GENERALIZED LINEAR MODELS IN LIFE INSURANCE: DECREMENTS AND RISK FACTOR ANALYSIS UNDER SOLVENCY II

Rocco Roberto Cerchiara
University of Calabria - Italy
cerchiara@unical.it

Matthew Edwards
Watson Wyatt Ltd
Matthew.Edwards@watsonwyatt.com

Alessandra Gambini
Watson Wyatt Italia Srl
Alessandra.Gambini@watsonwyatt.com

ABSTRACT

The need to accurately understand the factors underlying lapse, surrender and mortality (collectively termed 'decrements' for the purposes of this paper) rates in life insurance is increasing because of (inter alia) the effect of International Accounting Principles, Solvency II and MCEV reporting. The aim of this paper is to investigate the use of Generalized Linear Models to capture empirical dependencies between risk factors and to understand the best factors to allow for in a correct calibration of lapse risk.
After a brief description of theoretical aspects of Generalized Linear Models and their applications in analyzing for risk factors, we have investigated the lapse and surrender experience data of a large Italian bancassurer. The investigation covered the period from 1991 to 2007. The products concerned were life insurance savings policies, with a preponderance of single premium or recurrent single premium products.
The case study results show in the specific case of the chosen Italian company the importance of policy duration, as expected, but also show the sensitivity of lapse rates to calendar year of exposure, to product class and to policyholder age. The work has been carried out by analysing only readily available risk factor data and could hence be substantially improved by a refinement of the underlying database. The results presented in this paper are to be interpreted as being work in progress, to be further refined in the weeks leading up to the AFIR conference.

KEYWORDS

Generalized Linear Model, Poisson Model, risk factors, lapse risk, life insurance

1. Introduction

Insurers will generally want accurate knowledge of recent lapse, surrender and mortality (collectively termed 'decrements' for the purposes of this paper) rates for a number of purposes: primarily, pricing, statutory reserving, Market Consistent Embedded Value estimation, Economic Capital calculations, and standard formulas under the Quantitative Impact Studies (see CEIOPS, 2007, 2008). Also policyholder option take-up rates, such as annuity conversion rates or recurrency patterns of recurrent single premiums are an important element of such analyses. Some of these applications require 'best estimate' and entity-specific parameters, while others require prudent parameters; in both cases, however, it is important to know as accurately as possible the underlying reality.

Entity-specific assumptions are derived by analyzing the historical experience of the insurance enterprise in order to forecast the likely future experience. These results are generally also used to understand whether the past experience is a good indicator of the future, due for example to changed legal / fiscal / economical / financial environment, and future expectations will be assessed accordingly.

Frequent assessment of the adequacy of the analysis are recommended and indeed the MCEV Principles (see Market Consistent Embedded Value Principles, CFO Forum, June 2008) require a review of the analysis at least once a year.

One of the most important elements of entity-specific experience for the Italian market (and, indeed, many other developed insurance markets) is represented by the lapse behaviour of policyholders. Indeed we can observe that the incidence of lapse risk observed in the QIS3 results for the Italian market was about 52%. The nature of this risk can be distinguished in a irrational and a rational component (see FSA, 2007 for an interesting empirical study for UK market, that showed how lapse risk can be affected by factors as distribution channels, bad publicity, etc.). The irrational component is assumed to be underlying basis lapse behaviour, independent from the financial environment. Rational lapse behaviour is represented by the likely increase in the lapse frequency linked by explicit rational behaviour of policyholders following the evolution of the financial markets. It is based on the assumption that policyholders will be able to understand when the value of their lapse option is highest and will behave accordingly. This reactivity of policyholders is likely to depend on a range of factors, such as the type of client: for example institutional clients are obviously likely to be most cognizant of a policy's worth.

The Solvency II draft directive and the CFO Principles explicitly require an investigation of the correlation of policyholder behaviour, such as surrender rates, with other risk factors affecting the insurance entity, such as market risks (QIS4, TS.II.D.11 – 15). Generally lapse behaviour has been investigated as being dependent on product, duration in force and possibly policy generation.

Analyses of life insurance risks has been deeply analyzed in past literature. In the light of recent and expected changes in regulations (IFRS and Solvency II) regarding fair value evaluations (see International Actuarial Association, 2008) there have been several approaches to model stochastic mortality (see for example Ballotta and Habermann, 2006) and overall risk profile of life insurance, taking into account surrender risk and characteristics of Italian life insurance market, as shown in Savelli (1993), De Felice and Moriconi (2002) and Olivieri and Pitacco (2005).

Besides the scientific literature about retention analysis is very wide. It's worth mentioning Albizzati and Geman (1994) and Grosen and Jørgensen (1997, 2000), which represents the basic references for more recent papers. Bacinello et al. (2008) showed an interesting list of bibliography of papers, grouped according to the pricing methodology employed: Binomial or multinomial trees: Vannucci (2003), Bacinello (2005), Costabile et al. (2007); Partial differential equations and free boundary problems: Moore and Young (2005) and Shen and Xu (2005); Least Squares Monte Carlo (LSMC) simulation: Andreatta and Corradin (2003), Baione et al. (2006) and Bacinello et al. (2007).

The aim of this paper is not to price surrender options, but to investigate the use of Generalized Linear Models ("GLMs") in the context of lapse/surrender risk as a means to understand the relationships between risk factors and to calibrate the lapse/surrender risk as accurately as possible.

GLMs provide a powerful statistical analytic tool to investigate the underlying reality of the parameters available in the data in a way that can show how other, previously ignored parameters, such as age or financial conditions, may have affected observed policyholder behaviour.

Furthermore, such analyses allow better understanding of the likely variability in the decrement rates for the purpose of generating relevant sensitivity tests and calibrating internal models in so far as the stochastic modelling of longevity, lapse and mortality risk is concerned.

Finally this paper represents a work in progress devoted to show, using a real-life case study, how GLMs can provide financial services institutions with more accurate estimates of observed decrement experience, and how these analyses can be of particular value in understanding likely future trends and in informing dynamic lapse risk / market risk modeling, an area that is becoming increasingly important in the context of the European Solvency II project.

The paper is structured as follows. In Section 2 we describe the theoretical framework of GLMs. In Section 3, some examples of risk factors analysis with GLMs are presented, underlining the importance of volatility analysis of lapse rates for Internal model calibration. In Section 4 a case study on real data of an Italian life insurance company is developed. Finally Section 5 contains some concluding remarks.

