A two-part Beta regression approach for modelling surrenders and withdrawals in a life insurance portfolio

Fabio Baione[a], Davide Biancalana[b], Paolo De Angelis[c]

[a] *Department of Statistics, Sapienza University of Rome, Italy*
[b] *Sapienza University of Rome, Italy*
[c] *Department of Methods and Models for Economy Finance and Territory. Sapienza University of Rome, Italy*

Abstract

Beta regression is a flexible tool in modelling proportions and rates, but scarcely applied in actuarial field. In this paper, we propose its application in the context of policyholder behaviour and particularly to model surrenders and withdrawals. Surrender implies the expiration of the contract and denotes the payment of the surrender value, that is contractually defined. Withdrawal does not imply the termination of the contract and denotes the payment of a cash amount, left to the discretion of the policyholder, within the limits of the surrender value.
Moreover, the Actuarial Standard of Practice 52 states that, for surrender and withdrawal estimation, the actuary should take into account several risk factors that could influence the phenomenon. To this aim, we introduce a Two-Part Beta regression model, where the first part consists in the estimate of the number of surrenders and withdrawals by means of a Multinomial regression, as an extension of the Logistic regression model frequently used in the empirical literature just to estimate surrender.
Then, considering the uncertainty on the amount withdrawn, we express it as a proportion of surrender value; in this way, it assumes values continuously in the interval $(0, 1)$ and it is compliant with a Beta distribution. Therefore, in the second part, we propose the adoption of a Beta regression approach to model the proportion withdrawn of the surrender value.
Our final goal is to apply our model on a real life insurance portfolio providing

Email addresses: `fabio.baione@uniroma1.it` (Fabio Baione), `davide.biancalana@uniroma1.it` (Davide Biancalana), `paolo.deangelis@uniroma.it` (Paolo De Angelis)

the estimates of the number of surrenders and withdrawals as well as the corresponding cash amount for each risk class considered.

Keywords: Beta regression, Lapse, Multinomial regression, Policyholder behaviour, Surrender, Withdrawal.

1. Introduction

The interest in the analysis of the policyholder behaviour in the actuarial and risk management fields has dramatically increased over the last decade. Among the main reasons, there is the need to raise the accuracy of actuarial models both for regulatory purposes and for the greater awareness in the underwriting and management phases of insurance products.
Among others, traditional regression methods are commonly used by practitioners when the behaviour of the response variable is described as a function of other exogenous variables. Regression models adopted in actuarial life and non-life applications are Generalized Linear Models (see De Jong and Heller [2008]) or their extensions as Generalized Additive models, Hierarchical Generalized Linear Model, to mention the most used; one of the main motivations of such an extensive application is that the distribution of the response variable is a member of the exponential family, usually suitable for describing numbers or amounts linked to insurance events. However, it is interesting to highlight that in actuarial applications data may also be collected in the form of fractions, rates or proportions, continuously assuming value in the interval $(0,1)$. By means of example, one can refer to the portion withdrawn of the net cash value in a life insurance or pension plan, or the portion of loss or damage related to the property's value in a personal property coverage. In such cases, exponential distributions are not suitable for modelling the data and the assumption of a Beta distribution of the response variable is more appropriate.
To this aim, in the following paper, we introduce a methodological framework in the assessment of the likelihood that a policyholder will exercise the surrender and/or the withdrawal contractual options in a life insurance product looking at both frequency and severity associated with these events. The term surrender (or full withdrawal) means that policyholder voluntarily lapses his/her policy and early terminates the insurance contract to access to the cash surrender value. Moreover, many products (e.g. Unit Linked, Whole Life Variable Annuities, Universal life) allow the policyholder to take withdrawals (in some countries called partial surrender) from his/her policy. A withdrawal reduces, but does not set to

zero, the face amount of the contract and the surrender value as well. However, it is noteworthy that withdrawals do not involve an early termination of the contract and imply that only a portion of the cash value is depleted from the total cash value; therefore, the withdrawal amount can be measured as a proportion of the cash value being included in the interval $(0, 1)$.

The simultaneous existence of a surrender option and a withdrawal option in a life insurance contract significantly affects the estimate of a contract's future cash flow and, as far as we know, has never been discussed until now. An under/overestimate of the number of surrenders that early terminate the contract as well as of the amount surrendered or withdrawn may cause several effects: a change in the insurer's best estimate liabilities, an impact on the asset-liability strategy, a loss of future profits or the increase of losses in case of high financial guarantees.

The pricing of surrender options is mostly based on the assumption of a full rational behaviour of the policyholder, who optimally acts to maximize the terminal value of the policy and the use of financial mathematics techniques (see Bacinello [2003], Bacinello [2005]).

However, as real evidence contradicts rational behaviour, empirical research on the lapsation investigates irrational behaviour looking for the factors that influence the surrender behaviour of life insurance policyholder. The relationship between lapses and economic environment or insurance policy characteristics is generally performed by means of statistical approaches such as Generalized Linear Model (Haberman and Renshaw [1996], Kim [2005], Cerchiara [2009], Eling and Kiesenbauer [2014], Baione et al. [2021]), Classification tree (Milhaud et al. [2010]), and more recently Machine Learning (Aleandri [2018]). All the mentioned papers concern the estimate of the occurrence or not of surrenders but no one discusses about withdrawals simultaneously or alone, neither on the corresponding withdrawal amount.

In order to investigate a general framework for the estimation of surrender and withdrawal rates and the severity of the cash value withdrawn, we consider a two-part model (Duan et al. [1983]) to decompose the expected benefits between surrenders and withdrawals. Two-part models in actuarial sciences are often used for non-life premium rate-making or claims reserving and are usually called frequency-severity models. To this aim, we consider two multivariate regression models. The first one is a Multinomial regression (henceforth also MR) model used to estimate the probability of surrender or withdrawal; a basic theory of Multinomial model is described in Venables and Ripley [2003]. Hence, if a withdrawal event occurs, we use Beta regression (hereafter also BR) (see Cribari-Neto and Ferrari [2004]) to model the fraction of the face amount withdrawn.

BR defines a multivariate model, where the response variable ranges from 0 to 1, hence it is suitable to our aim.

In the current literature, BR has been applied in multiple fields of knowledge (e.g. medicine, economics, biology) but not in the life insurance.
The rest of the paper is organized as follows. Section 2 introduces the basic theory of BR. Section 3 describes the two-stage approach and how BR approach can be used for the estimate of withdrawal in order to get a full calculation of insurance benefit due to surrenders and withdrawals. Section 4 shows an application on a real life insurance portfolio and illustrates the results. Finally, Section 5 discusses the main findings and concludes.

2. An introduction to Beta regression

The Beta distribution has become more popular in recent years in modeling data bounded within open interval $(0,1)$, and is suitable to model rates and proportions therefore it is widely used in several fields of application.
Let $Y \sim \mathcal{B}e(a,b)$ be a Beta distributed random variable (henceforth r.v.) and its density function is given by:

$$\pi(y;a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\cdot\Gamma(b)} \cdot y^{a-1} \cdot (1-y)^{b-1}\ , \quad 0 < y < 1 \tag{2.1}$$

where $a > 0$, $b > 0$ and Γ is the Gamma function.

