2}{2a}$. To calibrate a model, natural estimators appear:

$$\hat{b} = average\ estimator\ of\ the\ historical\ lapse\ rate$$

Besides, if we write the lapse rate as:

$$r_{t+\Delta t} - r_t = a(b - r_t)\Delta t + \ \sigma \aleph(0, \Delta t)$$

We can have an estimation of the returning speed average (e.g. the rate variation between $[t; t+\Delta t]$), the empirical volatility and the average as, with r the lapse rate:

$$\hat{a} = \frac{E(r_{t+\Delta t} - r_{t)}}{\Delta t(\hat{b} - r_t)}$$

$$\hat{\sigma} = \sqrt{2\hat{a}Var[r]}$$

$$\hat{b} = E[r]$$

The dynamic lapse rate prevents here the rates from getting too high, due to a maximal and minimum value set in the model.

We get, based on the sample of Camelea historical lapses.

- $\hat{b} = 2.13\%$
- $\hat{a}$ =18.96%
- $Var[r]$ =0.012% (average squared deviation from the mean)
- $\hat{\sigma} = 0.68\%$

3.3.2. Creation of a dynamic surrender rate

In this scenario, we assume that the lapse rate only depends of the difference of the return rate offered by the market and the one offered by the NPGWM [36].The dynamic lapse rate reflects:

- The will of the policyholder to surrender his policy to reinvest in more profitable investment solutions because of increasing financial markets – for a guatanteed rate portfolio only; a non guatanteed rate portfolio, up to factor or a lag, replicates the financial market
- The will of the policyholder to reinvest (top –ups) in his policy because of rising markets – however, the policyholder does not surrender to reinvest his policy somewhere else because of a thin difference in offered interest rates
- The will of the policyholder to surrender because of decreasing markets rates, in order to protect his savings.

Considering α,γ, θ, φdata to determine and defined as follows, the goal is to create a function which will return [31]:

- If $\Delta = Benchmark\ market\ rate - NPGWM\ Served\ rate < \gamma$, with γ to determine, e.g. if the rate offered by one of NPGWM products is higher than the one proposed on the market (calculated with the Vasicek calculation), plus a certain quantity γ, then the policyholder is reinvesting in his policy. This is how we will model top-ups.

- If $\Delta = Benchmark\ market\ rate - NPGWM\ Served\ rate > \theta$, with θ to determine , e.g if the rate offered by one of NPGWM products is lower than the one proposed on the financial markets plus a certain quantity θ, then the policyholder will surrender –partially or fully- his policy. This is how we will model surrenders.

We can thus define a dynamic model modeling both top-ups and lapses only function of the evolution of the difference Δ of the market rate and the life-insurance portfolio served rate.

- If Δ< α, then $lapse_{rate} = lapse_{min}$ the historical minimum lapse rate (minimum lapse <0, e.g. maximum top-up rate)
- If α <Δ< γ, then $lapse_{rate} = lapse_{min} \times \frac{\Delta-\gamma}{\gamma-\alpha}$(top-up scenario)
- If γ<Δ< θ, then $lapse_{rate} = 0$ (absence of portfolio dynamic movements)
- If θ<Δ< φ, then $lapse_{rate} = lapse_{max} \times \frac{\Delta-\varphi}{\theta-\varphi}$ (surrendering scenario)
- If Δ>φ, then $lapse_{rate} = lapse_{max}$ (maximum surrender rate)

3.3.3. Surrendering borders

3.3.3.1. Maximum dynamic lapse rate

Let's define the maximum value for the lapse rate, with M: the maximum lapse rate value and α: the probability that the lapse rate gets higher than M.

In this context:

$$P(r_t > M) = \alpha$$

$$P\left(\aleph\left(b, \frac{\sigma^2}{2a}\right) > M\right) = \alpha$$

$$P\left(\aleph(0,1) > \frac{M-b}{\sigma/\sqrt{2a}}\right) = \alpha$$

e.g. $M = b + \frac{\sigma}{\sqrt{2a}} q_\alpha$, with q_α of a normal distribution.

The main advantage of such model is that we will be able to include in our parameterization the specifications of each insurance product. Indeed:

$$Dynamic_{LR} = Historical_{LR} \times (1 + dynamic)$$

Annualizing the monthly lapse rate to get the thresholds makes sense: In the insurance business, insurers are required to have a prudential approach on their risk; expressing the lapse rate on an annual basis signifies that we consider that the maximal monthly lapse rate will be the same on the dealing year.

$$12M = Historical_{LR} \times (1 + dynamic_{max})$$

$$dynamic_{max} = 12 \times \frac{b + \frac{\sigma}{\sqrt{2a}} q_\alpha}{Historical_{LRmax}} - 1$$

On a monthly basis, the $Historical_{LRmax}$ over the last 12 months was 5.28%, which returns a dynamic maximum lapse rate at 5.604%.

3.3.3.2. Surrendering thresholds

Another tricky point of this modeling is the determination of the threshold representative of the moment when policyholders start and end surrendering their policy.

3.3.3.2.1. Starting surrendering point

Basically, we can define a risk premium π function of the surrender charges and the evolution of Δ**.** However,it has been decided that surrender fees in NPGWM are equal or closed to zero (free-decision argument towards the policyholder). However, based on a 12 months average on historical data, we can reasonably assume that, in case of a Δ value inferior to 1.15% [29], policyholder will not surrender his policy. We also have to include in the expression of the threshold a stressed interest rate risk. If we consider σ_Δ the volatility of the difference between the benchmark rate and the life insurance portfolio one, we can express the threshold for which policyholders will start to surrender his policy as, and with ω a coefficient factor:

$$\pi = motivation_{surrender} + \omega \times \sigma_\Delta \qquad (*)$$

We will assess for now ω as the quantity for which the policyholder will not invest in the market because of a too high volatility. We can reasonably assess for now a volatility of the market rate around 20% (volatility for corporate bonds around 10%, and equities around 30%[32]). Thus, (*) becomes:

$$\pi = 1.15 + 0.2 \times \sigma_\Delta$$

On a 1000 scenarios basis (variation of 0.10% between 750 and 3000 scenarios), we get as σ_Δ average an amount of 8.12%, which sets to $\pi = 2.774\%$the moment when policyholders are going to start surrendering because of higher interest rate offered by the market. As comparison figure, the inflation rate in the Eurozone in December 2013 oscillates around 0.7% on an annual basis and around 1.6% in 2013).

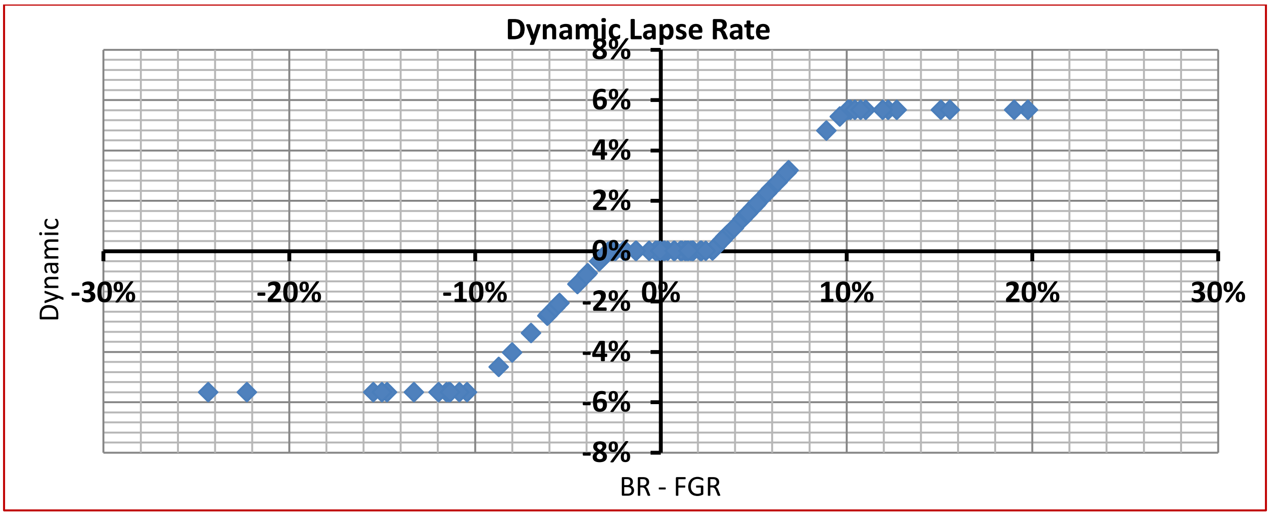
3.3.3.2.2. Ending surrendering point

Life insurance investors have a risk-prudential approach, e.g. they won't be lured by a high return with a high volatility. We assess, for now, to 10%the moment when policyholders will stop surrendering their policy to invest in the financial market [Annex 6]. Inasmuch as insurance products are supposed, proportionally, to replicate the market, the scenario where the difference between the market rate and the one offered by NPGWM be above 10% is highly unlikely.

3.3.3.3. Top-ups borders

Unfortunately, time was too short to model on a properly basis the top-ups. Assuming that top-ups react as lapse rates is a strong assumption. Indeed, "Loss aversion implies that one who loses $100 will lose more satisfaction than another person will gain satisfaction from a $100 windfall _ Daniel Kahnemann" [35] [Annex 5]

3.3.3.4. Dynamic lapse rate output



The NPGWM's life insurance business offer two types of investments: a guaranteed one and a no guaranteed one. Besides, a NPGWM product, Camelea, is provided with a stop-loss option in case of an excessive fall in the financial market. On a monthly basis, we assume in the modeling that the policyholder will surrender or reinvest in his policy in the same month.

- If the difference between the benchmark market rate (MSCI) and NPGWM served rate remains between -2.77% and 2.77%, the conjectural lapse rate is equal to 0.
- If the difference between the benchmark market rate (MSCI) and NPGWM served rate is equal to 6%, the estimated lapse rate is around 2%
- Conversely, if the difference between the benchmark market rate (MSCI) and NPGWM served rate is equal to -4%, policyholders will be reinvesting in their life insurance policies by 1.2%. In this scenario, the life insurance policy offers a more profitable return of investment than the financial markets.

3.4. Simulation of a Non-guaranteed rate portfolio

Lapse rates do not require a daily modeling, unlike the short term interest rate. Modeling it on a monthly basis seemed me accurate, implying a discrete version of the Vasicek model. The non-guaranteed rate model replicates the market (Reasonable assumption: Camelea investors are mainly investing in a mix between equities (60%), bonds and bond funds (40%)). The lapse rate models here the policyholder decision to switch his savings from his portfolio to a money-market fund (e.g. a closed-to-zero guaranteed rate fund).

3.4.1. Assumptions

Ann. Drift (m)	7,5%	Annual average of the market performance
Ann. Vol (sd)	20,0%	Annual volatility of the market
Timestep (dT)	0,0833	Pro-rata monthly basis
Cash Rendement	0,00%	Risk-free rate (in this case, money market fund returning rate, closed to zero)
a	-10,00%	Dynamic
b	-2,77%	Dynamic
g	2,77%	Dynamic
d	10,00%	Dynamic
MvMax	5,60%	Maximum lapse rate value
MvMin	-5,60%	Minimum lapse rate value
NumRuns	1 000	Number of simulated scenarios

Historical returns, from 1926 to 2010, for major asset classes in the United States[Annex 6] shows an average yield for bonds around 6% and 12% for equities, for a standard deviation of 20%. Considering that European historical returns were slightly under the US ones, assessing an annual drift of 7.5% with a standard deviation of 20% seems accurate and representative of the actual market trend.