2. Theoretical framework of generalised linear models

Traditional statistical and analytic techniques are of little use in this context, for various reasons. Correlations in the data between factors of interest will make one-way analyses of those factors incorrect. For instance, a traditional analysis based on univariate techniques will not be able to analyse at the same time age and policy duration. Insurers would normally expect to find correlations between the following factors within any one product class: policy duration, age, amount of benefit, sex. Because traditional analyses fail to take due account of the correlation between benefit amount and other factors, the effect of benefit amount on decrement probability will not be measured correctly. The traditional approach of weighting by benefit amount takes some account of amount, but in an incorrect way. In a cash flow projection, whether used in product pricing or embedded value calculations, the main driver is not in fact the lives or policies themselves but the benefit amounts, and it is therefore vital to model correctly the effect of benefit amount: a ‘high benefit amount, low off rates’ projection will give substantially different (and more accurate) results than a projection based on ‘average across amounts’ decrement rates.

Furthermore, traditional analyses are also restricted by concerns about time trends in the data: it is not generally considered ‘safe’ to use more than three-four years of data, especially for the mortality decrement.

GLMs provide a relatively simple and robust way to analyse the effect of many different factors on some observed event. For instance, the effect of policyholder age, benefit amount and duration on decrement rates within a particular product class can be measured. GLMs will take account automatically of all correlations in the data. This powerful aspect of such models entails many advantages: age and duration can be analysed together, giving useful information on a generally disregarded aspect of decrement experience.

The amount of the benefit can be used as a factor in the model, allowing the high/low amount decrement differential to be correctly measured and hence substantially improving the accuracy of cash flow projections.

Calendar year of exposure can be used as a factor, allowing the use of many years of data. This will both improve the robustness of the results, and provide more information about trends that may be of use in informing subsequent thinking about likely trends. This is particularly useful in the context of the mortality decrement, and may also help to investigate the effect of economic conditions on the lapse/surrender decrements.

GLMs also provide the ability to analyse ‘interactions’ between factors: for instance, how the duration effect might differ between high/medium/low benefit amount groups, or between different distribution channels.

For GLMs literature in actuarial field, it is important mentioning McCullagh and Nelder (1989) for a wide overview on GLMs, Fahrmair and Tutz (1996) in relation to unidimensional GLM, Mildenhall S. (1999) for Generalized Minimum Bias Models, Kaas et al. (2001) to use GLMs in Modern Risk Theory for pricing and reserving, Hardin and Hilbe (2001) which continue and extended the work of McCullagh and Nelder.

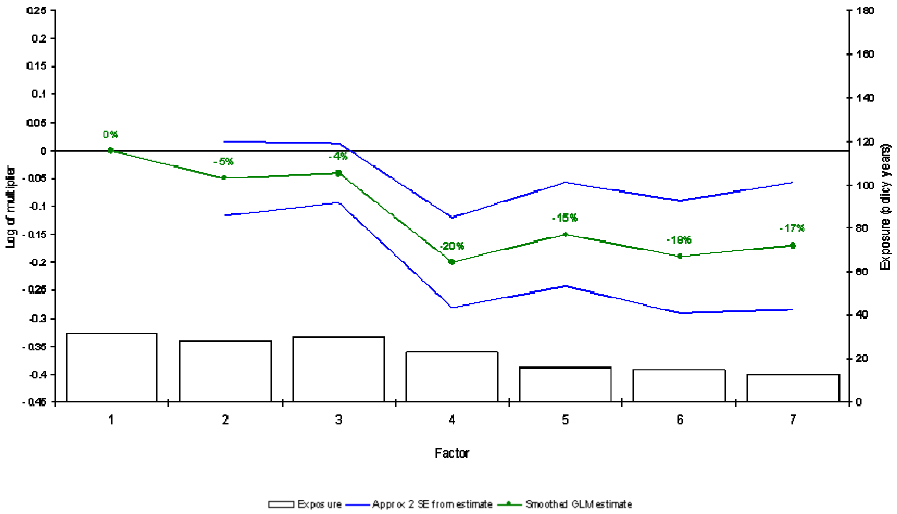
See Watson Wyatt (2007) for a practitioner’s guide to the technical and practical aspects of GLMs as applied in life and non-life insurance fields.

Finally we give a brief description of GLMs, using the framework described in Watson Wyatt (2007). GLM is a method that can model a number as a function of some factors. For instance, a GLM can model motor claim amounts as a function of driver age, car type, no claims discount, etc. and motor claim frequency (as a function of similar factors). Historically it is associated with non-life personal lines pricing (where there was a pressing need for multivariate analysis), but we show in the next sections how this approach could be useful also in life insurance.

A GLM will typically model the ‘observed amount’ (eg motor claims frequency, mortality rate, lapse/surrender rate) as:

$$\text{Amount or frequency} = \text{Base level} \times \text{Factor 1} \times \text{Factor 2} \ldots \quad (1)$$

while taking account automatically of correlations in the data. The GLM output will consist of a ‘base level’ number and, for each factor, a series of multiplicative coefficients showing the relative effect of belonging to a particular level of that factor. In the graph below, the green line (the middle one) shows these relativities for one factor (relative to the 'base' level of the factor, which here is level 1), with 95% confidence intervals in blue (the external lines) and exposure amounts in black at the bottom.



Graph 1: An example of GLM output

A GLM consists of a wide range of models that include linear models (LM) as a special case. The LM restriction assumptions of Normality, constant variance and additivity of effects are removed. Instead, the response variable is assumed to be a member of the exponential family of distributions. Also, the variance is permitted to vary with the mean of the distribution. Finally, the effect of the covariates on the response variable is assumed to be additive on a transformed scale. Thus the assumptions are:

(GLM1) *Random component*: Each component of $\underline{Y}$ is independent and is from one of the exponential family of distributions.

(GLM2) *Systematic component*: The p covariates are combined to give the linear predictor $\underline{\eta}$:

$$\underline{\eta} = X.\underline{\beta} \tag{2}$$

where $\underline{\beta}$ is a vector of parameters to be estimated and X is the matrix of covarietes (data).

(GLM3) *Link function*: The relationship between the random and systematic components is specified via a link function, g, that is differentiable and monotonic such that:

$$E[\underline{Y}] \equiv \underline{\mu} = g^{-1}(\underline{\eta}) \tag{3}$$

Most statistical texts denote the first expression in (GLM3) with $g(x)$ written on the left side of the equation; therefore, the systematic element is generally expressed on the right side as the inverse function, g^{-1}.

The use of a log-link function, typical in much GLM work, results in a simple multiplicative structure (obviously, the exponential transformation turns the additive structure in the underlying linear relationship into a multiplicative relationship): such a link function was assumed in the formula (1) shown above.

It is worth noting that, by virtue of the parameters being estimated as vectors, the values comprising each vector estimate (for instance, for the age vector, the individual coefficients quantifying the observed effect relating to each year of age) are not constrained to any relationship between themselves: to that extent, the GLM modelling procedure can be thought of as essentially non-parametric, as individual factor results are not constrained to any a priori parametric formulation (and, in particular, there is no linearity constraint within factor results).