The mean and variance of Y are, respectively,

$$E(Y) = \frac{a}{(a+b)}, \quad Var(Y) = \frac{a \cdot b}{(a+b)^2 \cdot (a+b+1)} \tag{2.2}$$

However, for regression purposes it is typically more useful to model the mean of the response. To such aim a different parametrisation for the Beta density is used. Let $\mu = a/(a+b)$ and $\varphi = (a+b)$, i.e. $a = \mu \cdot \varphi$ and $b = (1-\mu)\cdot\varphi$. It follows from Eq. (2.2) that

$$E(Y) = \mu, \quad Var(Y) = \frac{V(\mu)}{1+\varphi} \tag{2.3}$$

where $V(\mu) = \mu(1-\mu)$ is the variance function. Eq. (2.3) provides the meaning of μ and φ as the mean and precision parameter respectively. As can be seen, for larger values of φ the variance of Y is smaller. The density in the new parametrization is:

$$\pi(y;\mu,\varphi) = \frac{\Gamma(\varphi)}{\Gamma(\mu\cdot\varphi)\cdot\Gamma((1-\mu)\cdot\varphi)} \cdot y^{\mu\cdot\varphi-1}(1-y)^{(1-\mu)\cdot\varphi-1}, \quad 0<y<1, \tag{2.4}$$

where $0<\mu<1$ and $\varphi>0$.
Beta regression techniques have been introduced by Cribari-Neto and Ferrari [2004] and can be considered as an extension of the Generalized Linear Model theory when the continuous dependent variables can be assumed as Beta distributed.
Simple approaches to model continuous proportions like the transformation of the response and the adoption of linear regression models have drawbacks: parameters cannot be easily interpreted in terms of the original response and, when measures of proportions typically display asymmetry, inference based on the normality assumption can be misleading.
Hence, BR enables to define the statistical relationship between the conditional mean of Y and a row vector of independent covariates, $\mathbf{x}_i = (x_{i1},\ldots,x_{im})$. Let $(\mathbf{x}_i, y_i)$ be a member of a set of observations $(i=1,\ldots,n)$ where y_i is a random sample from Y, the dependent variable of regression equation, and $\mathbf{x}_i = (x_{i1},\ldots,x_{im})$ is the row-vector introduced above (covariates). The conditional mean for the i-th insured is given by:

$$\mu_i = E(Y|\mathbf{x}_i) = E(Y^{(i)}) = g^{-1}(\mathbf{x}_i\cdot\beta), \tag{2.5}$$

where β is the column-vector of regression coefficients and g is the link function introduced in GLM theory by Nelder and McCullagh [1989]. In some cases, the choice of the link function g is bounded with the features of statistical procedures. There are several possible link functions g in a Beta regression context. For instance, one can use the Logit, the Probit or the Complementary Log-Log link function among others. A review of these link functions is proposed in Nelder and McCullagh [1989], and in Atkinson [1985].
An estimate of β, by means of (2.5), is obtained by maximizing the following Beta log-likelihood function based on the y_i sample and :

$$\begin{aligned}\ell_B(\beta,\mathbf{x},\varphi) = \sum_{i=1}^{n} & \log(\Gamma(\varphi)) - \log(\Gamma(\mu_i\varphi)) - \log(\Gamma((1-\mu_i)\varphi)) \\ & +(\mu_i\varphi-1)\log(y_i) + [(1-\mu_i)\varphi-1]\log(1-y_i)\end{aligned} \tag{2.6}$$

Cribari-Neto and Ferrari [2004] provide closed-form expressions for the score function, for Fisher's information matrix and its inverse.

Moreover, tests of hypotheses on the regression parameters (e.g. Wald test, likelihood test etc.) can be performed using approximations from the asymptotic normality of the maximum likelihood estimator.

3. A two-part Beta regression approach for the estimate of the policyholder behaviour in case of surrender and withdrawal.

Considering that many actuarial problems involve the assessment of proportions and/or rates, we show how the BR can be used as a flexible and efficient tool into a specific insurance problem.
In the following, we show how the BR can be applied in a two-part model for the estimate of the total cash flow due to lapses, when the contracts includes a surrender and a withdrawal option.

3.1. A basic approach for surrender and withdrawal rates estimates

Letting s being the surrender probability per unit time (month, year, etc.), actuarial literature and practice provide several procedures of estimate. A common practice is to calculate surrender rates by policy count as the ratio between "Number of Contracts surrendered" and "Number of Contracts Exposed to Surrender". However, number of contracts completely lacks volume information, i.e. contracts with a small and high amount of exposure (e.g. premium, reserve, sum insured) are identically considered. As an alternative, face amount is considered when surrender rates are computed as the ratio between "Face Amount at Surrender" and "Face Amount Exposed to Surrender". Thus, the surrender value will reflect contract size increasing the accuracy in the cash flows modelling. However, an under/overestimate of the number of surrenders that early terminate the contract, could be observed.
Furthermore, if withdrawals are admissible, it is relevant to consider the probability of withdrawal in a unit of time, w. However, it is noteworthy that withdrawals do not involve an early termination of the contract and imply that only a portion of the cash value is depleted from the total cash value.
Moreover, in case of surrender, the uncertainty relates only to the probability of occurrence, whereas in case of withdrawal further uncertainty is introduced by the reduction of the total cash value. The latter can be defined as a percentage of decrement of the face amount exposed to surrender conditioned to withdrawal

events in a unit of time (hereafter referred to as "withdrawn percentage"). The withdrawn percentage is continuous and included between 0 and 1.
Let B be the r.v. benefits due to surrenders and withdrawals per unit of time of a policy with one unit face amount (henceforth Total Cash Flow). We will assume in the following that in a unit of time (month, year, etc,) a single surrender or withdrawal event can occur. Following Ospina and Ferrari [2010], the r.v. B is a mixture between a Bernoulli distribution and a Beta distribution. Specifically, we assume that the cumulative distribution function (hereafter cdf) of the r.v. B is:

$$F_B\left(b;s,w,\mu,\varphi\right) = (1-w)\cdot Ber\left(b;\frac{s}{1-w}\right) + w\cdot BETA\left(b;\mu,\varphi\right) \tag{3.1}$$

where $Ber\left(\cdot;\frac{s}{1-w}\right)$ represents the cdf of a Bernoulli r.v. with parameter $\frac{s}{1-w}$ and $BETA\left(b;\mu,\varphi\right)$ is the Beta cdf, whose density function is defined by Eq. (2.4). Therefore, B has a zero-and-one-inflated beta distribution (i.e $B \sim BEINF\left(s,w,\mu,\varphi\right)$) with parameters s, w, μ and φ if its density function with respect to the measure generated by the mixture is given by:

$$f_B\left(b;s,w,\mu,\varphi\right) = \begin{cases} 1-s-w & \text{if } b=0 \\ s & \text{if } b=1 \\ w\cdot\pi(b;\mu,\varphi) & \text{if } 0<b<1 \end{cases} \tag{3.2}$$

It is worth noting, that $Prob\left(B=0\right) = 1-s-w$ and $Prob\left(B=1\right) = s$; hence, the expected value of this r.v. B is:

$$E\left[B\right] = s + w\cdot\mu. \tag{3.3}$$

Note that $E\left[B\right]$ is the weighted average of the expected value of the Bernoulli distribution with parameter $\frac{s}{1-w}$ and the expected value μ of the Beta distribution (see Eq. (2.3)) with weights $1-w$ and w respectively.

3.2. The regression models for surrenders and withdrawals estimates

The basic probabilistic framework introduced above, represents a very simple example of one of the most relevant actuarial assumptions for life insurance policies. In particular, such assumption belongs to the broader class of policyholder behavior risk that commonly refers to uncertainty with regards to policyholder premium payment patterns, premium persistency, surrenders, lapses, partial withdrawals, among others (American Academy of Actuaries [2019]). These poli-

cyholder behavior assumptions can vary due to policyholder characteristics. The Actuarial Standard of Practice 52 (ASOP 52 [2018]) states that "in setting partial withdrawal and surrender assumptions, the actuary should consider the insured's age and gender, the policy duration, the existence of policy loans, and scheduled changes in premium and benefit amounts".
Therefore, considering that a real portfolio of policies is characterized by a relevant number of risk factors, a very large number of homogenous risk classes may be identified. In such cases, an individual estimation approach for each homogeneous risk class is not feasible and an estimation method, that uses all the information available in a single model, is more appropriate.
To this aim, academic literature has largely debated on the use of the multivariate regression method and, in particular, Logistic regression for lapse/surrender rates estimation (Haberman and Renshaw [1996], Kim [2005], Cerchiara [2009], Eling and Kiesenbauer [2014]). However, such models are only finalized in the estimate of the surrender probability, but they do not deal with the estimate of withdrawal rates and the related amount.
Our goal is to broaden the estimation model, so as to identify the expected value of B given a portfolio containing a number n of policies and a set m of covariates. To this aim, for each policy $i = 1,\dots,n$, we need to estimate the probabilities s and w and the expected withdrawn percentage μ, conditioned to the vector of independent covariates $\mathbf{x}_i = (x_{i1},\dots,x_{im})$. To face such estimation problem introduced in Eq. (3.3), we propose the following two-part regression approach:

- First, we focused on the estimation of the probability that the i-th insured withdraws w_i or surrenders s_i, by means of a Multinomial regression model (Venables and Ripley [2003]).