The cash return for a money-market fund is, as we saw above, closed to zero: in case of a rallying financial market, the investor reinvests in his portfolio; in case of the opposite scenario, he surrenders his savings to switch them in a money market fund(e.g. with a cash return equal to zero to avoid losing more money).

The others values used have been all determined above. However, we won't present and model in this note a special scenario for top-ups. We will assume that they act as lapses in case of rising yields on financial markets.

3.4.2. Scenario

We express the market return as follows:

$$Market\_return_T = \ln(1+m) + sd \times \sqrt{dT} \times dW_t$$

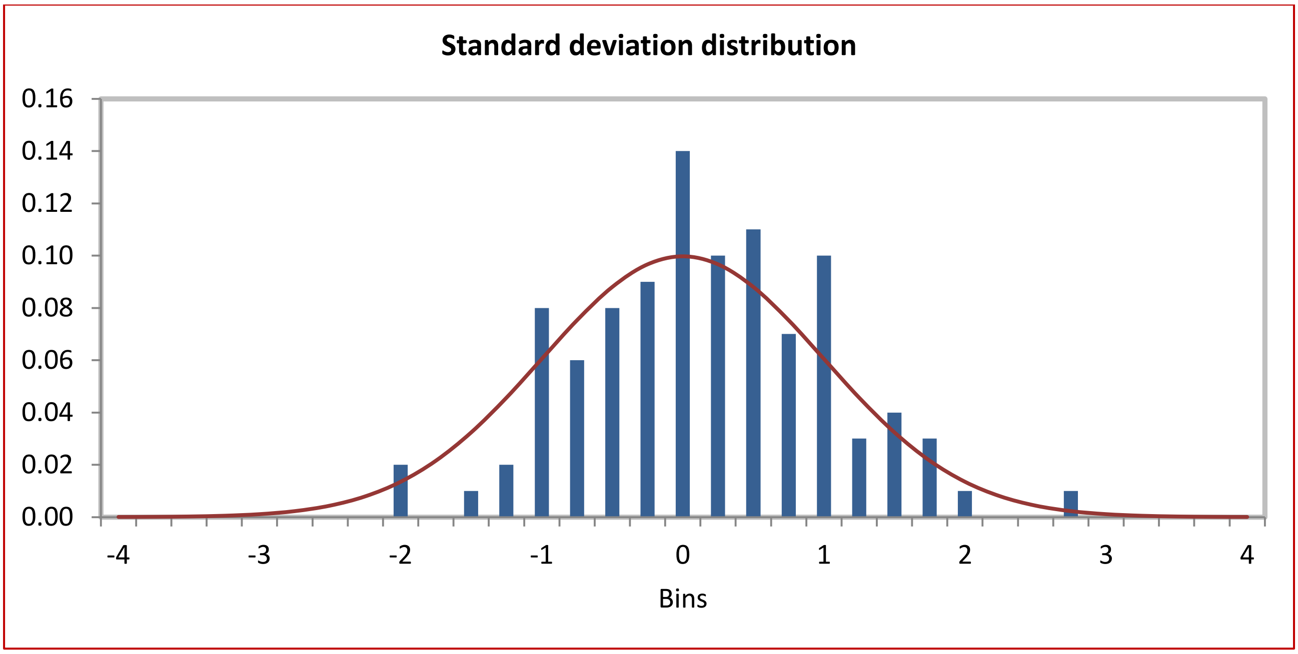
No guaranteed rate simplifies the Δ expression, which only depends of the market rate $\mathrm{Market\_return_T}$.

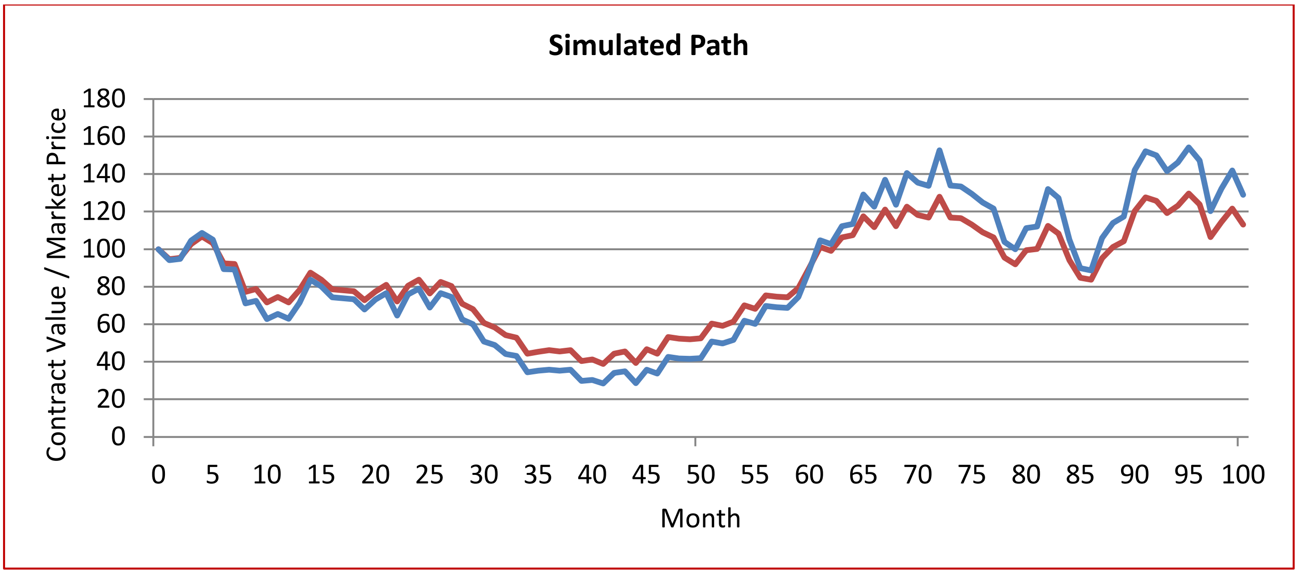
We do have an expression of the dynamic lapse rate based on the formula defined above, which gives us as a lapse-rate actualized contract value:

$$Contract\_value_T = Contract\_value_{T-1} \times e^{[market\_return_T + \ln(1+dynamic)]}$$

3.4.3. Non-guaranteed rate scenario outputs

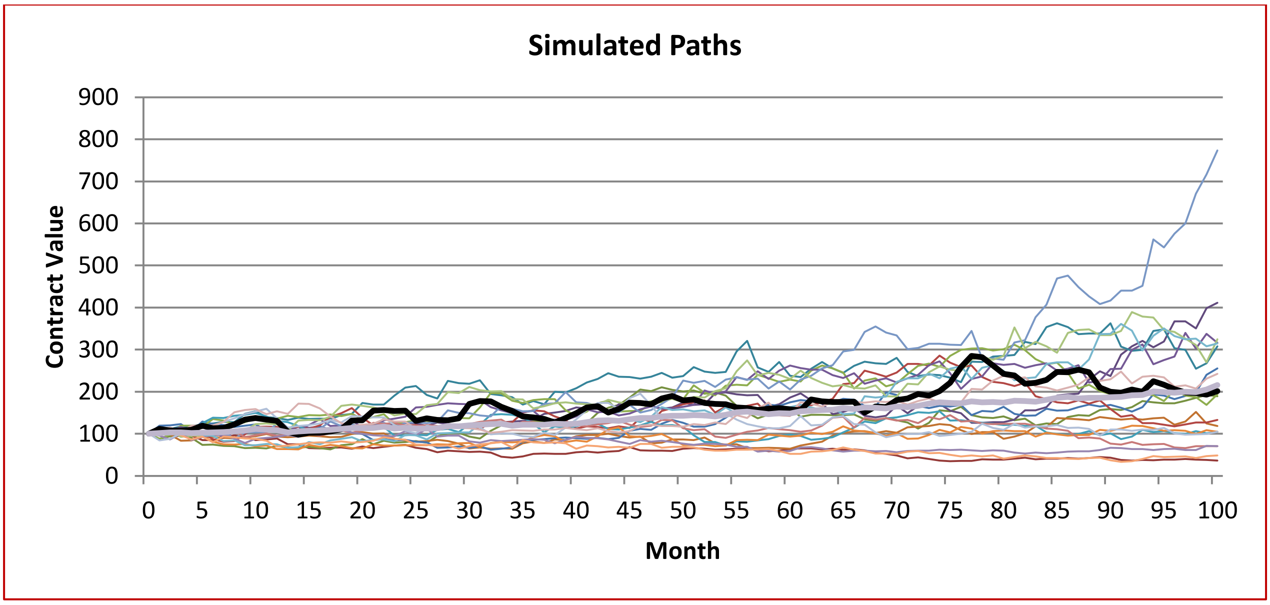
We get as first output the distribution of a discrete Brownian motion, in order to check that it follows correctly a normal law





For this simulated contract, when the market value (in red) increases, policyholders reinvest in their policy (top-ups); conversely, when it drops, they surrender, in order to protect their savings.

However, this only generates one sample of a contract value. With a VBA code [Annex 8], we generate 1,000 simulations in order to have averagely a contract value for varying levels of drift and volatility. We draw below an average contract value based on a sample of 20 contract simulations for a fixed drift and volatility.



The scenario presented here is extracted for the more likely-scenario on financial markets [Annex 7]: a drift equaled to 5% with a volatility of 20% [34]. The black line constitutes the average scenario.

We have an average summary of the evolution of the contract value actualized with the dynamic lapse rate and function of the monthly drift and the standard deviation of the market.

The output summary is as follows:

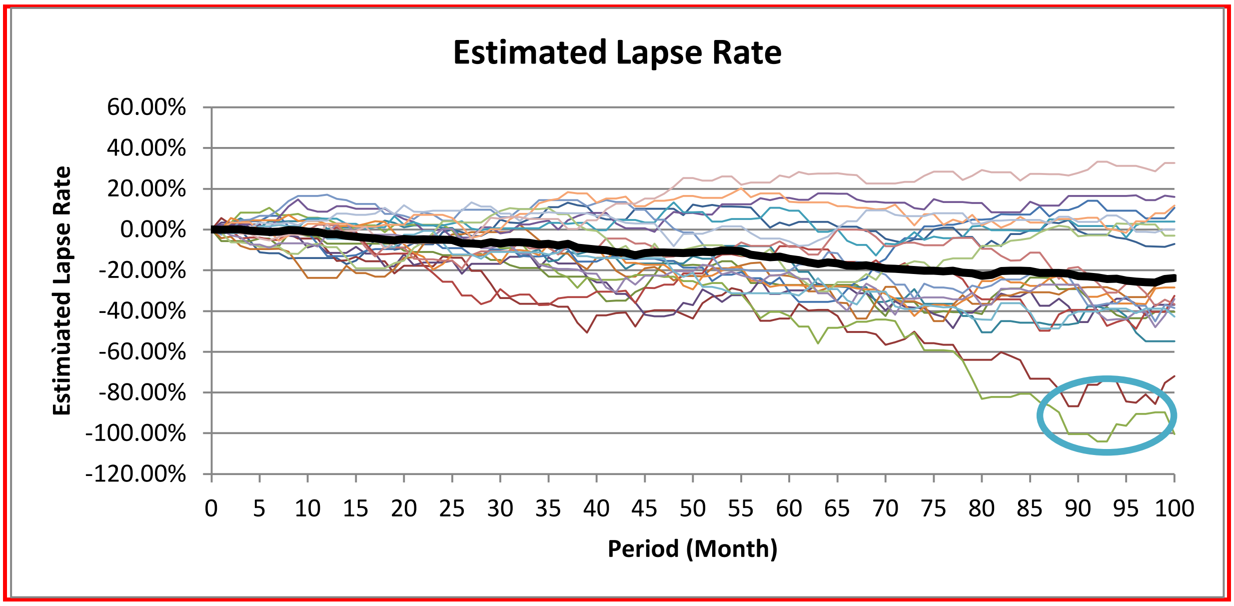
VBA SIMULATION					
RUN		Monthly Vol, Annualized (SD)			
		0%	10%	20%	30%
Monthly Ret, Annualized (Mu)	-25,0%	9,1	4,6	4,4	7,6
	-20,0%	15,6	9,7	9,4	14,3
	-15,0%	25,8	19,2	18,4	30,0
	-10,0%	41,6	34,1	38,3	58,3
	-5,0%	65,2	62,6	73,4	103,6
	0,0%	100,0	104,3	142,2	207,3
	5,0%	150,2	176,8	238,0	313,9
	10,0%	221,3	295,9	433,3	613,5
	15,0%	320,5	471,0	722,4	911,8
	20,0%	456,9	747,4	1 162,8	1 876,8
	25,0%	642,1	1 164,8	2 038,5	2 619,7