See Watson Wyatt (2007) for further details on methodology and parameters estimation. In case study of section 4 the “Pretium” Software of Watson Wyatt has been used to generate the GLM results.

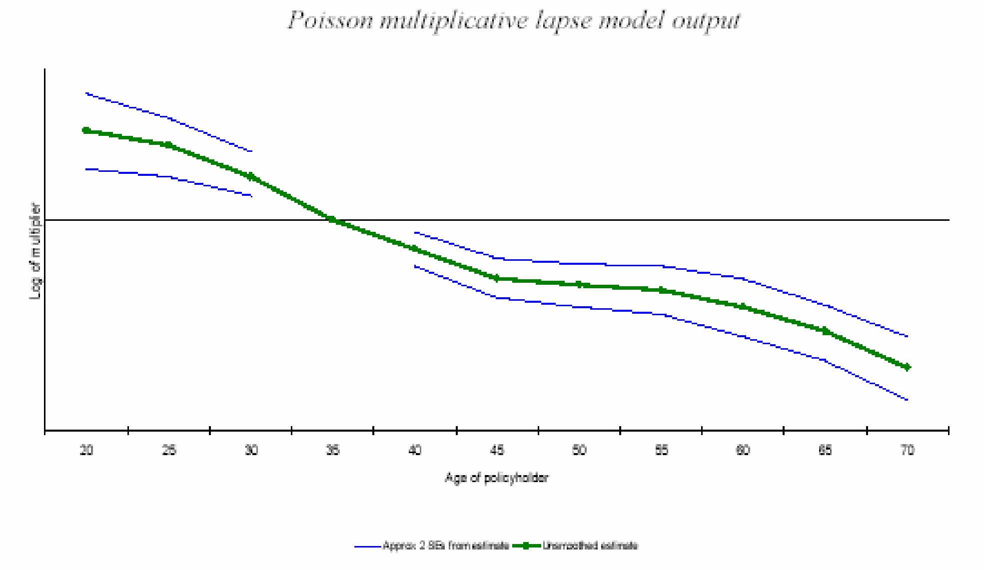
3. Selection of Risk Factors-products

In this section we outline GLM applications to the analysis of various risks, with particular reference to lapse risk (see Watson Wyatt, 2007).

3.1 Lapse analyses

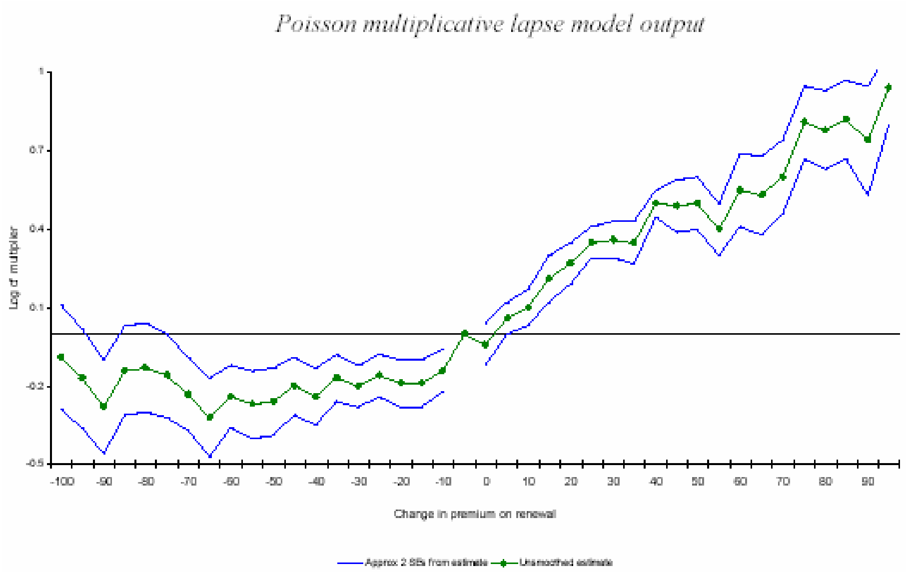
The typical model form for modelling retention (or lapse) and new business conversion is a logit link function and binomial error term (together referred to as a logistic model). The logit link function maps outcomes from the range of (0,1) to ($-\infty$, $+\infty$) and is consequently invariant to measuring successes or failures (in this context, therefore, invariant as regards the choice between measuring lapse probability and measuring retention probability). If the y-variate being modelled is generally close to zero, and if the results of a model are going to be used qualitatively rather than quantitatively, it is possible to use a multiplicative Poisson model form as an approximation given that the model output from a multiplicative GLM can be easier to explain to a non-technical audience.

The graph below shows sample GLM output for a lapse model in the context of non-life personal lines business. The green line on the graph demonstrates (on a log scale) the measured multiplicative effect of age of policyholder upon lapse rate. The effect is measured relative to an arbitrarily selected base level, and the results take into account the effect of all other factors analyzed by the GLM.



Graph 2: An example of GLM output for a lapse model

In this example, which is fairly typical for such business, it can be seen that young policyholders lapse considerably more than older policyholders, perhaps as a result of having more time and enthusiasm in searching for a better quotation, and perhaps also as a result of being generally less wealthy and therefore more interested in finding a competitive price.
This next graph shows the effect of premium change on lapse rate. This GLM output is from a UK Institute of Actuaries General Insurance Research Organisation (GIRO) study (see Bland, 1997) based on around 250,000 policies across several major UK insurers in 1996. The premium change is measured in ranges of monetary units (British pounds in this case), but the model could easily be based on the percentage change in premium. As would be expected, increases in premium increase lapses. The model, however, quantifies this accurately and enables investigations into potentially optimal rate increases to be undertaken. It can be seen in this case (as is often the case) that decreases in premium beyond a small threshold do not decrease lapses.