  MR is a classification method that generalizes binomial regression to multiclass problems, i.e. with more than two possible discrete outcomes. It is used to predict the probabilities of J different possible outcomes of a categorically distributed dependent variable Z, given a set of independent variables $\mathbf{x}_i$, as follows:

  $$Prob\,[Z = j|\mathbf{x}_i] = h_j^{-1}(\mathbf{x}_i,\Gamma),\ j = 1,\dots,J \text{ and } i = 1,\dots,n \tag{3.4}$$

  where Γ is a matrix of regression coefficients and h_j is a link function that should be selected so that the probabilities lie between 0 and 1 and sum over j to one. Different functional forms of h_j lead to multinomial ordered probit or logit model.

Since the multinomial conditional density for one observation of Z is:

$$f_Z(\mathbf{x}_i, \Gamma, z_1, \cdots, z_J) = \prod_{j=1}^{J} h_j^{-1}(\mathbf{x}_i, \Gamma)^{z_j} \tag{3.5}$$

where the generic z_j is equal to 1 if $Z = j$ and 0 otherwise,

the maximum likelihood estimate of the regression coefficients Γ are estimated by maximizing the following log-likelihood function:

$$\ell_Z(\mathbf{x}, \Gamma, z_1, \ldots, z_J) = \sum_{i=1}^{n} \sum_{j=1}^{J} z_{i,j} \log \left[h_j^{-1}(\mathbf{x}_i, \Gamma) \right] \tag{3.6}$$

- Second we need to model the fraction of the face amount withdrawn. To this aim, as in case of withdrawal B assumes values within open interval $(0,1)$, it should be considered as the response variable Y of a BR model as described in the previous Section 2.
  The BR model identifies, for each insured, the conditional expectation of the withdrawn percentage μ_i, by means of Eq. (2.5).

It is worth noting that the estimates of s_i, w_i, μ_i and φ allow an estimate of the conditional density $f_{B_i}(b; s_i, w_i, \mu_i, \varphi)$. Then, it is quite simple to perform a simulation model in order to get not only a point estimate, but also an interval estimate of the r.v. involved in our problem.

4. An application on a life insurance portfolio

In this section, we present an application of the proposed methodology to a real insurer database. The data set consists on seriatim data of single-premium whole life participating policies with zero interest rate, no minimum-guarantee and no surrender charge, from an Italian life insurance company between years 2009 and 2017.

4.1. Dataset analysis

In order to investigate the different behaviour of policyholders respect to the surrender/withdrawal events, the data set contains seriatim information on policies but limited to the following risk factors: Policy Duration (PD), Gender (G), Policyholder Underwriting Age (UA) and Mathematical Reserve (Res). It is worth noting that these characteristics are in line with the request of the Actuarial Standard of Practice No. 52 (ASOP 52 [2018]). Descriptive statistics of the portfolio for quantitative covariates are exhibited in Table 5.1. The average insurance contract is held by a 52.68 years old policyholder with 7.88 years duration and a mathematical reserve of 3.85 thousand euros. The *UA* distribution is sufficiently symmetric and lightly platykurtic while the *PD* is right skewed with a mode in 3. On the contrary, the distribution of the mathematical reserve shows a very large positive skewness and kurtosis suggesting highly asymmetric and right skewed distribution as appreciable by the values assumed by the median (2.13), the mean (3.85), the 0.95 quantile (11.30) and the maximum value (277.18). Moreover, our dataset contains surrenders and withdrawals for policies with policy duration greater than 2 years, as the contract conditions do not allow to lapse during the first two years, while the maximum observed policy duration is 21 years.

Table 4.1: Descriptive Statistics of the dataset

	Policy Duration (in years)	*Policyholder Underwriting Age (in years)*	*Math. Reserve (in k€)*	*Surrender (in k€)*	*Withdrawal (in k€)*
Mean	7.88	52.68	3.85	3.49	1.45
Median	7.00	53.00	2.13	2.07	0.62
Mode	3.00	63.00	1.22	1.50	0.50
St. Dev.	4.10	16.87	7.04	4.83	2.60
Coeff. of variation	52.02%	32.02%	182.81%	138.30%	179.71%
Kurtosis	-0.55	-0.94	348.83	60.49	55.47
Skewness	0.69	-0.03	13.93	6.04	6.01
Min	3.00	6.00	0.01	0.05	0.01
0.05 Quantile	3.00	25.00	0.31	0.31	0.10
0.95 Quantile	16.00	87.00	11.30	10.64	5.01
Max	21.00	94.00	277.18	84.87	46.00

In Table 5.2, a statistical summary of the dataset is reported. In the following tables and figures, *UA* and *Res* are grouped in bins just for sake of representation. We have selected the bins of the same size using groups of ten for *UA* and a thousand for *Res*. The latter considers a last class for values over ten thousand euro, corresponding to values over the 0.92 observed quantile. Policy duration is an integer variable, then all the levels are reported (from 3 to 21). It is worth noting that, with the exception of *G*, all the covariates are considered as numerical in the following regression models.

The surrender rates calculated per policy count are basically greater than the ones calculated per face amount, with the exception of *Res*, where there is a substantial alignment. The latter outcome is not unexpected, as contracts are grouped into more homogeneous classes per economic exposure; hence, as stated in Section 3.1, this condition implies a reduction of the differences between the two estimation methods.
The observed surrender rate by policy count is equal to $11.00\%(15,469/140,579)$, whereas by face amount is $9.99\%(54,052/541,195)$, with a difference of 1.01%. By comparing the surrender rates on each level of the variables, it is possible to observe that there are cases where this difference increases: by the way of example in case of level 4 of variable PD (the second most relevant level in terms of policy count and the third in terms of face amount) the surrender rate by policy count is equal to 15.13%, whereas by face amount is 12.75%, with a difference of 2.38%. This difference is attributable to a non-homogeneous portfolio per economic exposure, indeed if all the policies had the same face amount the surrender rates per policy and per face amount would be the same. To solve this issue, in setting surrender assumption an actuary should consider economic exposures (e.g. benefit amounts) among others features (ASOP 52 [2018]).
In this sense, it is useful to consider an economic exposure measure like mathematical reserve as covariate for the frequency component (MR).