In dark blue, the more likely scenarios

We do have here a tool to predict the evolution of the contract value according to the monthly market return and the monthly volatility. Cells in dark blue represent the more likely scenarios [Annex 5]; the figure colored in white represents the scenario drawn at the top of the page.

The table summarizes an average of the 1,000 contract values simulations, actualized with lapses, top-ups and market drift function of varying levels of market drift and volatility, after 100 months (≈ 8 years).

3.4.4. Projection of the lapse rate



For the same more likely scenario (μ=5% and sd =20%), we observe a decreasing lapse rate, which is in line with our assumptions. We modeled a rising economic trend of the financial markets. In this scenario, a negative lapse rate value reflects a top-up. As a consequence, the decreasing curve signifies an evolution of top-ups over the simulation period. In presence of financial markets keeping rising over time, investors do not surrender their policies; at the contrary, thanks to an economic rising trend, they reinvest in it.

In case of a modelling of decreasing market rates, we would have had an increasing curve, e.g. an increase of the lapse rate over time.

Note: In the simulation, we notice that a scenario is under -100% (circled in blue). It does not correspond to a full surrender, but to a top-up, where the policyholder was reinvested, 90 and 95 months after subscribing his policy, more than 100% of his original premium.

3.5. Simulation of a guaranteed rate portfolio

3.5.1. Scenario

The structure is the same as the one presented above, except for the cash returned: We have here as main assumptions a guaranteed rate of a certain amount (set it by default for model testing at 5%).

RendementGaranti	5,00%	Risk-free rate (in this case, NPGWM portfolio guaranteed rate)

Every month, when the market rises, the policyholder can surrender5.60% in his contract; when it decreases, he can reinvest maximum 5.60% of its contract.

Δhas a new expression which also depends of the guaranteed return rate:

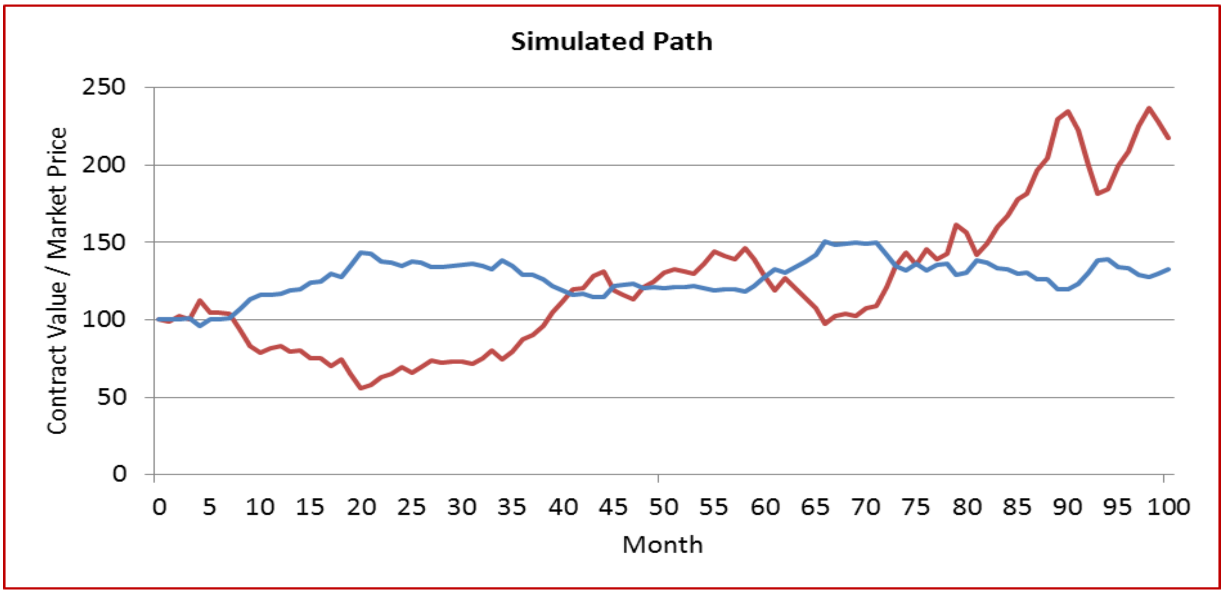
$$Market\_return_T = \ln(1+m) \times dT + sd \times \sqrt{dT} \times dW_t$$

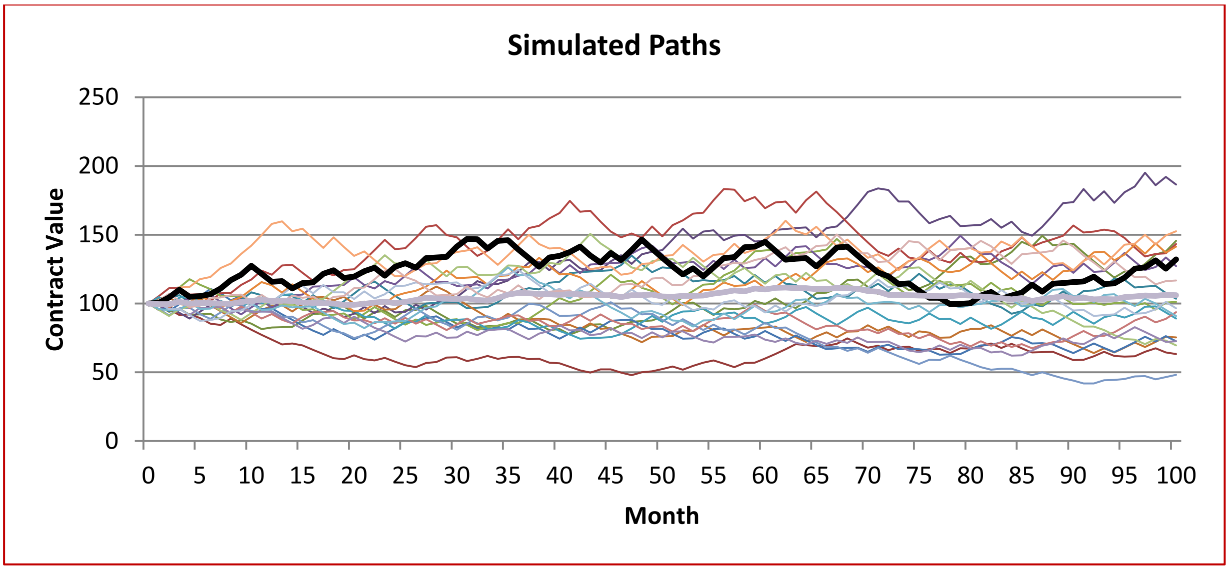
$$Portfolio\_return_T = \ln(1 + guaranteed\,rate) \times dT$$

$$\Delta = \boldsymbol{Market\_return_T} - \boldsymbol{Portfolio\_return_T}$$

3.5.2. Outputs

We use the same methodology for the guaranteed-rate outputs as the one previously presented above (red: market value; blue: contract value).





Average scenario in black

VBA SIMULATION					
RUN		Monthly Vol, Annualized (SD)			
		0,0%	10,0%	20,0%	30,0%
Monthly Ret, Annualized (Mu)	-25,0%	1 395,2	1 219,8	617,1	411,9
	-20,0%	721,4	769,6	446,8	326,5
	-15,0%	386,7	512,9	338,5	258,3
	-10,0%	214,0	327,0	243,4	212,8
	-5,0%	150,0	216,4	188,8	176,1
	0,0%	150,0	144,3	141,6	145,6
	5,0%	150,0	99,0	111,3	120,9
	10,0%	150,0	69,6	87,6	102,5
	15,0%	150,0	48,8	68,3	85,4
	20,0%	113,1	34,9	55,7	72,8
	25,0%	48,0	25,9	45,4	63,6

Results are not presented as the previous scenario: µ=-25% and sd= 0% represents a decreasing yield trend, implying a scenario with a lot of top-ups, the portfolio guaranteed rate being much more interesting than the market return . Conversely, for µ= 25% and sd= 30%, the market yield is much more interesting than the guaranteed portfolio return, implying a high lapse rate.

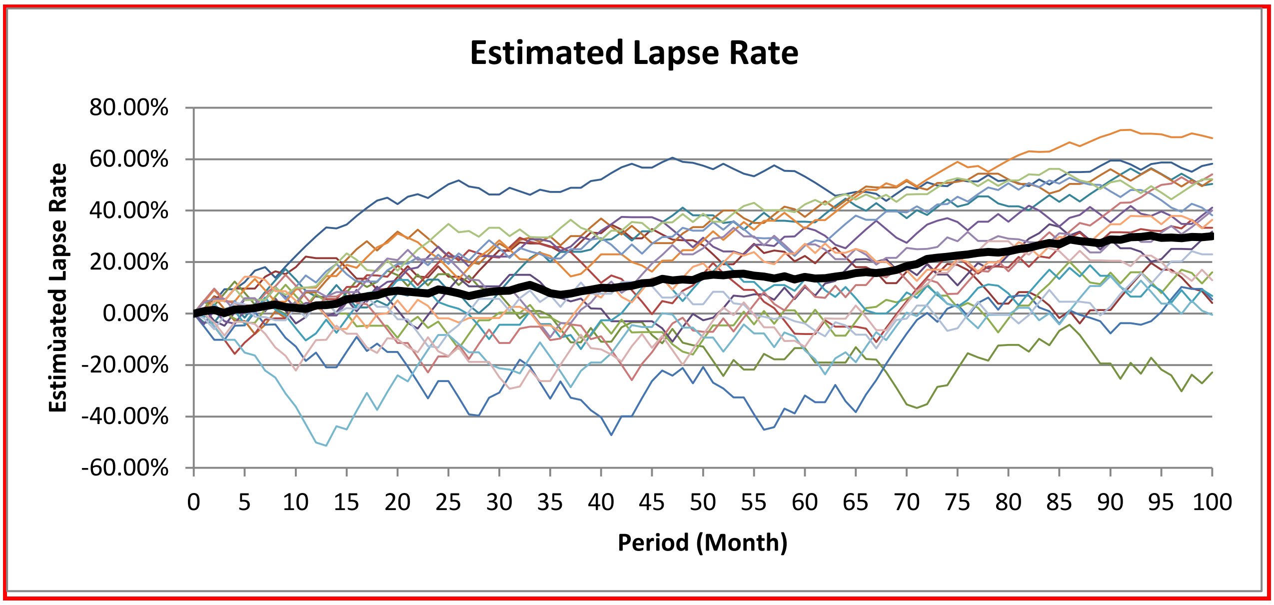
3.5.3. Projection of the lapse rate

We simulated here a contract value actualized with the lapse rate function of the market return evolution. We calculate the contract value free of lapse rate (e.g. only function of the guaranteed rate) over the simulation period.