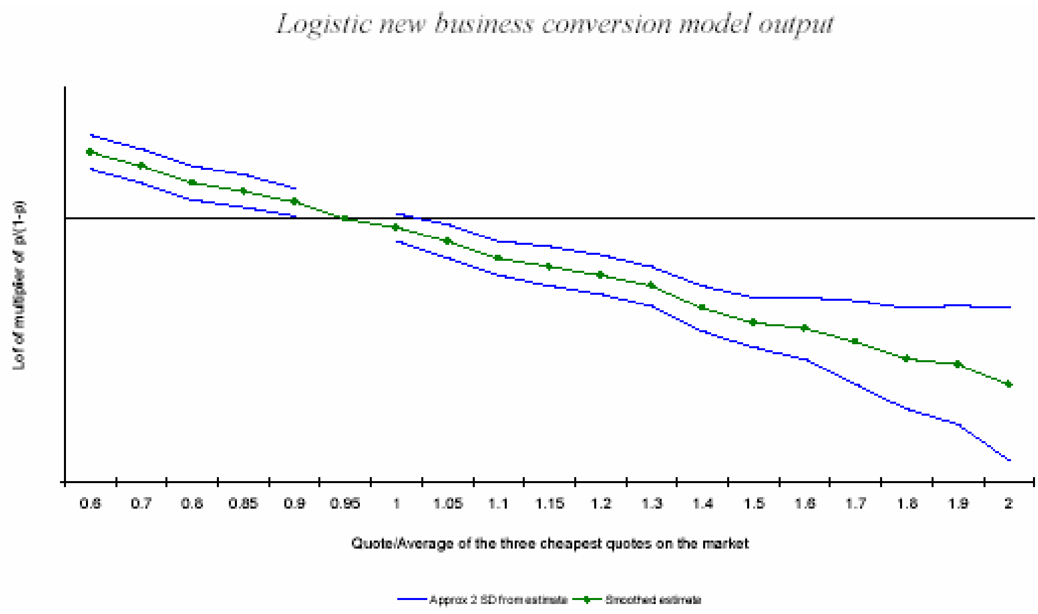


Graph 3: The effect of premium change on lapse rate

Measures of premium change should ideally consider whether customers have an inherent expectation of premium change. For example, customers with recent claims will anticipate a premium increase and may be prepared to accept their renewal offer rather than face the underwriting guides of a new company. Conversely, customers who are rolling off an accident surcharge, hitting a milestone age or a change in marital status may expect a decrease. A possible proxy for customer expectation is to adjust the premium change variable to be the ratio of proposed premium (based on new risk criteria and new rates) to adjusted proposed premium (new risk criteria based on last year's rates).

In addition to including premium change, absolute premium can also be considered as a factor in a model. This approach, though not theoretically incorrect, may make the model difficult to interpret since many other factors in the model will be a component of premium and therefore highly correlated with premium size. Adding absolute premium to the model may significantly alter the observed relativities for other factors which may make the results hard to interpret. One alternative to including absolute premium in such a case is to fit separate models for different bands of average premium.

The next graph below shows an example of the effect of competitiveness in a new business conversion model. The measure of competitiveness used in this case is the ratio of the proposed premium to the average of the three cheapest alterative premiums from a selection of alternative insurers. It can be seen that the less competitive the premium, the lower the conversion rate.



Graph 4: An example of the effect of competitiveness in a new business conversion model

A further analysis which can be undertaken is to superimpose the results of two models on one graph: one model that includes the competitiveness measure and one model that does not. The disparity between these two models will show how much of a factor's effects are simply price-related.

Retention analyses can also lead to operational actions which are unrelated to price. For example, in a highly rate-regulated state, consideration could be given to which segments of the population (given a restricted set of rates) are both profitable and most likely to renew in the future. Such a measure could help form new underwriting guides or targeted marketing and cross-sell campaigns.

Insurance expense analysis is another field of study that is often over-shadowed by loss analysis. If acquisition expenses are higher than renewal expenses then an understanding of likely retention (and therefore expected life of a policy) can be used to amortize the higher acquisition cost over the expected life of the policy.

It should be noted that the examples presented above relate to non-life personal lines; we discuss life insurance lapses in the case study described in chapter 4.

3.2 Longevity risk

An area where GLM investigations are very useful is in the analysis of mortality in respect of annuitants. Although the advantages of GLMs make them suitable for analyzing the mortality of any insurance class, attention has generally been on their application to annuitant mortality because (i) the increased mortality rates at typical annuitant/pension ages mitigate the high data requirements to some extent, and (ii) annuitant mortality has been an area of particular concern in many countries in recent years, while assured lives' mortality is generally regarded as reasonably well understood and as providing less likelihood of financially significant changes in the future.

A particular concern in relation to the mortality of annuitants/pensioners is the future improvement rate. GLM investigations with sufficient years of experience allow us to quantify the historical calendar year of exposure effect, hence informing the analyst's thinking about suitable future mortality improvement assumptions. By introducing interactions into the GLM, for instance interactions between annuity amount band and calendar year of exposure, or between sex and year of exposure, such trends can be considered in more detail than might otherwise be feasible. For very large portfolios, analysis of the year of birth effect can also be of interest, although care needs to be taken to correctly isolate the age effect before introducing year of birth into the model.

The risk factors generally considered in annuitant/pension mortality analyses are, in addition to age and sex, the following:

- Duration
- Calendar year of exposure
- Amount of benefit
- Geography, or other socio-economic measure
- Medical information (eg body-mass index)
- Early or late retirement
- Ill- health or normal retirement

The **new gender directive** of the EU will lead to much more importance being placed on the statistical justification of any risk factor used for pricing, in particular gender itself.

3.3. Other uses in Risk Management – Volatility analysis

A GLM analysis of lapse rates can assist in better understanding the natural volatility of lapse rates. The practical applications of such an analysis can manifold in risk management for example:

- Setting of distribution of parameters of internal risk models
- Understanding qualitatively and quantitatively the dependencies of the rational lapse policyholder behaviour modelled in the calculation of the life technical provision
- Setting appropriate liquidity levels in the financial management of asset underlying Italian traditional revalorization policies and setting corresponding asset allocation limits.

This volatility can be understood as being caused by four factors:

- Random fluctuations;
- Rational economic decisions by policyholders;
- Other decisions by policyholders linked to policyholder characteristics;
- Other 'completely unexplained' decisions by policyholders.

The random fluctuation aspect, whereby a small portfolio will have an apparently more volatile experience than a large portfolio, is of little conceptual or statistical interest, although it is an aspect that insurers (especially small insurers) should take due account of in capital calculations.

Regarding policyholders' rational economic decisions, there are two ways to use GLMs to improve understanding:

- By correctly identifying and quantifying the 'calendar year of exposure' effect in a GLM lapse analysis, it is possible to investigate relationships between observed calendar year trends and trends in the relevant financial markets. If such a relationship can be determined, the analyst will then be in a position to quantify the effect on surrenders if the relevant market measure (eg ten-year bond yields) moves by a particular amount. Since well-calibrated economic scenario generators can provide reasonably reliable market movements associated with particular probabilities (for instance, the 1-in-200 worst event used in some economic capital regimes), we can calibrate the surrender effect associated with the probability tail of interest.
- By introducing as a factor in the GLM a measure quantifying (probably approximately) the difference between the policy's surrender value and an estimated 'fair value' of the policy, calculated with reference to the yield curves relating to the particular year of exposure in question, the analyst can quantify to what extent this difference has influenced surrender rates. This information can then be used to estimate the surrender experience that would be expected under particular future economic scenarios.
- These two approaches are not necessarily mutually exclusive: it may be the case that having both of the factors described above provides a more predictive model than just one of the factors. This is likely to be true in particular if the model is attempting to treat substantially different product classes (eg protection policies, with no surrender values).