Table 4.2: Data Summary

Variable		Number of policies			Rate by policy count		Amounts (in k€)			Rate by face amount		Withdrawal %
Acronym	Levels	Policies	Withdrawal	Surrender	Withdrawal	Surrender	Math. Res.	Withdrawal	Surrender	Withdrawal	Surrender	
PD	3	18,953	917	3,278	4.84%	17.30%	83,227	1,373	14,379	1.65%	17.28%	25.71%
	4	17,170	839	2,598	4.89%	15.13%	75,935	1,318	9,680	1.74%	12.75%	28.47%
	5	15,604	883	2,449	5.66%	15.69%	73,117	1,399	8,912	1.91%	12.19%	25.84%
	6	15,619	830	2,275	5.31%	14.57%	76,576	1,429	8,264	1.87%	10.79%	27.55%
	7	11,821	551	901	4.66%	7.62%	64,305	1,042	3,558	1.62%	5.53%	26.10%
	8	7,573	262	653	3.46%	8.62%	38,218	581	2,233	1.52%	5.84%	28.70%
	9	7,593	312	653	4.11%	8.60%	22,864	477	1,949	2.09%	8.52%	38.71%
	10	8,340	357	494	4.28%	5.92%	21,408	358	1,152	1.67 %	5.38%	36.50%
	11	7,503	333	342	4.44%	4.56%	18,968	339	863	1.79%	4.55%	41.73%
	12	6,818	353	276	5.18%	4.05%	17,446	321	536	1.84%	3.07%	44.54%
	13	6,198	326	288	5.26%	4.65%	15,774	352	606	2.23%	3.84%	48.98%
	14	5,407	211	342	3.90%	6.33%	12,926	233	519	1.80%	4.01%	47.33%
	15	3,866	138	355	3.57%	9.18%	6,194	103	448	1.66%	7.23%	52.63%
	16	3,418	129	255	3.77%	7.46%	5,537	93	352	1.67%	6.36%	50.37%
	17	3,118	134	206	4.30%	6.61%	5,280	115	349	2.17%	6.61%	58.31%
	18	1,164	26	82	2.23%	7.04%	2,288	24	176	1.05%	7.68%	57.69%
	19	197	4	9	2.03%	4.57%	559	10	18	1.71%	3.28%	44.87%
	20	163	0	11	0.00%	6.75%	453	-	52	0.00%	11.53%	0.00%
	21	54	1	2	1.85%	3.70%	120	0	8	0.18%	6.77%	17.85%
G	F	71,667	3,443	7,422	4.80%	10.36%	263,552	4,506	24,620	1.71%	9.34%	28.52%
	M	68,912	3,163	8,047	4.59%	11.68%	277,643	5,061	29,432	1.82%	10.60%	30.89%
UA	< 30	13,990	782	1,180	5.59%	8.43%	21,511	612	1,914	2.85%	8.90%	49.23%
	$[30,40)$	21,796	1,089	1,989	5.00%	9.13%	46,746	1,035	4,321	2.21%	9.24%	38.40%
	$[40,50)$	25,950	1,308	2,832	5.04%	10.91%	84,470	1,820	8,014	2.15%	9.49%	33.87%
	$[50,60)$	25,419	1,202	3,031	4.73%	11.92%	102,820	2,044	10,863	1.99%	10.57%	32.13%
	$[60,70)$	26,428	1,166	3,185	4.41%	12.05%	118,901	1,907	13,204	1.60%	11.11%	27.48%
	≥ 70	26,996	1,059	3,252	3.92%	12.05%	166,746	2,149	15,736	1.29%	9.44%	22.44%
Res	$(0-1]$	28,115	1,600	3,241	5.69%	11.53%	15,058	549	1,790	3.65%	11.89%	60.95%
	$(1-2]$	35,122	1,293	3,957	3.68%	11.27%	54,411	985	6,007	1.81%	11.04%	50.68%
	$(2-3]$	28,139	879	2,996	3.12%	10.65%	68,593	766	7,333	1.12%	10.69%	35.03%
	$(3-4]$	14,058	668	1,561	4.75%	11.10%	50,144	670	5,553	1.34%	11.07%	28.00%
	$(4-5]$	7,029	348	803	4.95%	11.42%	32,991	485	3,787	1.47%	11.48%	29.71%
	$(5-6]$	7,027	324	823	4.61%	11.71%	38,296	516	4,467	1.35%	11.66%	29.25%
	$(6-7]$	3,938	265	405	6.73%	10.28%	25,516	481	2,617	1.89%	10.26%	28.04%
	$(7-8]$	3,093	175	288	5.66%	9.31%	23,547	388	2,195	1.65%	9.32%	29.23%
	$(8-9]$	1,616	97	163	6.00%	10.09%	13,858	235	1,400	1.69%	10.10%	28.35%
	$(9-10]$	1,326	104	120	7.84%	9.05%	12,529	293	1,135	2.34%	9.06%	29.82%
	$(10-inf)$	11,116	853	1,112	7.67%	10.00%	206,252	4,200	17,768	2.04%	8.61%	25.43%
Total		**140,579**	**6,606**	**15,469**	**4.70%**	**11.00%**	**541,195**	**9,567**	**54,052**	**1.77%**	**9.99%**	**29.72%**

The *Gender* analysis shows that males have a greater propensity to surrender (11.68%) than females (10.36%).
The withdrawal rate calculated per policy count (4.70%) is greater than the one calculated per face amount (1.77%). This is confirmed for each variables' level. It is worth noting that for *Res*, even though the contracts are grouped into more homogeneous risk classes, the estimates are not similar. This finding is not surprising because, as stated in Section 3.1, withdrawals introduce a component of uncertainty identified by the Beta component in Eq. (3.1).

Furthermore, the last column in Table 5.2 illustrates the withdrawn percentage and put in evidence different policyholder behaviour towards the withdrawal event. Considering that, the policies have not surrender charge, the withdrawn percentage is obtained considering the mathematical reserve of the withdrawn policies as denominator. As expected, policies with a lower (higher) face amount show a higher (lower) withdrawn percentage. A similar decreasing trend is shown by *UA*, that may be due to a lower exposure for younger individuals or to a greater demand for cash values at a young age. Whereas, we observe an increasing trend of the withdrawn percentage as *PD* increases. These different trends, observed for withdrawn percentage, suggest the need to deepen a multivariate analysis. To this aim BR represents a suitable model to perform this analysis.
In the following, our goal is the estimation of the elements in Eq. (3.3), through the two-part process. To perform statistical analysis, we split the dataset into a training and testing samples. The training sample is used to fit a predictive model, and its performance is evaluated on the test subset. The training set is randomly selected and consists of 75% of initial database.

4.2. Multinomial Logistic regression for surrender and withdrawal rates

The Multinomial Logistic regression provides the estimation of w_i, s_i conditioned to the vector of independent covariates $\mathbf{x}_i = (x_{i1}, \ldots, x_{im})$.

Then, considering a logit function for h, by Eq. (**??**) it holds:

$$w_i = \frac{e^{\left(\gamma_0^{(w)} + \sum_{k=1}^{\delta} \gamma_{PD,k}^{(w)} (PD_i)^k + \gamma_{G_i}^{(w)} (G_i) + \sum_{k=1}^{\delta} \gamma_{UA,k}^{(w)} (UA_i)^k + \sum_{k=1}^{\delta} \gamma_{Res,k}^{(w)} (Res_i)^k\right)}}{1 + \sum_j e^{\left(\gamma_0^{(j)} + \sum_{k=1}^{\delta} \gamma_{PD,k}^{(j)} (PD_i)^k + \gamma_{G_i}^{(j)} (G_i) + \sum_{k=1}^{\delta} \gamma_{UA,k}^{(j)} (UA_i)^k + \sum_{k=1}^{\delta} \gamma_{Res,k}^{(j)} (Res_i)^k\right)}}, \quad (4.1)$$

$$s_i = \frac{e^{\left(\gamma_0^{(s)}+\sum_{k=1}^{\delta}\gamma_{PD,k}^{(s)}(PD_i)^k+\gamma_{G_i}^{(s)}(G_i)+\sum_{k=1}^{\delta}\gamma_{UA,k}^{(s)}(UA_i)^k+\sum_{k=1}^{\delta}\gamma_{Res,k}^{(s)}(Res_i)^k\right)}}{1+\sum_j e^{\left(\gamma_0^{(j)}+\sum_{k=1}^{\delta}\gamma_{PD,k}^{(j)}(PD_i)^k+\gamma_{G_i}^{(j)}(G_i)+\sum_{k=1}^{\delta}\gamma_{UA,k}^{(j)}(UA_i)^k+\sum_{k=1}^{\delta}\gamma_{Res,k}^{(j)}(Res_i)^k\right)}}, \quad (4.2)$$

where $\gamma_{\cdot}^{(j)} \in \Gamma$, are the Multinomial Logistic regression coefficients for event $j = \{w, s\}$ and δ the degree of polynomial used for the numerical variables (i.e. *PD*, *Res* and *UA*). In this example, to avoid over parametrization, we set $\delta = 3$. The base levels for the categorical variable G is "Female". The coefficients' estimates are reported in Table 5.3.

Table 4.3: Results of the estimation using Multinomial Logistic regression model

Coefficient		Estimate		*p*-value	
Acronym	Symbol	Withdrawal	Surrender	Withdrawal	Surrender
Base	γ_0	-2.09E+00	-6.93E-01	$< 10^{-6}$	$< 10^{-6}$
G	$\gamma_{G=M}$	-5.28E-02	1.35E-01	$< 10^{-6}$	$< 10^{-6}$
PD	$\gamma_{PD,1}$	-1.24E-01	2.80E-02	$< 10^{-6}$	$< 10^{-6}$
PD	$\gamma_{PD,2}$	9.86E-03	-2.95E-02	$< 10^{-6}$	$< 10^{-6}$
PD	$\gamma_{PD,3}$	-3.21E-04	1.37E-03	$< 10^{-6}$	$< 10^{-6}$
UA	$\gamma_{UA,1}$	-2.21E-03	-4.92E-02	$< 10^{-6}$	$< 10^{-6}$
UA	$\gamma_{UA,2}$	-1.93E-04	1.28E-03	$< 10^{-6}$	$< 10^{-6}$
UA	$\gamma_{UA,3}$	1.18E-06	-9.26E-06	$< 10^{-6}$	$< 10^{-6}$
Res	$\gamma_{Res,1}$	5.65E-02	-8.89E-02	$< 10^{-6}$	$< 10^{-6}$
Res	$\gamma_{Res,2}$	-7.05E-04	2.98E-03	$< 10^{-6}$	$< 10^{-6}$
Res	$\gamma_{Res,3}$	1.95E-06	-2.53E-05	$< 10^{-6}$	$< 10^{-6}$

As one can see all the coefficients are highly significant, hence all selected factors give a relevant contribution to the explanation of the policyholder behaviour respect to withdrawals and surrenders occurrence.