We can now express, for each simulated scenario, the expected lapse rate over time in function of the both contract value as, for the month n.

$$LapseRate_{projected} = 1 - \frac{Contract\ value\ with\ surrender}{Contract\ value\ without\ surrender}$$

The black lines represents the average lapse rate of the 20 scenarios dawned.



Projected lapse rate for a market return of 5% and a volatility of 20%

The evolution of the lapse rate makes sense considering our assumptions. Indeed, we presumed an annual return of 5% per year, e.g. a rising trend. Logically, from a rising trend of the market trend results a rising lapses trend. People are in this scenario surrendering their guatanteed – rate policy because of more profitable investment solutions offered on financial markets (rising economic trend). A fixed served rate in an increasing market rates period increases the lapse rate on a guatanteed rate portfolios, more profitable investments being available on the financial markets.

3.6. Dynamic modelling, Advantages and limits

3.6.1. Advantages

The contract value is only a function of the market rate and requires only a few historical data (market drift and market volatility, but also lapse rate volatility and drift are necessary to model the dynamic lapse rate) and attached parameters. The approach is different from an historical one here. According to the market evolution, we predict what the lapse rate is going to be (indirectly, through the contract value), while the historical approach assumes investors are always go to act as they used to do in the past. We have now to calibrate the model with the company data to assess the lapse rate what the contract actualized with the lapse rate value will be in the future. Thus, the conjectural lapse rate is distinctly defined.

Moreover, we do not use historical data to draw an assumption on the average lapse rate (e.g. by default, we do not assume that people surrender or reinvest in their policy for no reasons, because it does not make economic sense –for instance, assuming an average lapse rate over the whole simulation period signifies that the product is meant to disappear over time). Instead, we use historical data to model relationships between lapse rates and market returns.

3.6.2. Limits

In the model, the policyholder surrenders or reinvests in his policy simultaneously with the market, which constitutes a strong assumption (policyholders are not all finance-aware for most of them; In NPGWM, dedicated funds do not represent a large part of the company business). Besides, it supposes that policyholders surrender in one time its policy, but do not surrenders the month after in case of a continuous fall of its portfolio. We may need to insert a lag between the time the market evolution and the insured will to surrender or reinvest in his policy (e.g. look deeper in the investor behavior [33]).

3.6.3. Conclusion

Our survival analysis showed us that the evolution of financial markets was one of the main driver regarding the policyholder decision to surrender. Besides, we noticed afterwards a correlation between market returns (European MSCI market rate) and the Camelea historical surrender rate. When the difference of both rates is higher than a certain threshold, policyholders start surrendering (lapses) invest their savings in financial markets (higher yield)in case of rising market or protect them in case of falling ones.Conversely, when the same difference is lower than another threshold, the policyholder reinvest in its policy (top-ups) – scenario when the Camelea portfolio offers a higher investment return than the market or more security (less volatile market).

4. Conclusion

With the Solvency II directive, life insurers are compelled to work with a prudential – risk approach. Under this regulatory framework, they must have a precise mapping of their risks.
The 2013 laspse SCR of PEL highlighted the fact that the lapse rate was one the biggest risks the life insurer faces. This ranking makes its modelling and understanding essential.

A survival analysis on the Camelea surrender data determined which factors are determinant in the policyholder decision to surrender or not: While the market rates, the age, the gender, or again the wealth of a policyholder directly impact the surrender risk, unemployment rate, risk profile ... have no impact.

Fitting a semi – parametrical model, the Cox Proportional Hazards model, introduced by Cox in 1972, gave us some interesting results regarding the ability for each job occupation to face the surrender risk. The study highlighted that the life insurer has to manage carefully the interest rate he is offering to its clients. Indeed, market yields impact a lot the policyholder's decision to surrender or not his policy.
Conversely, it might worth to target wealthy policyholders, the study showing that the wealthier they are, the less they surrender. Age does not impact significantly, but marginally the surrender risk. It makes sense that the older the policyholder gets, the more he retired – especially after 60, the average age for retirement. Finally, Female policyholders surrender less than male policyholders.

Being aware of all these behaviours regarding the surrender risk is an important information for the life insurer. Indeed, this one will be aware, starting today, and for any new policyholder investing within the Camelea portfolio, of its tendency to surrender his life insurance portfolio.

A correlation between equity market rates and surrender rates was a motivation to set a model predicting the surrender rate based on a dynamic lapse rate, function of the company interest served rate and the market one.

This type of modeling based on a dynamic scenario offers a lot of advantages. Firstly, investors do not surrender their contract for no reasons based on a scenario with a regular average lapse rate (underlying economically speaking that the portfolio is meant to disappear over time); they surrender in reaction of a higher yield on financial markets. Secondly, we do not use historical data to make an assumption on the average lapse rate but to model a relationship between lapse rates and market returns. Finally, we can estimate an average lapse trend over a 100 months period based on the market evolutions.

However, this model also shows its limits. We did not take into account macro data to define an investor profile as reference, and shock the dynamic lapse rate to model different classes of investors. Indeed, as we could have seen at the beginning of this chapter, policyholders do not react similarly in front of the surrender decision, and their socio – professional category has a large impact on this decision.
Besides, we do not model either the difference between arbitration and lapse decision, the political or Luxembourg legal hazard on the life insurance rules. These will constitutes the improvements to bring on the model.

5. Bibliography

[1] Commissariat aux Assurances, 2012/2013, Rapport annuel

[2] Lloyd's , What is Solvency II

[3] Deloitte, Solvency II requirements, Pillars & Embedded Value

[4] JC Flowers
http://www.ft.com/cms/s/0/5abb3802-9e15-11e1-9456-00144feabdc0.html#axzz3Ek7lMJxM

[5]EIOPA, 28 January 2013, Technnical specification on the Long Term Guarantee Assessment (Part 1), doc 13/061

[5 bis]EIOPA, 28 January 2013, Technnical specification on the Long Term Guarantee Assessment (Part 1), doc 13/061, pages 81 to 92

[6] , European Commission, 5 July 2010, QIS 5 Technical specifications, Financial institutions

[7] The European Insurance CFO Forum Market Consistent Embedded Value Principles, June 2008

[8]: Fiona Steele, Multilevel Discrete-time Event History Analysis , Centre for Multilevel Modelling Graduate School of Education University of Bristol

[9]: Lucy Fike, 2008, Censoring in Time-to-Event Analysis, the analysis factor

[10]: Mary Lunn, 2007,Definitions and censoring, Department of Statistics, University of Oxford

[11] Jianqing Fan, Non- and Semi- Parametric Modeling in Survival analysis, Princeton university

[12]: Bianca L. De Stavola, 2013, The Cox model, introduction and history, London School

[13]Victor Fardell, Ewen Gallic, 30 January 2013, Le modele de Cox, Faculté de Sciences economiques de Rennes

[14] Chapter 3 the cox proportional hazards model, Singapore University

[15] Introduction to survival analysis using R, Handout2b, Graduate Institute of Statistics National Central University Taiwan

[16] Cox – proportional hazards regression for survival data, John Fox, February 2002

[17] John Fox and Sanfors Weisberg, 23 February 2011, Cox – proportional hazards regression for survival data in R

[18] Jiezhi Qi, Comparison of proportional hazards and Accelerated Failure time models, Department of Mathematics and Statistics, University of Saskatchewan

[19] Survival analysis, Cox proportional hazards model, Keil university, Germany

[20] Chapter 5 extended and stratified Cox, Singapore university

[21] Aris Perperoglou, A fast routine for fitting Cox models with time – varying effects, Leiden University

[22] G. David Garson, North Carolina State University, Parametrical survival analysis

[23] Dr Daowen Zhang , Regression models for survival data; the Accelerated Failure Time model, Statistics department, University of British Columbia

[24] Patrick Roston, Estimating a smooth baseline hazard function for the Cox model, Hub for Trials Methodology Research, University College London, 13 September 2011

[25] Terry M. Therneau, Patricia M. Grambsch; Thomas R. Flemming, Martingale-based residuals for surivival models, University of Washington, Seattle

[26] Andre Berchtold, 8 April 2010, Statistics: Research, Teaching, Consulting

[27]Wiley, 2000, Expert Trading Systems: Modeling Financial Markets with Kernel Regression, chapter 2

[28] Étalonnage des modèles stochastiques de taux d'intérêt, Canadian Institute of Actuaries

[29]Hoffmann, Post, Pennings, 2011, Individual Investors and the Financial Crisis:How Perceptions Change, Drive Behavior, and Impact Performance

[30] Modèles financiers et analyses de risque dynamiques en assurance; Modélisations des obligations: présentation et utilisation en assurance; Support de cours 2013 2013; Frédéric Planchet; version2.3; Janvier 2013.

[31]Yériel Dynovisz, 2012, Réconciliation des différentes mesures de la performance d'un portefeuille de contrats d'épargne sous les référentiels MCEV, Solvabilité 2 et IFRS, mémoire d'actuariat

[32]AdrienSuru, 2011, Modélisation du rachat et parallèle avec la Physique, Memoire d'actuariatRemark: Allianz takes as risk premium a constant value of 1% according to an internal survey; the weight of Allianz life insurance business is about €5.4M in 2012, similar figure than NPGWM Assets Under Management

[33] Roger G. Ibbotson, The Equity Risk Premium

[34] PIMCO _ Forecasting Equity Returns in the New Normal http://www.pimco.com/EN/Insights/Pages/Forecasting-Equity-Returns-in-the-New-Normal-.aspx

[35] Daniel Kahnemann,Loss aversion

[36] Jeremy Kent, Ed Morgan, Dynamic policyholder behaviour, Staple Inn Actuarial Society, 18 November 2008

JRSS B, 1972, Cox, D.R. Regression models and life tables (with discussion)

Takis Konstantopoulos, Notes on survival models, Heriot – Watt University

V. Nguyen, Time – varying covariates, survival analysis, UCIrvine

Stephen J Walters, What is a Cox model, ScHARR, University of Scheffield

Surrender triggers in life insurance: What main features affect the surrender behavior in a classical economic context?