Regarding the third factor affecting lapse volatility, other policyholder decisions linked to policyholder characteristics: in effect, these are decisions for which there is some proxy available in the data (for instance, policyholder age). The effect of this element can be gauged by comparing the predictiveness of a lapse model taking account of these factors with the predictiveness of a model without the factors. Such tests of predictiveness, to be done rigorously, need to be done after splitting the data into a modeling portion (eg 70-80% of the data) and a testing portion (the remaining 20-30%).

Regarding the fourth effect, this can be regarded as representing all influences on policyholder behaviour not described by the available factors, and over and beyond that which is reasonably attributable to random fluctuations. The presence of such effects can be seen to some extent by consideration of model diagnostics such as the distribution of residuals (in effect, the 'observed – predicted' amounts in respect of every data point), which can indicate hidden heterogeneity in the experience.

4. The case study

We have investigated the lapse and surrender experience data of a large Italian bancassurer. The investigation covered the period from 1991 to 2007. The products concerned were life insurance savings policies, with a preponderance of single premiums or recurrent single premium products. The data covered a total of 279,000 lapses/surrenders and 6,129,000 policy-years of exposure.

The risk factors available for our analysis were:

- Product
- Year of event (obviously equal to year of exposure)
- Duration
- Year of policy inception.

We also had a separate dataset supplied by the same company in a different format, which allowed the analysis of policyholder age and sex but not product class. Some doubts about the consistency of these two datasets led us to concentrate on that which seemed more robust (the data described above), and not to use this second dataset containing age and sex information. However, for interest, we comment below on the observed age and sex effects as they are of interest.

For the sake of easy illustration, given that the context of this modeling was to provide indicative results and not to provide necessarily accurate predictive estimates, we have modelled the lapse/surrender experience using a Poisson probability assumption and a log-link function. Because the dataset was provided to some degree 'ready grouped', rather than policy-by-policy, we have had to conduct this analysis without being able to:

- Use a completely rigorous method to categorize products into reasonably homogeneous product groups; rather, our method has had to involve using the same data both to decide on the composition of these groups and to quantify the lapse/surrender effect associated with those groups. This procedure will have exaggerated the apparent product lapse/surrender effect.

- Split the data into independent modeling and testing portions. Typically, we would prefer to test the predictiveness of our model by quantifying all model factors on a large modeling portion of the data (70-80%), and then testing the predictiveness of that model on the remaining 'testing' portion (20-30%).

In a "real world" analysis the methodology would be refined to solve the above mentioned limits as well as with reference to additional risk factors where available, such as client segment, sum insured, geographical distribution, gender, age etc, which have not been included owing to time constraints at the data-preparation stage. Of these non-modelled factors, a factor in respect of sum insured (or some other indicator of policyholder wealth, such as premiums or reserves) would be likely to be of great significance.

Initial model runs using all available factors at an extreme level of detail led us to remove year of policy inception from the model (obviously, the three factors of exposure year, duration and policy inception year describe two elements of information, and therefore one of the factors is largely redundant). We also regrouped the large number of products into a smaller number of product classes with similar observed lapse/surrender levels. As noted above, this was done in an approximate manner. This product grouping was refined using a second iteration of the model, leading to the derivation of a reasonably robust third iteration.

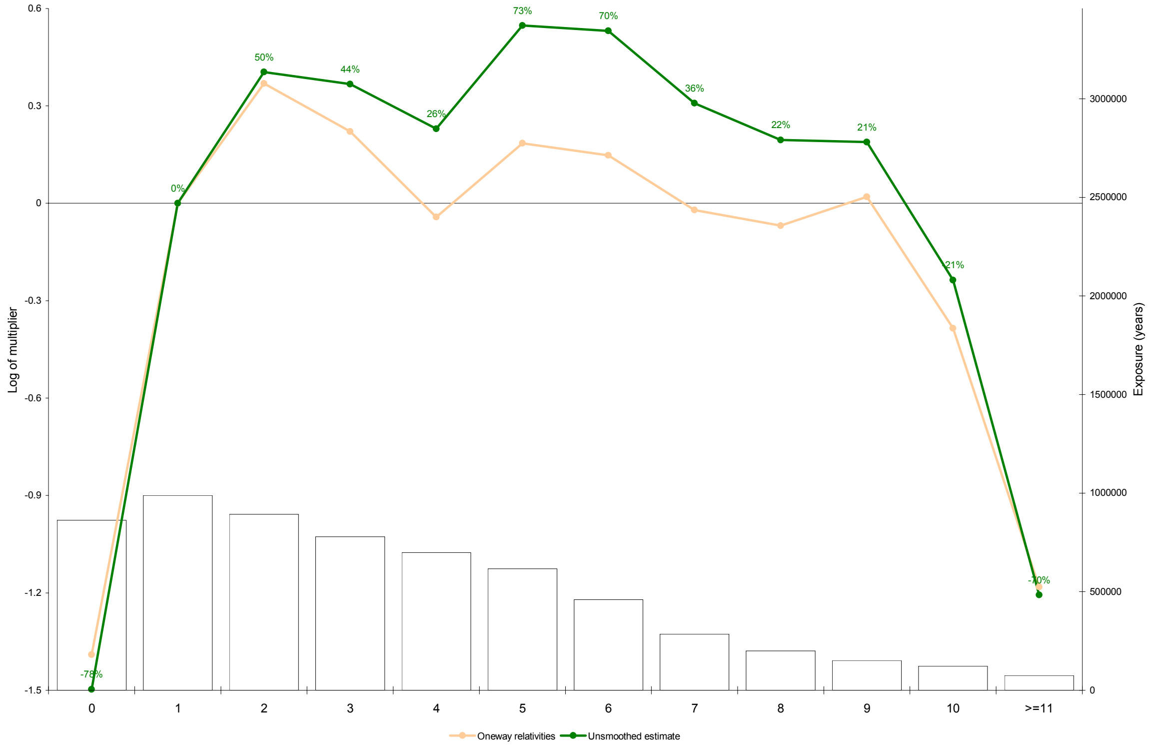
This third iteration model gave us results as illustrated graphically below. By way of introduction, these graphs should be interpreted as follows:

- The graphs show the relative effect of belonging to different levels of a particular factor (eg the relative effect of a particular calendar year of exposure), in addition to the effect of other factors (eg product) in the model but not shown in the graph in question.
- For each band, the black bars (at bottom) show exposure (quantified on right-hand axis)
- The GLM results are in green, and are relative to the base level for the factor. By definition, the base level result is seen as 0%. A result of, for instance, 50% would mean that the lapse/surrender probability for that level is 150% of the lapse/surrender of 'base level' policies, taking account of all other factor results in the model.
- The blue lines give 95% confidence interval either side (although these are shown only in the later graph relating to policyholder age, where the policy-by-policy nature of the data has allowed correct confidence interval calculations).
- The yellow line is what would have been generated by a 'one-way' analysis: ie considering just the factor in question, without any other factors. Thus, the yellow line on the duration graph shows how lapse/surrender experience would typically be understood. The difference between this line and the green GLM line, which better represents the underlying reality, arises from correlation between the factor shown in the graph and all other factors in the model.