As mentioned above, in order to evaluate the models, a validation dataset representing 25% of the global data is served to verify the prediction quality of the models. To assess the classification accuracy in the multinomial regression model, we adopt the Receiver Operating Characteristic (ROC) curve, typically used in binary classification, and we use the area under the curve (AUC) as validation measure. AUC ranges in value from 0 to 1, where a value of 0 indicates a perfectly inaccurate test and a value of 1 reflects a perfectly accurate test. However, considering that we deal with a multi-class dataset (i.e. "Withdrawal", "Surrender", "Other"), it is necessary to extend ROC curve and AUC to face with multiple classes classifications problem. Moreover, the dataset is imbalanced with most of the data falling in "Other" (84.3%) and the remaining between "Surrender" (11.00%) and "Withdrawal" (4.70%). One can draw a ROC curve by considering each element of the indicator matrix as a binary prediction (micro-averaging), in such a case. Micro-averaging treats the entire set of data as an aggregate result, in order to convert multiclass prediction into binary prediction and compute the metric average. This measure is recommended in multi-class classification setup where class levels are imbalanced. In Figure 5.1, the ROC curve is plotted with True Positive Rate (TPR) against the False Positive Rate (FPR), where TPR is on the y-axis and FPR is on the x-axis. AUC measures the entire two-dimensional area underneath the entire ROC curve and is equal to 91.47% showing high classification accuracy.

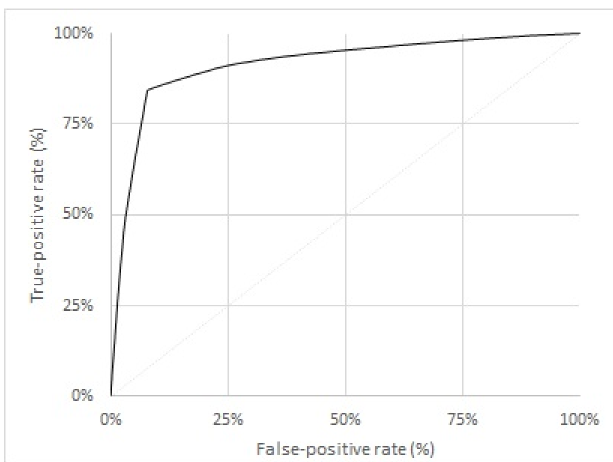


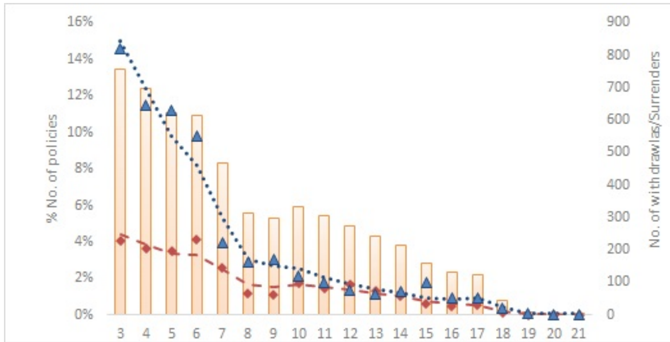
Figure 4.1: Micro-average Receiver Operating Characteristic curve (ROC)

In Table 5.4, for sake of simplicity, we show the calculus of the number of surrenders and withdrawals for *UA* classes.
It is worth noting that the estimates are very close to the observed values, indeed the differences are in the range $(-1.69\%, 2.61\%)$ for surrenders and $(-2.58\%, 5.95\%)$ for withdrawals. As one can see, considering the total number of surrenders and withdrawals, the error is negligible (0.22%) for surrenders and (1.18%) for withdrawals.

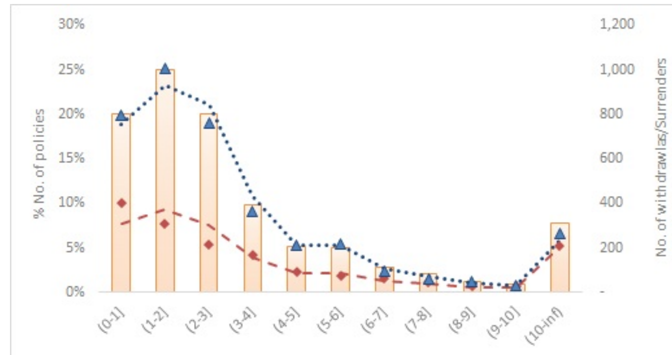
Table 4.4: Total Surrender and Withdrawal events for Policyholder Underwriting Age

UA	Surrender			Withdrawal		
	Observed	Fitted	Δ%	Observed	Fitted	Δ%
< 30	291	291	0.04%	195	190	-2.58%
$[30, 40)$	498	498	0.04%	268	284	5.95%
$[40, 50)$	708	713	0.77%	322	330	2.47%
$[50, 60)$	754	750	-0.55%	300	295	-1.67%
$[60, 70)$	796	817	2.61%	289	281	2.69%
≥ 70	815	801	-1.69%	262	275	5.04%
Total	3,862	3,871	0.22%	1,636	1,655	1.18%

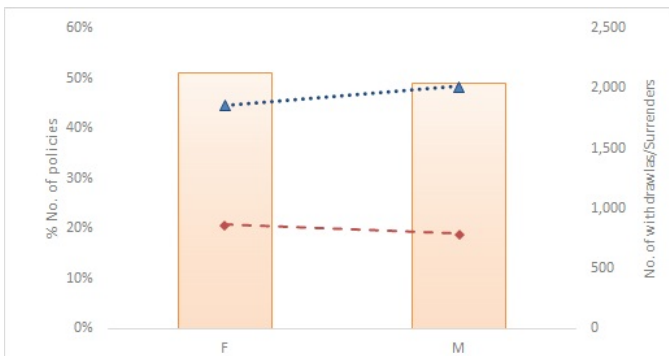
In Figure 5.2 a comparison between observed and fitted number of surrenders and withdrawals, for each covariate is reported.



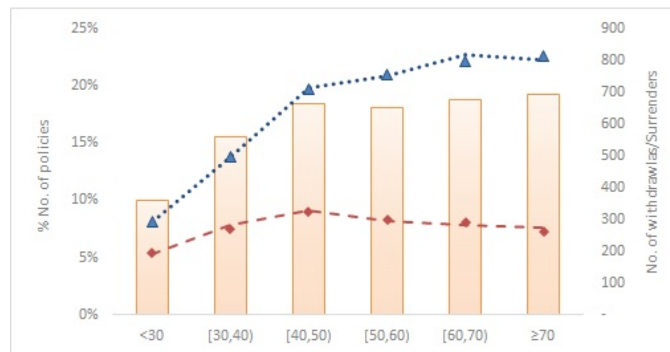
(a) Policy Duration



(b) Mathematical Reserve in k€



(c) Gender



(d) Policyholder Underwriting Age

Figure 4.2: Multinomial Logistic regression goodness of fit. Primary y-axis: Number of policies in percentage (Histograms). Secondary y-axis: Observed number of withdrawls (Diamonds) and surrenders (Triangles). Fitted number of withdrawals (Dashed line) and surrenders (Dotted line)

4.3. Beta regression for expected withdrawn percentage estimate

The BR provides the estimation of the expected withdrawn percentage conditioned to the vector of independent covariates $\mathbf{x}_i = (x_{i1}, \dots, x_{im})$, (see Eq. (2.5)). To avoid confusion, for BR we use the same risk factors as used for the Multinomial Logistic regression model. For the link function g, we test functions suggested in Cribari-Neto and Ferrari [2004] and we select the one that generates the greatest log-likelihood value as reported in Table 5.5 As can be seen, the Complementary log-log shows the highest log-likelihood. Thus, we define the BR model by the following equation:

Table 4.5: Link function selection for BR: log-likelihood values

g	$\ell_B(\mu_i, \varphi)$
Logit	7,182
Probit	7,176
Complementary log-log	7,226
Log-log	7,133

$$\mu_i = 1 - e^{-e^{\left(\beta_0 + \Sigma_{k=1}^{\delta} \beta_{PD,k}(PD_i)^k + \beta_{G_i}(G_i) + \Sigma_{k=1}^{\delta} \beta_{UA,k}(UA_i)^k + \Sigma_{k=1}^{\delta} \beta_{Res,k}(Res_i)^k\right)}} \tag{4.3}$$

The maximum likelihood estimates of all the coefficients β and precision parameter φ are reported in Table 5.6, as well as the p-value for asymptotic Wald test. It is worth noting that all the parameters are significant.