Ya-fang Yan, Survival Anlaysis Final Report, Data Analysis

Anne- Gael Lugand, 2010, Evaluation des barèmes de provisions du contrat Dépendance du BCAC, memoire d'actuariat

How to simulate a Cox proportional hazards model with change point, 2013, stats.stackexchange

Sanjiv R. Das, Rangarajan K. Sundaram, Suresh M. Sundaresan, August 2003, A simple unified model for pricing derivatives, securities with equity, interest – rate and default risk

Frederic Planchet, Juillet 2013, Modèles de durée; Statistiques des modèles non - paramétriques, Support de cours 2013 2014

BNP Paribas Assurances, Modèles financiers et analyses de risques dynamiques en assurance

Davind Wang, 21 May 2008, Pricing varaible annyuity with embedded guarantees, FSA, Milliman

Xavier Milhaud, 11 July 2011, Segmentation et modelisation des comportements de rachat en Assurance Vie, Mémoire d'actuariat, AXA Global Life

Arthur Charpentier, Christophe Dutang, December 2012L'actuariat avec R

Samuel H. Cox, Yijia Lin, 15 November 2013, Annuity Lapse rate modeling: Tobit or not Tobit

Changki Kim, 2005, Surrender rate impacts on asset/liability management. Asia-Pacific Journal of Risk and Insurance, 1

Nick Wilkinson, 2008, An introduction to behavioral economics

6. Annexes

6.1. Annex 1 : Excell Life irregularities

http://www.commassu.lu/FR/documents/excell-life-international.asp

COMMISSARIAT AUX ASSURANCES

HOME | NOTRE MISSION | ACTUALITÉS | DOCUMENTS | STATISTIQUES | OPÉRATEURS

COMMISSARIAT AUX ASSURANCES

LIENS CONTACT

Documents

- Lois
- Circulaires
- Rapports Annuels
- Règlements
- Notes d'information
- **Excell Life International S.A.**

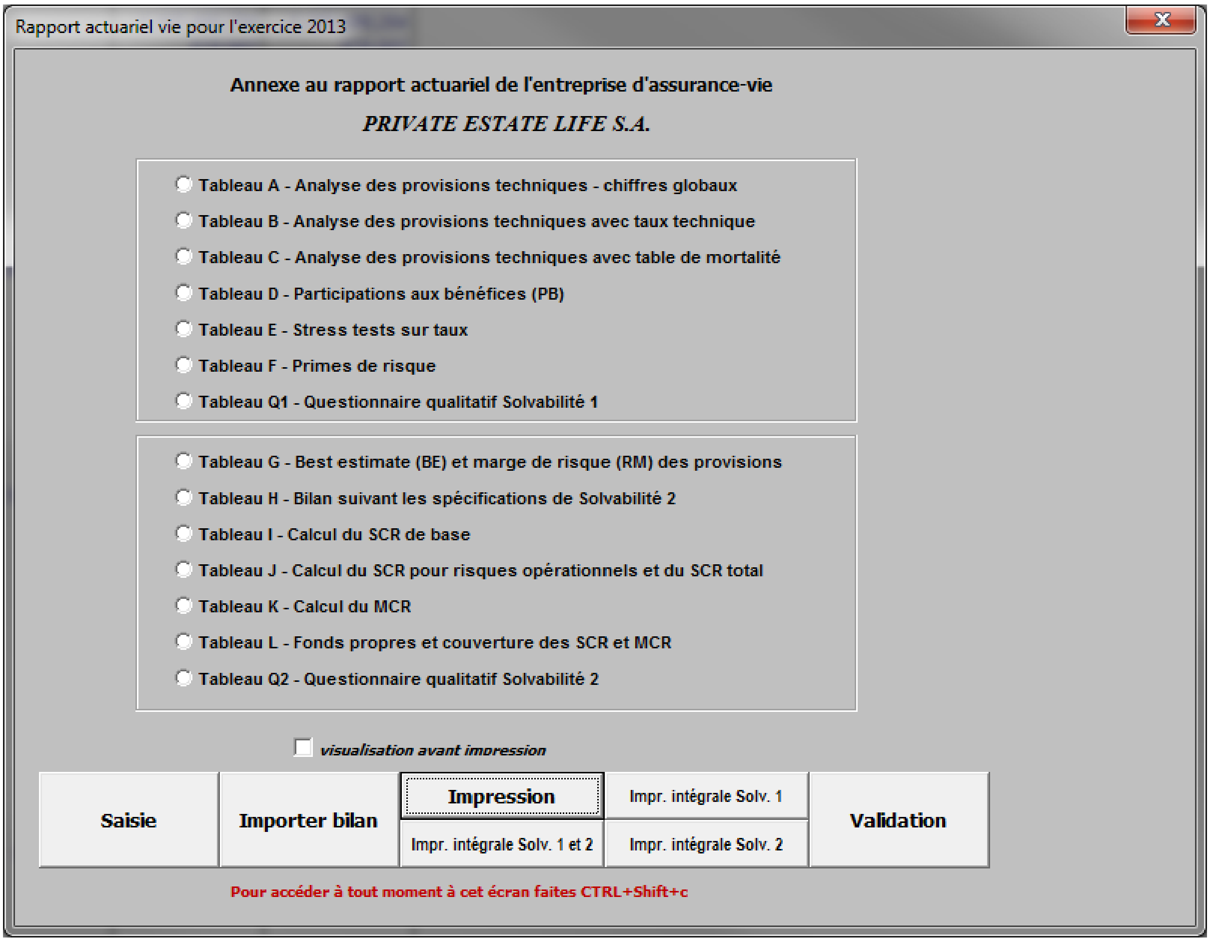
Excell Life International S.A.		
05/08/2010	Mesures prises à l'égard d'EXCELL LIFE INTERNATIONAL S.A. (5 août 2010)	download
15/02/2011	Mesures prises à l'égard d'EXCELL LIFE INTERNATIONAL S.A. (15 février 2011)	download
19/04/2011	Non renouvellement des mesures de restriction à l'égard d'EXCELL LIFE INTERNATIONAL S.A.(19 avril 2011)	download
17/11/2011	Sanctions prises à l'égard d'EXCELL LIFE INTERNATIONAL S.A.(17 novembre 2011)	download
05/06/2012	Arrêté ministériel du 05.06.2012 portant retrait de l'agrément de Excell Life International S.A.	download
13/06/2012	Mesures prises à l'égard de EXCELL LIFE INTERNATIONAL S.A. (13 juin 2012)	download
27/06/2012	Retrait d'agrément de EXCELL LIFE INTERNATIONAL S.A.	download
12/07/2012	Jugement de mise en liquidation de Excell Life International S.A. du 12.07.2012	download
17/07/2012	Publication au Mémorial B du 17.07.2012 de l'arrêté ministériel de retrait de l'agrément de Excell Life International S.A.	download
21/08/2012	Mise en place d'un site Internet pour Excell Life International S.A. en liquidation	download
21/08/2012	Site internet mis en place par les liquidateurs de Excell Life International S.A.	download

6.2. Annex 2: CAA quarterly request ; asset classification on guaranteed-rate portfolios

Nom du groupe contrepartie	Type de contrepartie	Exposition brute directe
BEI - EIB	émetteur public	1
Communauté Européenne	émetteur public	2
Autriche	émetteur public	3
Pologne	émetteur public	4
Pays-Bas	émetteur public	5
Danemark	émetteur public	6
Finlande	émetteur public	7
Luxembourg	émetteur public	8
France	émetteur public	9
Etats-Unis	émetteur public	10
Groupe dont fait partie PRIVATE ESTATE LIFE S.A.	émetteurs intragroupe	
Nykredit	groupe bancaire/conglomérat financier	1
DNB ASA	groupe bancaire/conglomérat financier	2
RBC	groupe bancaire/conglomérat financier	3
Nordea	groupe bancaire/conglomérat financier	4
Danske Bank	groupe bancaire/conglomérat financier	5
Dexia	groupe bancaire/conglomérat financier	6
GE CAPITAL EURO FUNDING _ Autre groupe bancaire	groupe bancaire/conglomérat financier	7
ING	groupe bancaire/conglomérat financier	8
SWEDBANK HYPOTEK AB _ Autre groupe bancaire	groupe bancaire/conglomérat financier	9
L-BANK BW FOERDERBANK _ Autre groupe bancaire	groupe bancaire/conglomérat financier	10
	groupe de (ré)assurances	
	groupe de (ré)assurances	
	groupe de (ré)assurances	
	groupe de (ré)assurances	
	groupe de (ré)assurances	
	groupe de (ré)assurances	
	groupe de (ré)assurances	
	groupe de (ré)assurances	
	groupe de (ré)assurances	
	groupe de (ré)assurances	
Autobahn Schnell AG	autres	1
Volkswagen	autres	2
TDC AS	autres	3
Carlsberg Breweries	autres	4
Glaxosmithkline Capital	autres	5
Asfinag	autres	6
Total Capital SA	autres	7
NV Nederlandse Gasunie	autres	8
Mondelez International	autres	9
RWE Finance BV	autres	10

Exposition figures have been changed

6.3. Annex 3: CAA Solvency II reporting spreadsheet



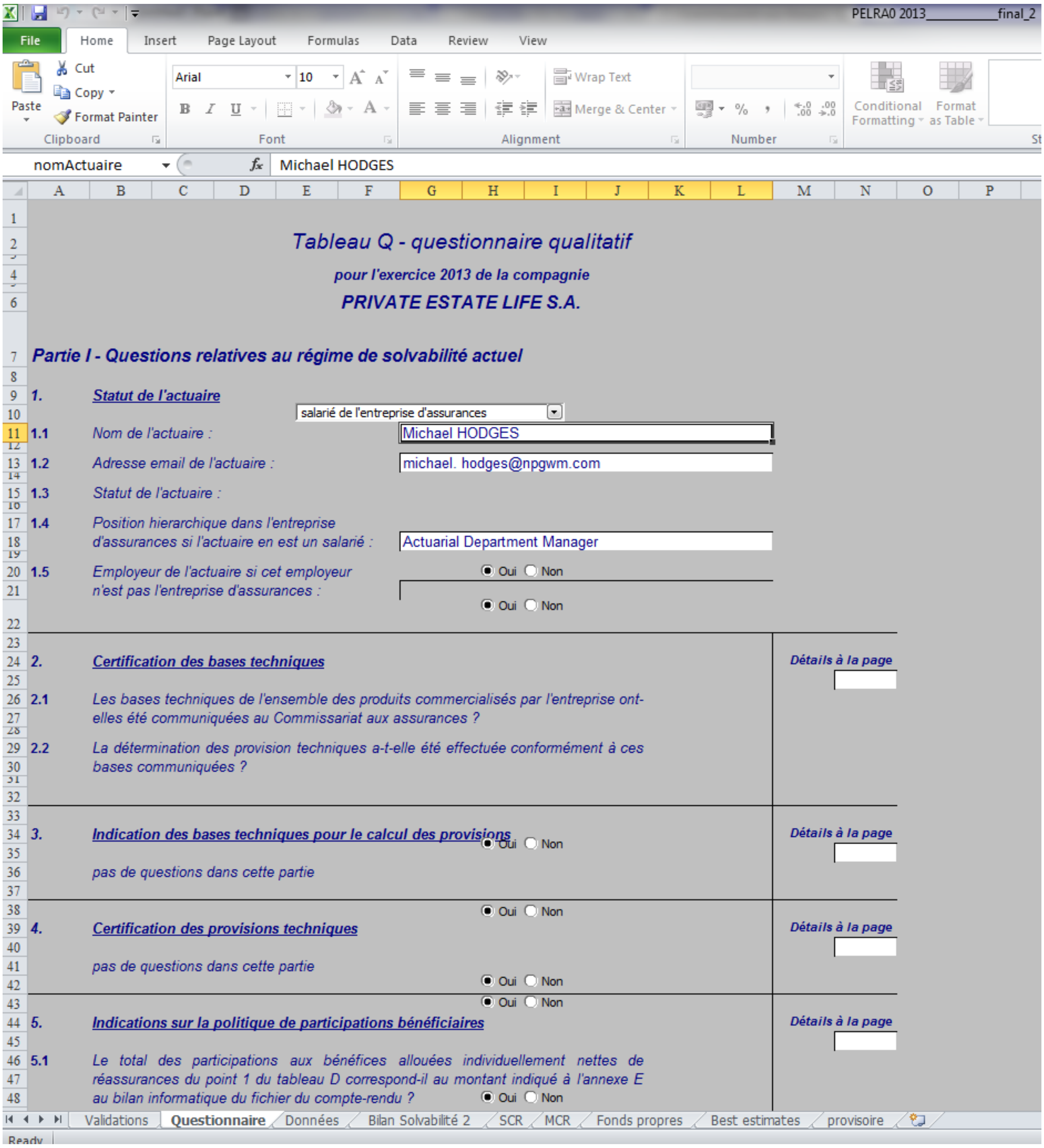


Tableau Q - questionnaire qualitatif

pour l'exercice 2013 de la compagnie

PRIVATE ESTATE LIFE S.A.