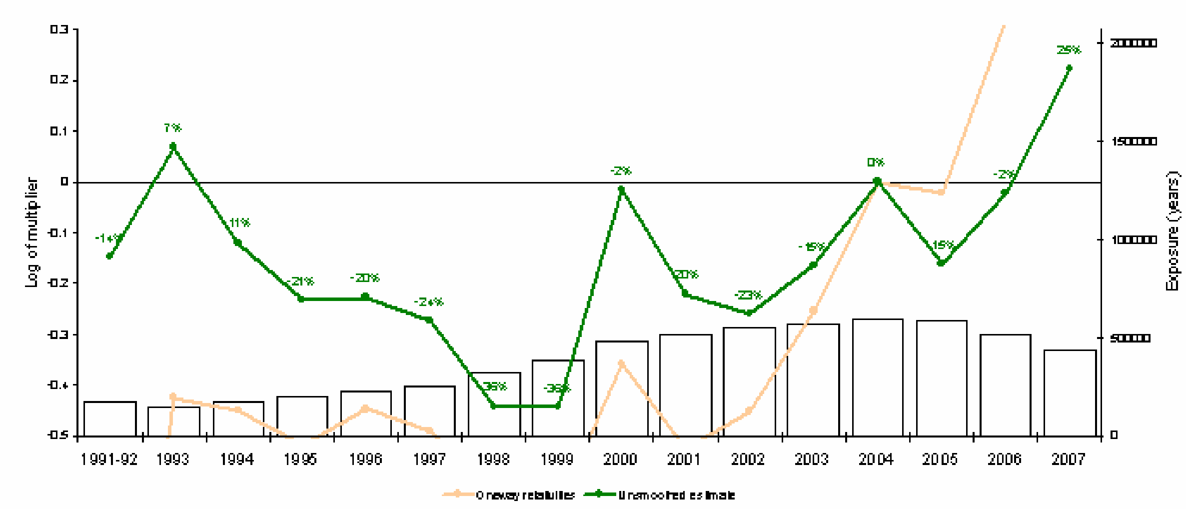
This model is predicting lapse/surrender rates according to the following formula:

> Probability of lapse/surrender = 3.1% × multiplicative factor relating to calendar year of exposure × multiplicative factor relating to policy duration × multiplicative factor relating to product class.

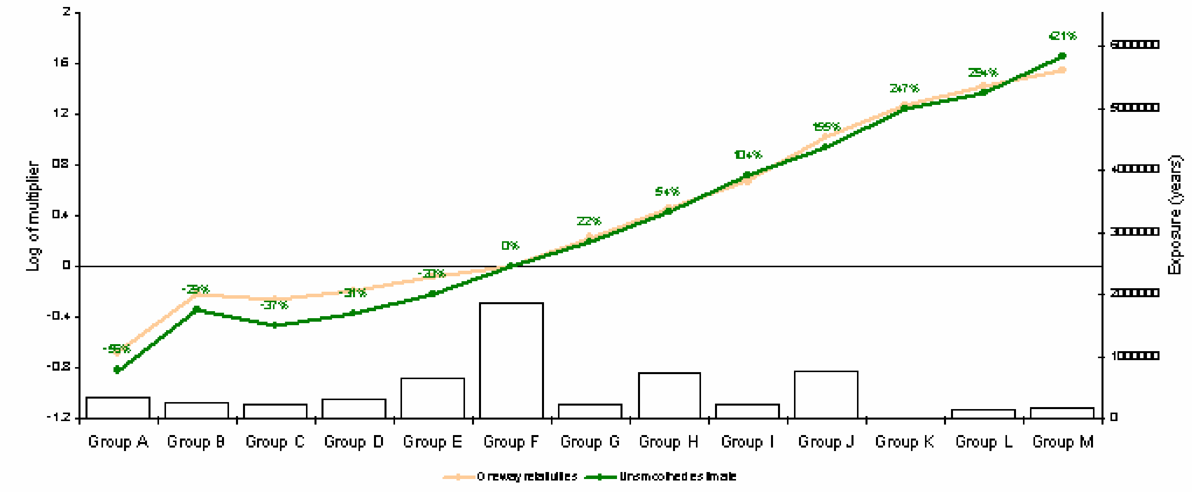
The graphs below show the values in respect of these three multiplicative factors.



Graph 5: Effect of policy duration



Graph 6: Effect of calendar year of exposure



Graph 7: Effect of product group

The graphs clearly show the very strong influence of policyholder duration, as expected, with the peaks at two and five years being consistent with what would be expected for policy condition and potentially (particularly considering elder inception generations) tax reasons

Of interest also are the following points regarding policy duration:

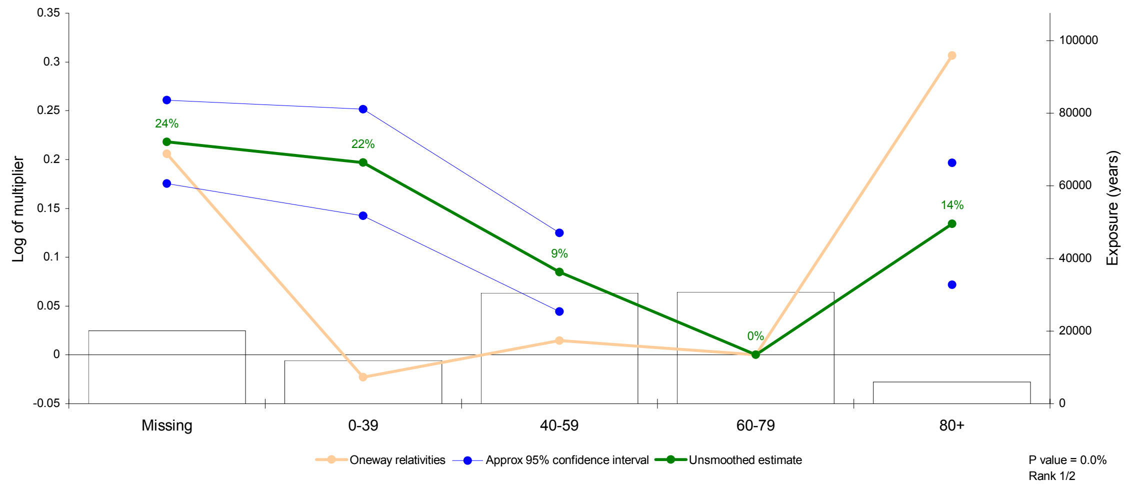
- The very low experience beyond year 10, which might be interpreted as a form of automatic policyholder selection: those policyholders who have not surrendered in the first ten years seem particularly unlikely to countenance surrender thereafter. There are likely also to be product considerations in this effect.
- The one-way analysis (seen in yellow) gives substantially different results in years 3-9. This might be considered a critical period as regards profitability, since it is generally over such a period that companies seek to recoup the initial expenses in excess of initial loadings. Misunderstanding of the surrender probabilities in this duration interval could therefore impinge substantially on product profitability.