In Table 5.7, for sake of simplicity, we show the calculus of the cash flow estimates for withdrawals grouped for *UA* classes.
It is worth noting that the estimates are very close to the observed values, indeed the differences are in the range $(-2.51\%, 6.31\%)$. As one can see, considering the total cash flow amount the error is negligible (2.83%).

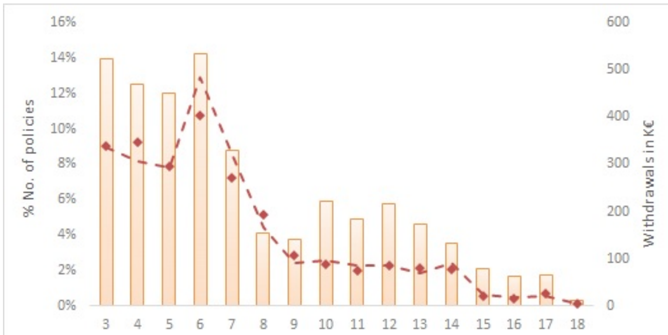
In Figure 5.3, a comparison between observed (triangles) and fitted (dotted line) withdrawal amounts for each risk factor is reported, where the exposure is the percentage of number of policies for each covariate.

Table 4.6: Coefficients estimates of the Beta regression

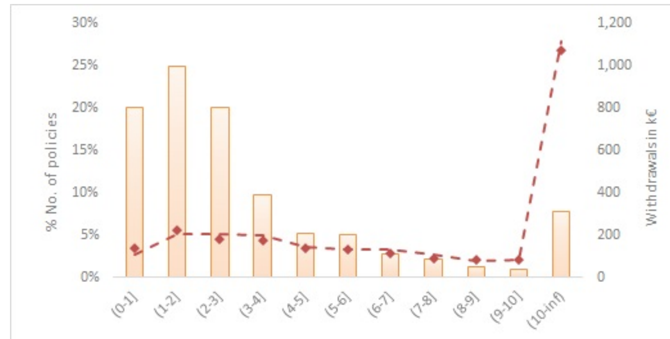
Acronym	Symbol	Estimate	p-value
Base	β_0	-6.83E-01	1.14E-05
G	$\beta_{G=M}$	8.37E-02	$< 10^{-6}$
PD	$\beta_{PD,1}$	-1.30E-01	$< 10^{-6}$
PD	$\beta_{PD,2}$	2.22E-02	$< 10^{-6}$
PD	$\beta_{PD,3}$	-7.66E-04	$< 10^{-6}$
UA	$\beta_{UA,1}$	1.51E-02	8.70E-02
UA	$\beta_{UA,2}$	-4.34E-04	9.34E-03
UA	$\beta_{UA,3}$	2.52E-06	1.23E-02
RES	$\beta_{Res,1}$	-1.59E-02	$< 10^{-6}$
RES	$\beta_{Res,2}$	1.61E-04	$< 10^{-6}$
RES	$\beta_{Res,3}$	-4.52E-07	$< 10^{-6}$
Precision	φ	2.88E-02	$< 10^{-6}$

Table 4.7: Withdrawal amount in k€for Policyholder Underwriting Age estimated

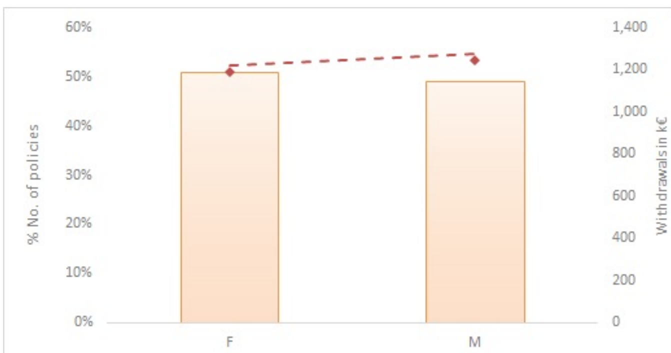
UA	Observed	Fitted	Δ%
<30	134.17	131.55	-1.95%
$[30,40)$	251.41	257.12	2.27%
$[40,50)$	458.90	480.71	4.75%
$[50,60)$	528.07	514.82	-2.51%
$[60,70)$	475.91	505.94	6.31%
≥ 70	583.20	610.44	4.67%
Total	2,431.66	2,500.58	2.83%



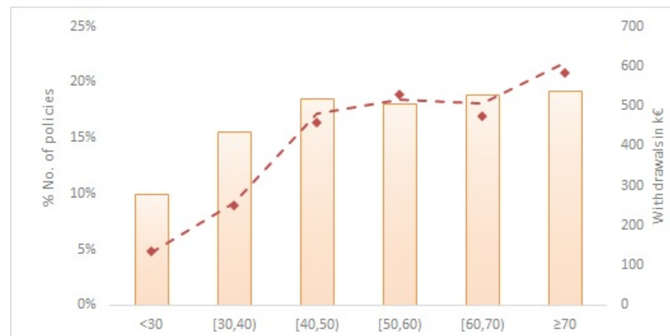
(a) Policy Duration



(b) Mathematical Reserve in k€



(c) Gender



(d) Policyholder Underwriting Age

Figure 4.3: Withdrawal Amounts. Primary y-axis: Number of policies in percentage (Histograms). Secondary y-axis: Observed (Diamonds) and fitted (Dashed line) withdrawal amounts

In order to measure the estimation accuracy, we introduce the goodness-of-fit (GoF) measure called Nash-Sutcliffe Efficiency (NSE). The latter is a normalized statistic that determines the relative magnitude of the residual variance compared to the measured data variance. Given a dependent variable y and its estimate $\hat{y}$, NSE can be defined as: $NSE = 1 - \frac{\sum(y-\hat{y})^2}{\sum(y-E(y))^2}$.
The NSE can be interpreted as test statistic for the accuracy of model predictions. The NSE ranges from $-\infty$ to 1: if $NSE = 1$, there is a perfect match of the modeled to the observed data; if $NSE = 0$, the model predictions are as accurate as the mean of the observed data, if $-\infty < NSE < 0$, the observed mean is a better predictor than the model. It means that 0 is a critical value to accept or reject the model. We obtain on the test set an NSE of 22.38%, which is compliant with acceptance.

Finally, to test the goodness of fit, we also introduce a variant of the graphical half-normal plot method. Following Atkinson [1985], when the distribution of residuals is not known, half-normal plots with simulated envelopes are a helpful diagnostic tool. The main idea is to enhance the usual half-normal plot by adding a simulated envelope, which can be used to decide whether the observed residuals are consistent with the fitted model. Half-normal plots with a simulated envelope can be produced as follows:

- fit the model and generate a simulated sample of I independent observations using the fitted model, as if it were the true model;
- fit the model to the generated sample, and compute the ordered absolute values of the residuals;
- repeat steps (1) and (2) π times;
- consider the I sets of the π order statistics; for each set compute its average, minimum and maximum values;
- plot these values and the ordered residuals of the original sample against the half-normal scores $\phi^{-1}\left((i+I-1/8)/(2I+1/2)\right)$

The minimum and maximum values of the ψ order statistics yield the envelope. Atkinson suggests to use $\psi = 19$, so that the probability that a given absolute residual will fall beyond the upper band provided by envelope is approximately equal to $1/20 = 5\%$. Observations corresponding to absolute residuals outside the limits provided by the simulated envelope are worthy of further investigation. Additionally, if a considerable proportion of points falls outside the envelope, then one has evidence against the adequacy of the fitted model. Figure 5.4 shows the simulated half-normal plot that states evidence in favour of the model accuracy.