Partie I - Questions relatives au régime de solvabilité actuel

1. Statut de l'actuaire — salarié de l'entreprise d'assurances

1.1 Nom de l'actuaire : Michael HODGES

1.2 Adresse email de l'actuaire : michael. hodges@npgwm.com

1.3 Statut de l'actuaire :

1.4 Position hierarchique dans l'entreprise d'assurances si l'actuaire en est un salarié : Actuarial Department Manager

1.5 Employeur de l'actuaire si cet employeur n'est pas l'entreprise d'assurances : ◉ Oui ○ Non

◉ Oui ○ Non

2. Certification des bases techniques — Détails à la page

2.1 Les bases techniques de l'ensemble des produits commercialisés par l'entreprise ont-elles été communiquées au Commissariat aux assurances ?

2.2 La détermination des provision techniques a-t-elle été effectuée conformément à ces bases communiquées ?

3. Indication des bases techniques pour le calcul des provisions ◉ Oui ○ Non — Détails à la page

pas de questions dans cette partie

◉ Oui ○ Non

4. Certification des provisions techniques — Détails à la page

pas de questions dans cette partie

◉ Oui ○ Non

◉ Oui ○ Non

5. Indications sur la politique de participations bénéficiaires — Détails à la page

5.1 Le total des participations aux bénéfices allouées individuellement nettes de réassurances du point 1 du tableau D correspond-il au montant indiqué à l'annexe E au bilan informatique du fichier du compte-rendu ? ◉ Oui ○ Non

6.4. Annex 4: Survival model, R code

```
############################
###########Survival Analysis
####################################
####################################

#### 1 st Data set ####

dat <-
read.csv("C:/Users/Zachou/Desktop/FINAL/
New/COXF3.csv",header=T, sep=";")

summary(dat)

#library(simPH)

library(survival)

library(KMsurv)

## Data illustration

dat[200:205,1:14]

##correlation

cor(dat[,4:14])

######### Fitting a Cox model #########

##Cox model

N <- 9159

M0 <- coxph(Surv(Duration,
Event)~Portfolio+Still+DeltaU+DeltaS+Risk+J
ob+Age+Gender,data=dat)

M0

summary(M0)

##Cox model with significant covariates

M0 <- coxph(Surv(Duration,
Event)~Portfolio+Still+DeltaS+Age+Gender,d
ata=dat)

M0

summary(M0)

#### Survival & cumulative hazard ####

##survival function _ Kaplan -Meier
estimate

EM0 <- survfit(M0)

quantile(EM0, probs=c(0.25,0.5,0.75),
conf.int=FALSE)

#pseudo-observation

plot(EM0,main="Survival function within
the Camelea protfolio", xlab="Time (in
years)", ylab="Survival", lwd=2, col="green")

## Impact of surrenders ##

#Partial surrenders

Data_Still <- subset(dat,Still==0)

survie_ptf <-
Surv(Data_Still$Duration,Data_Still$Event)

M0_Still <-
coxph(survie_ptf~Portfolio+DeltaS+Age+Ge
nder, data=Data_Still)

M0_Still

EMO_Still <- survfit(M0_Still)

#Full surrenders

Data_Not <- subset(dat,Still==1)

not_ptf <-
Surv(Data_Not$Duration,Data_Not$Event)

M0_Not <-
coxph(not_ptf~Portfolio+DeltaS+Age+Gend
er, data=Data_Not)

M0_Not

EM0_Not <- survfit(M0_Not)

# Survival plot

plot(EM0,main="Survival function: Impact of
surrenders", xlab="Time (in years)",
ylab="Survival", lwd=2, col="green")

lines(EMO_Still, col="red2")

lines(EM0_Not, col="blue2")

legend (x="bottomleft", lwd=2,
col=c("green","red","blue"),
legend=c("Global Survival
function","Survival - Partial surrenders
only","Survival - Full surrenders only"))

## Cumulative hazard

#Cumulative bazeline hazard

expcoef <- exp(coef(M0))

Lambda1 <-basehaz(M0, centered = FALSE)

summary(Lambda1)

Lambda1A <- expcoef[1]*Lambda1$hazard
#savings

Lambda1B <- expcoef[3]*Lambda1$hazard
#SX5T

Lambda1C <- expcoef[4]*Lambda1$hazard
#Age

Lambda1D <- expcoef[5]*Lambda1$hazard
#Gender

plot(hazard ~ time, main = "Cox PH Estimate
- Cumulative hazard", type="s",
xlab="Duration", ylab="Cumulative hazard",
ylim=c(0,10),lwd=4, data=Lambda1)

lines(Lambda1$time, Lambda1A, lwd=2,
col="red")

lines(Lambda1$time, Lambda1B, lwd=2,
col="green")

lines(Lambda1$time, Lambda1C, lwd=2,
col="blue")

lines(Lambda1$time, Lambda1D, lwd=2,
col="orange")

legend (x="topleft", lwd=2,
col=c("black","red","green","blue",
"orange"), legend=c("pseudo-
observation","Portfolio","DeltaS","Age","Ge
nder"))

######## Checking COX PH model
assumptions ########

# Schoenfeld residuals

coxres <- cox.zph(M0)

par(mfrow=c(3,2))

plot(coxres, main="Shoenfeld residuals")

cox.zph(M0)

##martingale residuals vs non-dichotomic
covariates

res.m <- residuals(M0,type="martingale")

res.m

par(mfrow=c(1,1))

X <-as.matrix(dat[,c("Still","DeltaS")])
```

```
for (j in 1:2) {

plot(X[,j],res.m,xlab=c("Still","Market rate movements")[j],ylab="residuals", main="Martingale residuals")

abline(h=0,lty=2,col="red")

lines(lowess(X[,j],res.m,iter=0), col="green2", lwd=2)}

# Partial residuals

b <- coef(M0)[c(2,3)]

par(mfrow=c(1,1))

for (j in 1:2) {

plot(X[,j], b[j]*X[,j]+res.m, main="Link: exponential test", xlab=c("Still","Market rate movements")[j], ylab="component+residuals")

abline(lm(b[j]*X[,j]+res.m~X[,j]), lty=2, col="red2")

lines(lowess(X[,j],b[j]*X[,j]+res.m,iter=0), col="green2", lwd=2)}

#DfBeta

DFM0 <- residuals(M0, type='dfbeta')

par (mfrow=c(2,3))

for (j in 1:5) {

plot(DFM0[,j], ylab=names(coef(M0))[j])

abline(h=0,lty=2, lwd=2,col="red2")}

################################# Time dependent covariates ####################################
####################################

###### 2 st Data set ######

dat <- read.csv("C:/Users/Zachou/Desktop/FINAL/New/COXT1.csv",header=T, sep=";")

summary(dat)

library(survival)

dat[200:205,1:11]

## fitting a Cox model ##

M0 <- coxph(Surv(start,stop, Event)~Portfolio+DeltaS+Age+Gender,data=dat)

M0

summary(M0)

## Check CoxPH assumptions ##

#Schoenfeld residuals

cox.zph(M0)

#Obs# Too many classes in Savings => need to redefine the covariate

#DfBeta residuals

DFM0 <- residuals(M0, type='dfbeta')

par (mfrow=c(2,2))

for (j in 1:4) {

plot(DFM0[,j], ylab=names(coef(M0))[j], main="DfBeta residuals")

abline(h=0,lty=2, lwd=2,col="red2")}

####### 3 rd Data set #######

## Recoded Portfolio covariates into 4 classes

dat <- read.csv("C:/Users/Zachou/Desktop/FINAL/New/COXT2.csv",header=T, sep=";")

summary(dat)

library(survival)

## Fitting a Cox model ##

M0 <- coxph(Surv(start,stop, Event)~Portfolio+DeltaS+Age+Gender,data=dat)

M0

summary(M0)

## Checking assumptions ##

#Schoenfeld residuals

cox.zph(M0)

#good results: no time dependent covariates: HP verified

#DfBeta residuals

DFM0 <- residuals(M0, type='dfbeta')

par (mfrow=c(2,2))

for (j in 1:4) {

plot(DFM0[,j], ylab=names(coef(M0))[j], main="DfBeta residuals")

abline(h=0,lty=2, lwd=2,col="red2")}

#################################### ####### Study on Jobs ##############################

###JOB1

dat1 <- read.csv("C:/Users/Zachou/Desktop/FINAL/New/JobA/Job1.csv",header=T, sep=";")

M1 <- coxph(Surv(start,stop, Event)~Portfolio+Still+DeltaS+Age+Gender, data=dat1)

M1

summary(M1)

# OUT

M1 <- coxph(Surv(start,stop, Event)~DeltaS, data=dat1)

M1

cox.zph(M1)

###JOB2

dat2 <- read.csv("C:/Users/Zachou/Desktop/FINAL/New/JobA/Job2.csv",header=T, sep=";")

M2 <- coxph(Surv(start,stop, Event)~Portfolio+Still+DeltaS+Age+Gender, data=dat2)

M2

summary(M2)

#OUT

M2 <- coxph(Surv(start,stop, Event)~Gender+DeltaS, data=dat2)

M2

cox.zph(M2)
```

```
###JOB3

dat3 <-
read.csv("C:/Users/Zachou/Desktop/FINAL/
New/JobA/Job3.csv",header=T, sep=";")

M3 <- coxph(Surv(start,stop,
Event)~Portfolio+Still+DeltaS+Age+Gender,
data=dat3)

M3

#OUT

M3 <- coxph(Surv(Duration,
Event)~Portfolio+DeltaS, data=dat3)

M3

summary(M3)

cox.zph(M3)


###JOB4

dat4 <-
read.csv("C:/Users/Zachou/Desktop/FINAL/
New/JobA/Job4.csv",header=T, sep=";")

M4 <- coxph(Surv(start,stop,
Event)~Portfolio+Still+DeltaS+Age+Gender,
data=dat4)

M4

#OUT

M4 <- coxph(Surv(Duration,
Event)~Portfolio, data=dat4)

M4

summary(M4)

cox.zph(M4)


###JOB5

dat5 <-
read.csv("C:/Users/Zachou/Desktop/FINAL/
New/JobA/Job5.csv",header=T, sep=";")

M5 <- coxph(Surv(start,stop,
Event)~Portfolio+Still+DeltaS+Age+Gender,
data=dat5)

M5

#OUT

M5 <- coxph(Surv(Duration,
Event)~Portfolio+Gender, data=dat5)

M5

summary(M5)

cox.zph(M5)


###JOB6

dat6 <-
read.csv("C:/Users/Zachou/Desktop/FINAL/
New/JobA/Job6.csv",header=T, sep=";")

M6 <- coxph(Surv(start,stop,
Event)~Portfolio+Still+DeltaS+Age+Gender,
data=dat6)

M6

#OUT

M6 <- coxph(Surv(start,stop,
Event)~DeltaS+Age+Gender, data=dat6)

M6

cox.zph(M6)


###JOB7

dat7 <-
read.csv("C:/Users/Zachou/Desktop/FINAL/
New/JobA/Job7.csv",header=T, sep=";")

M7 <- coxph(Surv(start,stop,
Event)~Portfolio+Still+DeltaS+Age+Gender,
data=dat7)

M7

#OUT

M7 <- coxph(Surv(start,stop, Event)~DeltaS,
data=dat7)

M7

cox.zph(M7)


###JOB8

dat8 <-
read.csv("C:/Users/Zachou/Desktop/FINAL/
New/JobA/Job8.csv",header=T, sep=";")

M8 <- coxph(Surv(start,stop,
Event)~Portfolio+Still+DeltaS+Age+Gender,
data=dat8)

M8

#OUT

M8 <- coxph(Surv(start,stop,
Event)~DeltaS+Age, data=dat8)

M8

summary(M8)

cox.zph(M8)


###JOB9

dat9 <-
read.csv("C:/Users/Zachou/Desktop/FINAL/
New/JobA/Job9.csv",header=T, sep=";")

M9 <- coxph(Surv(start,stop,
Event)~Portfolio+Still+DeltaS+Age+Gender,
data=dat9)

M9

#OUT

M9 <- coxph(Surv(start,stop,
Event)~DeltaS+Age, data=dat9)

M9

summary(M9)

cox.zph(M9)


###JOB10

dat10 <-
read.csv("C:/Users/Zachou/Desktop/FINAL/
New/JobA/Job10.csv",header=T, sep=";")

M10 <- coxph(Surv(start,stop,
Event)~Portfolio+Still+DeltaS+Age+Gender,
data=dat10)

M10

#OUT

M10 <- coxph(Surv(start,stop,
Event)~Portfolio+DeltaS+Age+Gender,
data=dat10)

M10

summary(M10)

cox.zph(M10)


###JOB11

dat11 <-
read.csv("C:/Users/Zachou/Desktop/FINAL/
New/JobA/Job11.csv",header=T, sep=";")

library(survival)

M11 <- coxph(Surv(start,stop,
Event)~Portfolio+Still+DeltaS+Age+Gender,
data=dat11)

M11

#OUT
```

```
M11 <- coxph(Surv(start,stop,
Event)~DeltaS, data=dat11)

M11

summary(M11)

cox.zph(M11)

####################################
#### Fitting AFT model
####################################
####

dat <-
read.csv("C:/Users/Zachou/Desktop/FINAL/
New/COXF3.csv",header=T, sep=";")

summary(dat)

library(survival)

attach(dat)



###Testing parametric models over data

AFT0_1 <- survreg(Surv(Duration,
Event)~Portfolio+DeltaS+Age+Gender,
data=dat, dist="lognormal")

summary(AFT0_1)

AFT0_2 <- survreg(Surv(Duration,
Event)~Portfolio+DeltaS+Age+Gender,
data=dat, dist="weibull")

summary(AFT0_2)

AFT0_3 <- survreg(Surv(Duration,
Event)~Portfolio+DeltaS+Age+Gender,
data=dat, dist="loglogistic")

summary(AFT0_3)

AFT0_4 <- survreg(Surv(Duration,
Event)~Portfolio+DeltaS+Age+Gender,
data=dat, dist="gaussian")

summary(AFT0_4)

AFT0_5 <- survreg(Surv(Duration,
Event)~Portfolio+DeltaS+Age+Gender,
data=dat, dist="exponential")

summary(AFT0_5)

AFT0_6 <- survreg(Surv(Duration,
Event)~Portfolio+DeltaS+Age+Gender,
data=dat, dist="t")

summary(AFT0_6)

AFT0_7 <- survreg(Surv(Duration,
Event)~Portfolio+DeltaS+Age+Gender,
data=dat, dist="extreme")

summary(AFT0_7)

### Model comparison

anova(AFT0_1,AFT0_2,AFT0_3,AFT0_4,AFT0
_5, AFT0_6, AFT0_7)

## => model selected: weibull (AIC test;
lowest value)



################### Parametrical
Analysis
####################################
#################

## Weibull AFT model

weibull.aft <- survreg(Surv(Duration,
Event)~Portfolio+DeltaS+Age+Gender,
data=dat, dist="weibull")

summary(weibull.aft)



#DfBetas

DFM0 <- residuals(weibull.aft,
type='dfbeta')

par (mfrow=c(2,2))

for (j in 1:4) {

plot(DFM0[,j], ylab=names(coef(M0))[j],
main="Beta residuals")

abline(h=0,lty=2, col="red")}



######## Survival analysis ########

## Estimated Survival curves ##

par(mfrow=c(1,1))

#Intercept

curve(pweibull(x,scale=exp(coef(weibull.aft)
[1]), shape=1/weibull.aft$scale,
lower.tail=FALSE),from=0,
to=max(dat$Duration), col="black", lwd=5,
main="Survival function"
,ylab=expression(hat(S)(t)), xlab="Time(in
years)")

#Portfolio

curve(pweibull(x,scale=exp(coef(weibull.aft)
[1]+coef(weibull.aft)[2]),
shape=1/weibull.aft$scale,
lower.tail=FALSE),from=0,
to=max(dat$Duration), col="green2",add=T
)

#DeltaS

curve(pweibull(x,scale=exp(coef(weibull.aft)
[1]+coef(weibull.aft)[4]),
shape=1/weibull.aft$scale,
lower.tail=FALSE),from=0,
to=max(dat$Duration), col="blue2",add=T)

#Age

curve(pweibull(x,scale=exp(coef(weibull.aft)
[1]+coef(weibull.aft)[5]),
shape=1/weibull.aft$scale,
lower.tail=FALSE),from=0,
to=max(dat$Duration),
col="yellow2",add=T)

#Gender

curve(pweibull(x,scale=exp(coef(weibull.aft)
[1]+coef(weibull.aft)[6]),
shape=1/weibull.aft$scale,
lower.tail=FALSE),from=0,
to=max(dat$Duration), col="red2",add=T)

legend (x="bottomleft", lwd=2,
col=c("black","green","blue","yellow",
"red"), legend=c("pseudo-
observation","Portfolio","DeltaS","Age","Ge
nder"))



## Plotting Estimated Weibull densities ##

curve(dweibull(x,
scale=exp(coef(weibull.aft)[1]),shape=1/wei
bull.aft$scale), from=0,
to=max(dat$Duration),
ylab="Density",ylim=c(0,0.4),
xlab="Duration",main="Estimated Weibull
densities",axes=F, lwd=5, col="black")

axis(1,cex.axis=.8)

axis(2,cex.axis=.8)

box()

curve(dweibull(x,
scale=exp(coef(weibull.aft)[1]+coef(weibull.
aft)[2]),shape=1/weibull.aft$scale), from=0,
to=max(dat$Duration), add=T, col="red2")

curve(dweibull(x,
scale=exp(coef(weibull.aft)[1]+coef(weibull.
aft)[4]),shape=1/weibull.aft$scale), from=0,
to=max(dat$Duration), add=T,col="green2")

curve(dweibull(x,
exp(coef(weibull.aft)[1]+coef(weibull.aft)[5]
),shape=1/weibull.aft$scale), from=0,
to=max(dat$Duration),
add=T,col="yellow2")

curve(dweibull(x,
exp(coef(weibull.aft)[1]+coef(weibull.aft)[6]
```