The calendar year of exposure factor has a large effect on surrender probabilities. It is possible to define the 'explanatory power' of any factor as being the extent to which moving from one extreme to the other affects the modelled quantity. In the case of this factor, the 1999 low of -37% and the 2007 high of +26% translate into an 'explanatory power' of 1.26/0.63 = 2.

For example the high 2007 effect is consistent with a generally increased incidence of surrenders in 2007 of savings policies in the Italian life insurance market. In particular capitalization policies, which had been acquired in past years by institutional investors based on the relatively high level of unrealised capital gains present in Italian "gestioni separate", whilst the yields available on the market where low. In a market situation where interest rates have been picking up again and past reserves of unrealised capital gains have been gradually used up, these policies will have been surrendered at a greater rate than in previous years.

The graph 7, in respect of product group, shows a very strong effect. Calculation of the explanatory power of this factor, analogous to the calculation done in respect of the calendar year of exposure factor, gives a result of 5.21/0.54 = 9.65. However, this statistic is misleading in two respects. First, we might have chosen a larger or smaller number of product groups and hence arrived at a quite different range. Secondly, our categorization of products into product groups with reference to their observed lapse/surrender characteristics in an initial model used the same data then used to quantify the effect of belonging to the eventual product groups: this will have exaggerated the observed product effect.

We have also considered the effect of policyholder age, as shown in the graph below. This is shown more for 'general interest', as for data reasons the policyholder age results were derived from a dataset different from that underlying the above results. For reference, the model from which the results shown graphically below were produced included the factors policy duration and calendar year of exposure, but not product group (which was not available in the dataset used for this separate investigation).

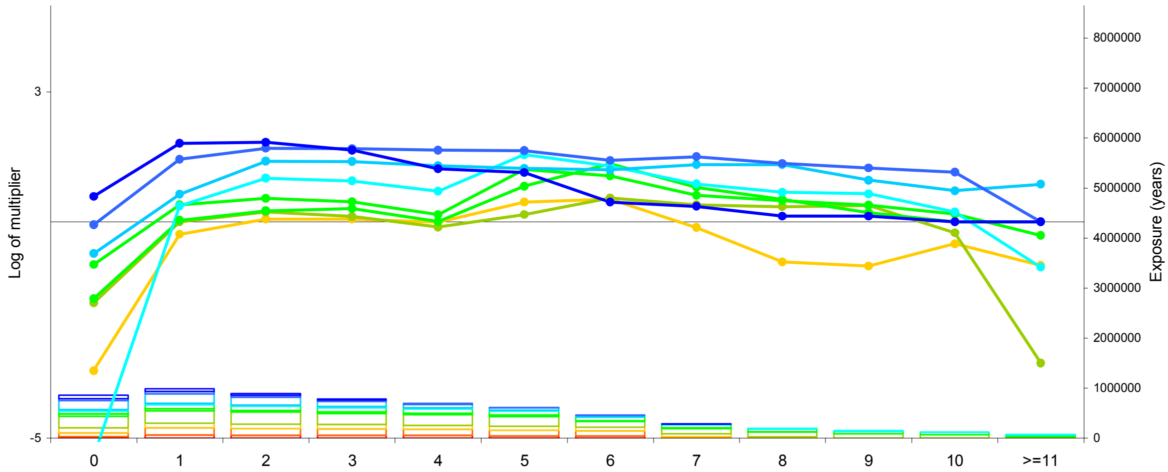


Graph 8: Effect of policyholder age

For reference, it is also worth noting that the 'policyholder age' dataset also included policyholder sex. We included policyholder sex as a factor in the GLM run on this dataset, but this factor proved to be insignificant (ie there was no statistically significant effect of sex on policyholder lapse/surrender behaviour).

Having conducted initial modeling on the data, it is then interesting to seek further insights via more detailed follow-up investigations. For this portfolio, we investigated the interaction between product class and (i) duration and (ii) calendar year of exposure. Setting up an interaction in the model allows the GLM to calculate a series of (for example) duration coefficients for each product class.

The interaction between duration and product class gave the following results:

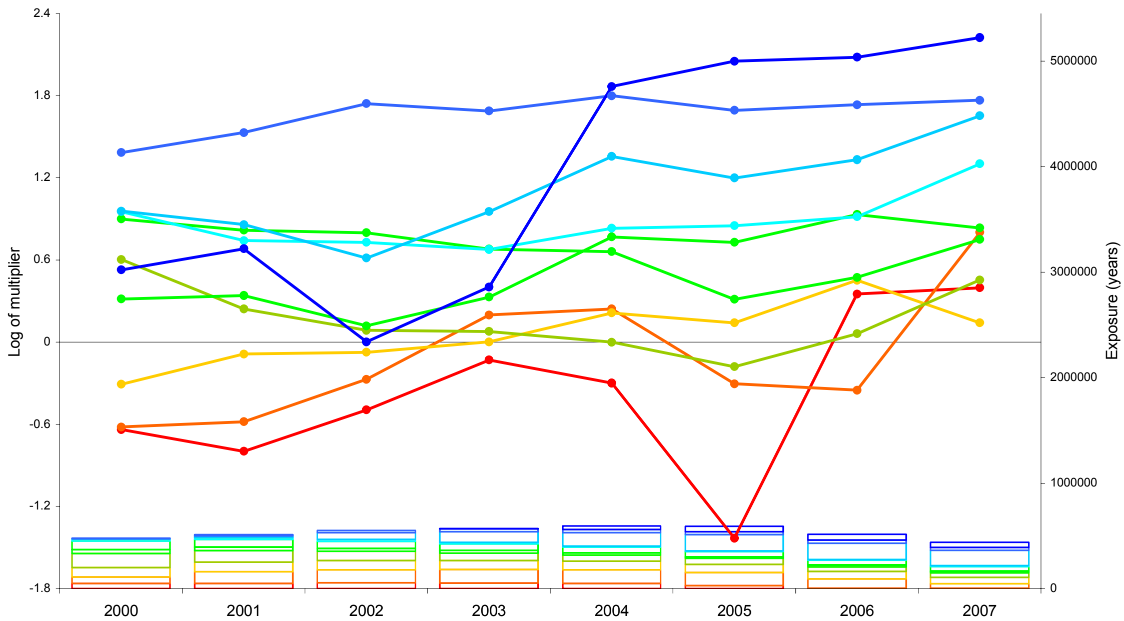


Graph 9: Interaction between duration and product group

The above graph shows the duration 'curve' separately for each product group. In the interests of legibility, we have removed the labels from these curves.