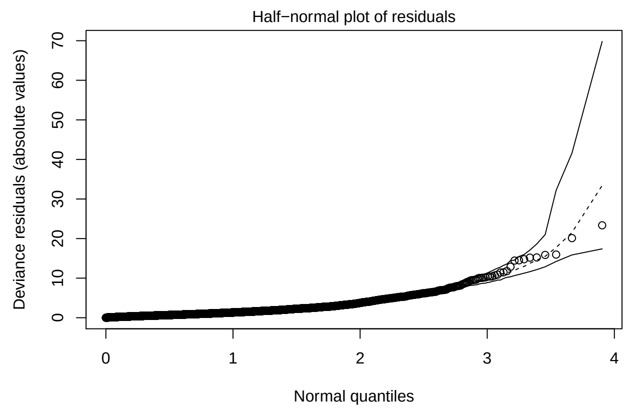
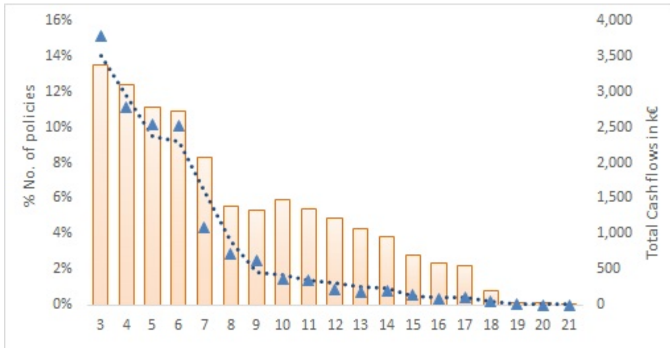


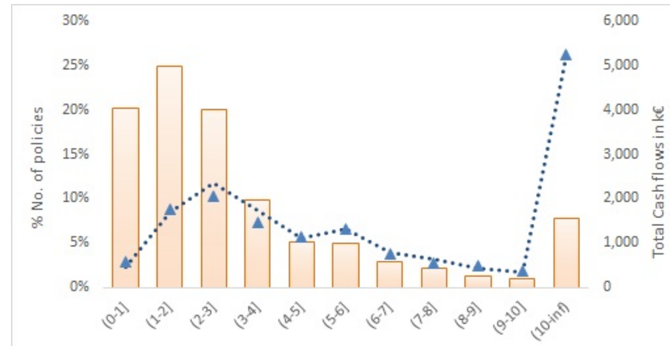
Figure 4.4: Half-Normal plot

4.4. Final results

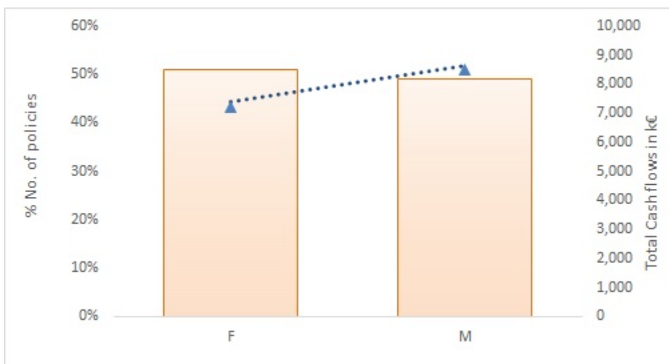
Once estimated the surrender and withdrawal rates and the expected percentage withdrawn for each *i*-th insured, i.e. s_i, w_i, μ_i, $i = 1, \ldots, n$, it is easy to compute the expected Total Cash Flow by means of (3.3). Figure 5.5 shows the comparison between Total Cash Flow observed (triangles) and fitted (dotted line), for each rating factor.



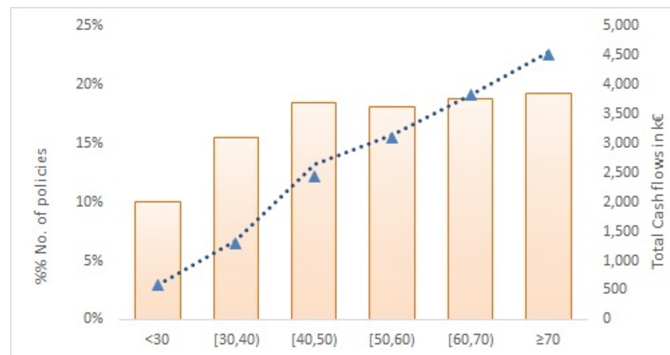
(a) Policy Duration



(b) Mathematical Reserve



(c) Gender



(d) Policyholder Underwriting Age

Figure 4.5: Total Cash Flow fitting analysis. Primary y-axis: Number of policies in percentage (Histograms). Secondary y-axis: Observed (Triangles) and fitted (Dotted line) cash flow amounts

Finally, in Table 5.8, for sake of simplicity we show the calculus of the Total Cash Flow estimates grouped by *UA* classes.
It is worth noting that the estimates are very close to the observed values, indeed

Table 4.8: Total Cash Flow for Policyholder Underwriting Age

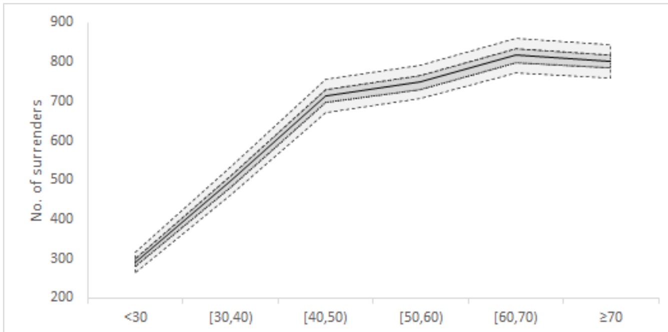
UA	Total Cash Flow in k€		Δ%
	Observed	Fitted	
< 30	599.22	586.64	-2.10%
$[30, 40)$	1,295.75	1,345.89	3.87%
$[40, 50)$	2,420.02	2,634.54	8.86%
$[50, 60)$	3,088.92	3,128.23	1.27%
$[60, 70)$	3,809.79	3,820.36	0.28%
≥ 70	4,506.50	4,544.79	0.85%
Total	15,720.20	16,060.49	2.16%

the differences are in the range $(-2.10\%, 8.86\%)$. As one can see, considering the total cash flow amount the error is negligible (2.16%). Furthermore, we calculate on the test set NSE measure to compare the fitted cash flows to the observed values getting 8.6%, that is compliant with acceptance.
Lastly, to demonstrate additional benefits deriving from our approach we describe a simulation model to calculate the probability distribution of the number of surrenders, withdrawals as well as the corresponding amounts. It is relevant to observe that the simulation approach is carried out at policyholder level. Indeed, as previously stated, the estimates of s_i, w_i, μ_i and φ allow an estimate of the conditional density in Eq. (3.2). This means that we can perform the analysis of the predictive distribution considering different levels of aggregation, starting from a single policyholder to the total portfolio.
Let K be the number of iteration and κ be the iteration index. For a generic policyholder $i = 1, \ldots, n$, for each iteration κ the simulation is performed as follows:

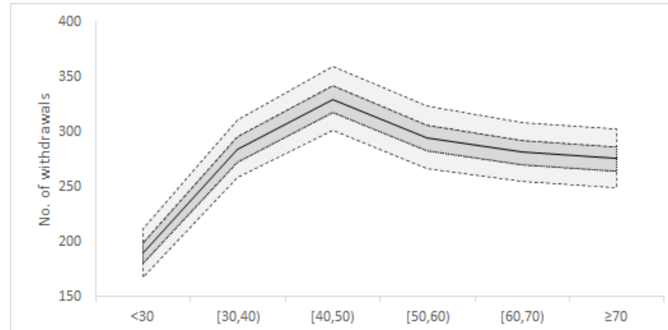
- generate a random number $u_i^{(\kappa)}$ from a uniform distribution $U \sim Unif(0, 1)$;
- if $u_i^{(\kappa)} < w_i$, a withdrawal events occurs and we sample from a $BETA(\cdot; \mu_i, \varphi)$ to get a pseudo realization of the withdrawal percentage $\hat{b}_i^{(\kappa)} \in (0, 1)$;