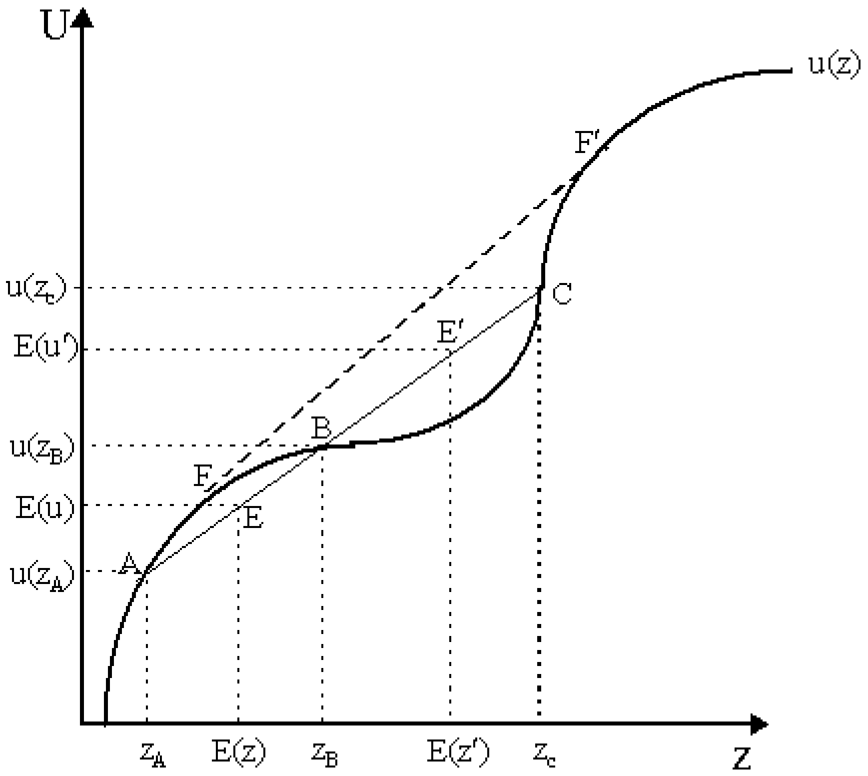
```
),shape=1/weibull.aft$scale), from=0,
to=max(dat$Duration),add=T,col="blue2")

legend (x="topleft", lwd=2,
col=c("black","red","green","yellow",
"blue"), legend=c("pseudo-
observation","Portfolio","DeltaS","Age","Ge
nder"))
```