This graph shows a surprising similarity in duration effects (other than in overall magnitude) between the product groups. However, this similarity may result from the way in which the product groups were defined in our exercise (by grouping products with similar overall levels of lapse/surrender behaviour according to the initial GLM). If the product groups had been created with reference to their qualitative characteristics (in particular, type or level of guarantee) then we might have seen much greater differences between the duration effects associated with each product group.

We also modelled an interaction between product group and calendar year of exposure, restricting the scope of this investigation to the years 2000 onwards (as initial 'all years' results were meaningless owing to the large number of products not present before 2000). This interaction is shown in the following graph:



Graph 10: Interaction between calendar year of exposure and product group

This graph shows the calendar year of exposure 'curve' for the various product classes. With a couple of exceptions (which consideration of the underlying data show unlikely to relate to random fluctuations), these are generally similar (other than, of course, in overall magnitude).

5. Conclusions

We propose a GLM approach, typically used in general insurance, to investigate the risk factors underlying the underwriting risk of life insurance. Furthermore, such analyses allow better understanding of the likely variability in the decrement rates for the purpose of generating relevant sensitivity tests and calibrating internal models in so far as the stochastic modeling of lapse risk (and other factors too).

This paper represents a work in progress devoted to better understanding, using data of an Italian life insurance company, how GLMs could improve the accuracy of parameter future trends estimations and in informing dynamic lapse risk / market risk modeling, an area that is becoming increasingly important in the context of the European Solvency II project.

The case study results reaffirm the importance of policy duration, as expected, but also show the sensitivity of lapse rates to calendar year of exposure, to product class, and to policyholder age. This particular case study showed that the duration and calendar year of exposure effects were relatively insensitive to product class, although this may not be a conclusion which would be replicated in other companies' experience.

References

Albizzati M. O., Geman H. (1994): Interest rate risk management and valuation of the surrender option in life insurance policies. The Journal of Risk and Insurance. Vol. 61, pp. 616–637

Bacinello A.R. (2005): Endogenous Model of Surrender Conditions in Equity-Linked Life Insurance, Insurance: Mathematics and Economics. Vol. 37, 270-296

Bacinello A. R., Biffis E. and Millossovich P. (2007). Pricing life insurance contracts with early exercise features. Working paper, Tanaka Business School, Imperial College London

Bacinello A.R., Biffis E., Millossovich P. (2008): Regression-Based Algorithms for Life Insurance contracts with Surrender Guarantees, Proceedings of MAF, Venice, Italy

Baione F., De Angelis P., Fortunati A. (2006): On a fair value model for participating life insurance policies. Investment Management and Financial Innovations. Vol. 3, pp.105–115

Ballotta L. and Haberman S. (2006): The fair valuation problem of guaranteed annuity options: The stochastic mortality environment case. Insurance: Mathematics and Economics, Elsevier, vol. 38(1), pages 195-214

Bland R.H. et al. (1997): Institute of Actuaries GIRO Customer Selection and Retention Working Party, ISBN 0 901066 45 1

CFO Forum (June 2008), Market Consistent Embedded Value Principles

CEIOPS (2007): QIS3 - Technical Specifications

CEIOPS (2008): QIS4 - Technical Specifications

Costabile M, Massabò I., Russo E. (2007): A binomial model for valuing equity-linked policies embedding surrender options.. To appear on Insurance: Mathematics & Economics
De Felice M., Moriconi F. (2002): Finanza dell'assicurazione sulla vita. Principi per l'asset-liability management e per la misurazione dell'embedded value. Giornale dell'Istituto Italiano degli Attuari, LXV - n. 1-2, Rome, Italy
European Commission (2007): Proposta di Direttiva del Parlamento Europeo e del Consiglio relativa all'accesso alle attività di assicurazione e di riassicurazione e al loro esercizio
Fahrmair L., Tutz G. (1996): Multivariate statistical modelling based on generalized linear models, Springer Verlag, New York
FSA (2007): Survey of the persistency of life and pension policies. Financial Services Authority, London
Grosen A., Jørgensen P. L. (1997): Valuation of early exercisable interest rate guarantees. The Journal of Risk and Insurance, Vol. 64, 481–503
Grosen A., Jørgensen P. L. (2000): Fair valuation of life insurance liabilities: the impact of interest rate guarantees, surrender options, and bonus policies. Insurance: Mathematics & Economics. Vol. 26, pp. 37–57
Hardin J., Hilbe J. (2001): Generalized Linear Models and Extensions, Stata Press
International Actuarial Association (2008): Second Exposure Draft of International Actuarial Standard of Practice – Disclosure of Information about Insurance Risk under IFRS A Practice Guideline
Kaas R., Goovaerts M., Dhaene J., Denuit M. (2001): Modern actuarial risk theory, Kluwer Academic Publishers, Boston
McCullagh P., Nelder J.A. (1989): Generalized linear models, Second edition, Chapman and Hall, New York
Mildenhall S. (1999): A systematic relationship between minimum bias and generalized linear models, Proceedings of the Casualty Actuarial Society, LXXXVI
Moore, K. S. and V. R. Young (2005): Optimal design of a perpetual equity-indexed annuity. North American Actuarial Journal , vol. 9: 57–72
Olivieri A., Pitacco E. (2005): La valutazione nelle assicurazioni vita – Profili attuariali, Egea, Milan, Italy
Savelli N. (1993): Un modello di teoria del rischio, Collana ASA, Lint, Trieste
Shen W. and H. Xu (2005): The valuation of unit-linked policies with or without surrender options. Insurance: Mathematics and Economics, vol. 36: 79–92
Vannucci E. (2003): An evaluation of the riskiness of unit linked policies with minimal return guarantees. In Proceedings of the 6th Spanish-Italian Meeting on Financial Mathematics, pp. 569–582.
Watson Wyatt (2007): A Practitioner 's Guide to Generalized Linear Models