- if $u_{i,\kappa} \geq w_i$ to establish if the policyholder surrenders or not we sample from a $Ber\left(\cdot; \frac{s_i}{1-w_i}\right)$ to get a pseudo realization $\hat{b}_i^{(\kappa)} \in \{0,1\}$

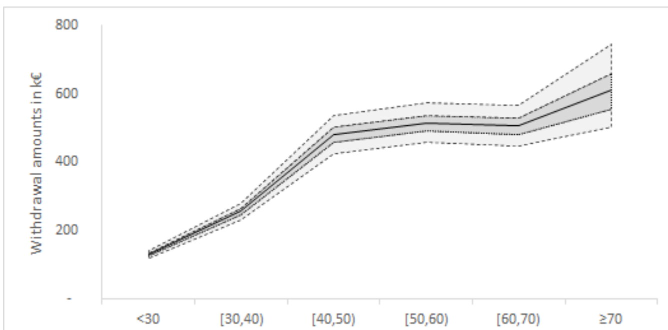
We apply this simulation method to the test set composed by $n = 35,122$ and by setting the number of iterations $K = 5,000$.
In Figure 5.6 we show the interval estimates of the number of surrenders and withdrawals, as well as the amount of withdrawals and total cash flows by *UA* classes. The black line represents the point estimate of the expected value already shown in Tables 5.4, 5.7, and 5.8, whereas the areas in dotted lines represents the confidence interval with an error probability of 10% and 50%, respectively.



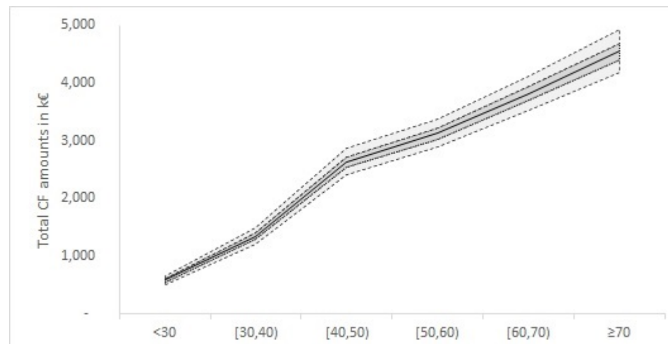
(a) Number of surrenders



(b) Number of withdrawals



(c) Amount withdrawn



(d) Total cash flow amount

Figure 4.6: Interval estimates of the two-part model. mean (solid line), confidence interval error probability of 10% (light gray) and 50% (dark gray)

Finally, in Table 5.9 is exposed a statistical summary of the sample distribution of the r.v. of interest.

Table 4.9: Statistical summary of simulated distributions

	No. of surrenders	*No. of withdrawals*	*Withdrawals (in k€)*	*Cash flows (in k€)*
Mean	3,872.45	1,655.65	2,501.00	16,064.90
Median	3,873.00	1,655.00	2,496.08	16,070.03
Mode	3,877.00	1,655.00	2,496.00	16,177.00
St. dev	58.47	39.92	96.72	365.67
Coeff. Of Variation	1.51%	2.41%	3.87%	2.28%
Kurtosis	0.09	0.01	0.13	0.02
Skweness	0.03	0.02	0.27	0.07
Min	3,657.00	1,506.00	2,199.80	14,833.85
0.05 Quantile	3,776.00	1,591.00	2,350.98	15,472.39
0.95 Quantile	3,968.00	1,723.00	2,667.90	16,665.13
Max	4,110.00	1,802.00	2,877.64	17,653.01

5. Final comments

This paper deals with an application of a two-part Beta regression in a life actuarial context. BR is very suitable when practitioners deals with rates and proportions modelling. Although BR is largely discussed by academic literature and used by practitioners, as far as we know, it has been scarcely used in the actuarial field.

To this aim, we introduce a two-part model to investigate the effects on the expected benefits on a contract when surrender and withdrawal options are both eligible, as in some life insurance contracts or pension funds. In this context the accuracy in the estimation of number of in force policies as well as the cash flow amounts is relevant; the latter must take into account the proportion of the cash value depleted from the total cash value when a withdrawal event occurs. Hence, it is useful to investigate the behaviour of the policyholders towards their risk features.

To this aim, we have suggested to estimate in a first stage the number of surrenders

and withdrawals by means of a Multinomial Logistic regression; then, in the second stage, the percentage of cash amount paid for withdrawals, which is included between 0 and 1, is modelled by a Beta regression.
Furthermore, based on a real data set, we provide the estimates of the number of surrenders and withdrawals as well as the corresponding cash values. Our findings confirm that the policyholder behaviour is affected by policyholder and contract features, for this kind of events. This is also proved by each statistical test, where high significance for each risk factor is observed, for both regression models proposed. Moreover, the expected cash flows outcomes show an acceptable goodness of fit measures.

References

Aleandri M., 2018. Modeling Dynamic Policyholder Behaviour through Machine Learning Techniques, Sapienza University of Rome https://www.dss.uniroma1.it/it/system/files/pubblicazioni/CopiaPubblicazione.pdf Accessed 10 Sep 2019

American Academy of Actuaries, (2019). Life Principle-Based reserves (PBR) Assumptions Resource Manual https://www.actuary.org/sites/default/files/files/publications/PBR_Assumptions_Resource_Manual_012919.pdf Accessed 19 Nov 2020

Task Force on Principle-Based Reserves of the Life Committee of the Actuarial Standards Board, 2017. Actuarial Standard of Practice No. 52: Principle-Based Reserves for Life Products under the NAIC Valuation Manual, https://www.actuarialstandardsboard.org/wp-content/uploads/2017/10/asop052_189.pdf

Atkinson, A. C., 1985. An Introduction to Graphical Methods of Diagnostic Regression Analysis. Oxford University Press, New York.

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

Bacinello, A. R., 2005. Endogenous model of surrender conditions in equity-linked life insurance. Insurance: Mathematics and Economics 37(2), 270-296.

Baione, F., Biancalana, D., De Angelis, P., 2021. An application of Sigmoid and Double-Sigmoid functions for dynamic policyholder behaviour. Decisions Econ Finan 44, 5–22.

Cerchiara, R. R., Edwards, M., Gambini, A., 2009. Generalized Linear Models in Life Insurance: Decrements and Risk factor analysis under Solvency II. Working Paper. AFIR Colloquium Rome.

Cribari-Neto, F., Ferrari, S.L.P., 2004. Beta Regression for Modelling Rates and Proportions. Journal of Applied Statistics, 31(7), 799-815.

De Jong P., Heller G.Z., 2008. Generalized linear models for insurance data, Cambridge: Cambridge University Press.

Duan, N., Manning, W.G., Morris, C.N., Newhouse, J.P., 1983. A comparison of alternative models for the demand for medical care. J Business Econ Statist.1:115-126.

Eling, M., Kiesenbauer, D., 2014. What policy features determine life insurance lapse: An analysis of the German market. Journal of Risk and Insurance, (81)2, 241-269.

Haberman, S. and Renshaw, A.E., 1996. Generalised linear models and actuarial science, Journal of the Royal Statistical Society, Series D, 45, 407-436.

Kim, C., 2005. Modeling surrender and lapse rates with economic variables. N. Am. Actuar. J., 9(4), 56–70.

Milhaud X., Loisel S., Maume-Deschamps V., 2010. Surrender triggers in life insurance: what main features affect the surrender behavior in a classical economic context?, Bulletin Francais d'Actuariat (Institute des Actuaires), 11(22).

Nelder J.A., McCullagh, P., 1989. Generalized Linear Models. 2nd edition. Chapman & Hall, London.

Ospina, R., Ferrari S.L., 2010. Inflated beta distributions. Stat Papers 51, 111.

Venables, W.N., Ripley,B.D., 2003. Modern Applied Statistics with S. 4th edition. Springer, New York.