Plotting estimated hazard function

```
curve(dweibull(x,
scale=exp(coef(weibull.aft)[1]),shape=1/wei
bull.aft$scale)/pweibull(x,scale=exp(coef(we
ibull.aft)[1]),shape=1/weibull.aft$scale,
lower.tail=FALSE), from=0,
to=max(dat$Duration), ylab="Hazard",
xlab="Duration",main="Estimated hazard
function",axes=F, lwd=5, col="black")

axis(1,cex.axis=.8)

axis(2,cex.axis=.8)

box()

curve(dweibull(x,
scale=exp(coef(weibull.aft)[1]+coef(weibull.
aft)[2]),shape=1/weibull.aft$scale)/pweibull
(x,scale=exp(coef(weibull.aft)[1]+coef(weibu
ll.aft)[2]),shape=1/weibull.aft$scale,
lower.tail=FALSE), from=0,
to=max(dat$Duration), add=T, col="red2")

curve(dweibull(x,
scale=exp(coef(weibull.aft)[1]+coef(weibull.
aft)[4]),shape=1/weibull.aft$scale)/pweibull
(x,scale=exp(coef(weibull.aft)[1]+coef(weibu
ll.aft)[4]),shape=1/weibull.aft$scale,
lower.tail=FALSE), from=0,
to=max(dat$Duration), add=T,
col="green2")

curve(dweibull(x,
scale=exp(coef(weibull.aft)[1]+coef(weibull.
aft)[5]),shape=1/weibull.aft$scale)/pweibull
(x,scale=exp(coef(weibull.aft)[1]+coef(weibu
ll.aft)[5]),shape=1/weibull.aft$scale,
lower.tail=FALSE), from=0,
to=max(dat$Duration), add=T,
col="yellow2")

curve(dweibull(x,
scale=exp(coef(weibull.aft)[1]+coef(weibull.
aft)[6]),shape=1/weibull.aft$scale)/pweibull
(x,scale=exp(coef(weibull.aft)[1]+coef(weibu
ll.aft)[6]),shape=1/weibull.aft$scale,
lower.tail=FALSE), from=0,
to=max(dat$Duration), add=T, col="blue2")

legend (x="topleft", lwd=2,
col=c("black","red","green","yellow",
"blue"), legend=c("pseudo-
observation","Portfolio","DeltaS","Age","Ge
nder"))
```

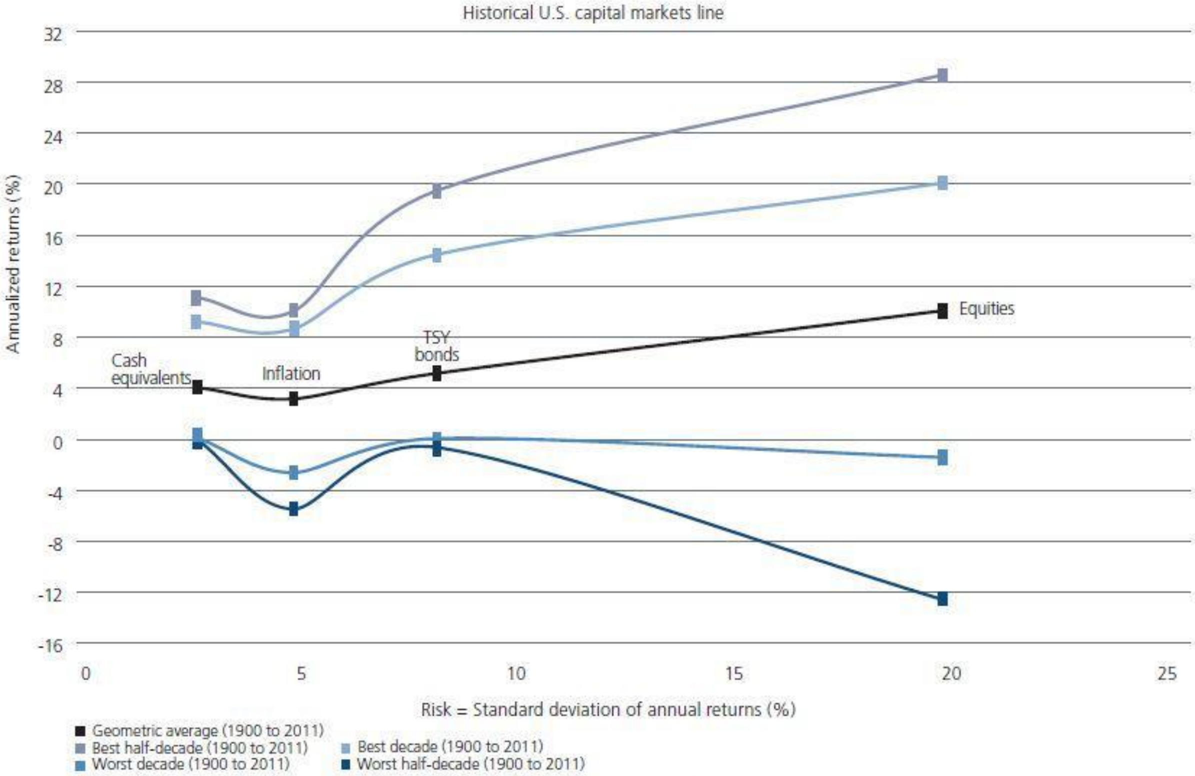
6.5. Annex 5: The loss aversion effect; The investor psychology

http://en.wikipedia.org/wiki/Loss_aversion



6.6. Annex 6: Market evolutions (Ibbotson)

FIGURE 2: HISTORICAL RETURNS FROM EQUITIES ARE HIGH AND VOLATILE



Source: Dimson et al., *Triumph of the Optimists* (2000), Ibbotson Associates (2012), PIMCO calculations
Note: Cash equivalents = 3-month Treasury bills; Inflation = U.S. CPI; Treasury bonds = 10-20 yrs. maturity Treasuries; Equities = S&P 500

The guaranteed rate of an insurance portfolio is averagely set on the return of sovereign bonds yield. This chart supports the hypothesis that the difference Δ between the benchark rate and the served one can not be (in average) be above 10%.

6.7. Annex 7: Ibbotson Index Series, historical returns

Table 1. Ibbotson Index Series: Summary Statistics of Annual Total Return, 1926–2010

Series	Geometric Mean	Arithmetic Mean	Standard Deviation
Large Company Stocks	9.9%	11.9%	20.4%
Small Company Stocks	12.1	16.7	32.6
Long-Term Corporate Bonds	5.9	6.2	8.3
Long-Term Government Bonds	5.5	5.9	9.5
Intermediate-Term Government Bonds	5.4	5.5	5.7
U.S. Treasury Bills	3.6	3.7	3.1
Inflation	3.0	3.1	4.2

Source: Ibbotson® SBBI®, *2011 Classic Yearbook: Market Results for Stocks, Bonds, Bills, and Inflation, 1926–2010* (Chicago: Morningstar, 2011).

6.8. Annex 8: Dynamic lapse rate model, VBA Code

6.8.1. Non guaranteed rate scenario

```
Dim BR, FGR, a, b, g, d, MvMax, MvMin, Dynamic, ContractValue, MeanContractValue, Mu, SD, z, V1, V2 As Double

Dim i, j, ColNum, RowNum, NumRunsAs Integer

Worksheets("Camelea").Range("N4:Q14").ClearContents

Worksheets("Camelea").Range("AD3:AW102").ClearContents

NumRuns = Worksheets("Camelea").Range("B12").Value

FGR = Worksheets("Camelea").Range("B5").Value

FGR = Log(1 + FGR) / 12

a = Worksheets("Camelea").Range("B6").Value

b = Worksheets("Camelea").Range("B7").Value

g = Worksheets("Camelea").Range("B8").Value

d = Worksheets("Camelea").Range("B9").Value

MvMax = Worksheets("Camelea").Range("B10").Value

MvMin = Worksheets("Camelea").Range("B11").Value

For RowNum = 4 To 14

  Mu = Worksheets("Camelea").Cells(RowNum, 13).Value

  For ColNum = 14 To 17

    SD = Worksheets("Camelea").Cells(3, ColNum).Value

MeanContractValue = 0

    For j = 1 ToNumRuns

ContractValue = 100

      For i = 1 To 100

        'Generate normal random variable

        z = 2

        Do While z >= 1

        V1 = 2 * Rnd - 1

        V2 = 2 * Rnd - 1

        z = V1 ^ 2 + V2 ^ 2

        Loop

z = Sqr(-2 * Log(z) / z)

        z = V2 * z

'Simulate Prices


BR = Log(1 + Mu) / 12 + SD * Sqr(1 / 12) * z

        If BR - FGR < a Then

          Dynamic = MvMax

        Else

          If BR - FGR < b Then

            Dynamic = MvMax * (BR - FGR - b) / (a - b)

          Else

            If BR - FGR < g Then

              Dynamic = 0

            Else

              If BR - FGR < d Then

                Dynamic = MvMin * (BR - FGR - g) / (d - g)

              Else

                If BR - FGR > d Then

                  Dynamic = MvMin

                End If

              End If

            End If

          End If

        End If

        Dynamic = Dynamic * (-1)

ContractValue = ContractValue * Exp(BR + Log(1 + Dynamic))

          If j < 21 Then

            If RowNum = 10 Then

              If ColNum = 16 Then

Worksheets("Camelea").Cells(i + 2, j + 29).Value = ContractValue

              End If

            End If

          End If

        Next i

MeanContractValue = MeanContractValue + ContractValue

    Next j


MeanContractValue = MeanContractValue / NumRuns

Worksheets("Camelea").Cells(RowNum, ColNum).Value = MeanContractValue

  Next ColNum

Next RowNum

End Sub
```

6.8.2. Guaranteed rate scenario

```
Dim BR, FGR, a, b, g, d, MvMax, MvMin,
Dynamic, ContractValue,
MeanContractValue, Mu, SD, z, V1, V2 As
Double

Dim i, j, ColNum, RowNum, NumRunsAs
Integer

Worksheets("Garanti").Range("N4:Q14").Cle
arContents

Worksheets("Garanti").Range("AD3:AW102
").ClearContents

NumRuns =
Worksheets("Garanti").Range("B12").Value

FGR =
Worksheets("Garanti").Range("B5").Value

FGR = Log(1 + FGR) / 12

a =
Worksheets("Garanti").Range("B6").Value

b =
Worksheets("Garanti").Range("B7").Value

g =
Worksheets("Garanti").Range("B8").Value

d =
Worksheets("Garanti").Range("B9").Value

MvMax =
Worksheets("Garanti").Range("B10").Value

MvMin =
Worksheets("Garanti").Range("B11").Value

For RowNum = 4 To 14

  Mu =
Worksheets("Garanti").Cells(RowNum,
13).Value

  For ColNum = 14 To 17

    SD = Worksheets("Garanti").Cells(3,
ColNum).Value

MeanContractValue = 0

    For j = 1 ToNumRuns

ContractValue = 100

      For i = 1 To 100

        'Generate normal random variable

        z = 2

        Do While z >= 1

        V1 = 2 * Rnd - 1

        V2 = 2 * Rnd - 1

        z = V1 ^ 2 + V2 ^ 2

        Loop

z = Sqr(-2 * Log(z) / z)

        z = V2 * z

'Simulate Prices

        BR = Log(1 + Mu) / 12 + SD * Sqr(1 /
12) * z

        If BR - FGR < a Then

          Dynamic = MvMax

        Else

          If BR - FGR < b Then

            Dynamic = MvMax * (BR - FGR
- b) / (a - b)

          Else

            If BR - FGR < g Then

              Dynamic = 0

            Else

              If BR - FGR < d Then

                Dynamic = MvMin * (BR -
FGR - g) / (d - g)

              Else

                If BR - FGR > d Then

                  Dynamic = MvMin

                End If

              End If

            End If

          End If

        End If

ContractValue = ContractValue * Exp(Log(1
+ FGR) + Log(1 + Dynamic))

If j < 21 Then

            If RowNum = 10 Then

              If ColNum = 16 Then

Worksheets("Garanti").Cells(i + 2, j +
29).Value = ContractValue

              End If

            End If

          End If


        Next i

MeanContractValue = MeanContractValue +
ContractValue

    Next j

MeanContractValue = MeanContractValue /
NumRuns

Worksheets("Garanti").Cells(RowNum,
ColNum).Value = MeanContractValue

  Next ColNum

Next RowNum

End Sub
```