Preprint

Lapse risk in life insurance contracts

E. Barucci, T. Colozza, D. Marazzina and E.Rroji

Abstract We investigate the drivers of lapses in life insurance contracts of a large Italian insurance company. We consider both traditional (with profit or participating) and unit-linked policies. We develop two different types of analyses. First of all, we investigate the determinants of lapse decisions by policyholders looking at microdata on each contract. Then, through a panel study, we include macroeconomic variables. We show that the determinants of lapses for the two types of contracts are quite different. No evidence supporting the Interest Rate Hypothesis, i.e. a positive correlation between interest and lapse rates, is detected. Instead some evidence that a positive correlation of lapse rates with personal financial/economic difficulties, also known as the Emergency Fund Hypothesis, emerges. Some behavioral finance insights come out.

Keywords Lapses · Generalized Linear Models · Survival Analysis · Dynamic Panel

1 Introduction

Life insurance contracts are mostly characterized by market risk, that is the risk related to the investment of funds in financial assets. Non financial risks of a life insurance contract are associated with the possibility that the policyholder decides to withdraw from the contract (lapse risk) and with the death event (mortality risk). While the latter can be modeled through classical actuarial tools, such as mortality tables, lapse risk is more subtle as it builds upon behavioral and psychological motivations that are difficult to model.

In this paper we analyze a unique database provided by a large Italian insurance company over a ten year horizon (2008-2017), a period that includes the recent financial and Euro crisis. We develop two different types of analysis. Firstly, we investigate how some features of the policyholder and of the

Department of Mathematics, Politecnico di Milano, Italy

contract affect the lapse decision. Secondly, we analyze the lapse phenomenon for the company as a whole through a panel analysis by aggregating data at the regional level and using economic information at that level. We analyze both traditional contracts (participating policies) and insurance products with a significant financial component such as unit-linked contracts.

Three main hypotheses on the decision to withdraw from a contract by the policyholder have been investigated in the literature, see [3,11]. The Interest Rate Hypothesis (IRH) relates the lapse decision to financial market conditions and to the value of the contract: the policyholder withdraws from the contract in response to changes in market interest rates, and in particular the lapse rate should positively depend on the interest rate as its increase renders other safe options more interesting and premiums of new policies cheaper. This hypothesis is integrated with the hypothesis that the decision to lapse is driven by the possibility to replace the contract through another contract at better conditions (Policy Replacement Hypothesis, PRH). Economic and liquidity conditions may also affect the decision of a policyholder to withdraw from a life contract. According to the Emergency Fund Hypothesis (EFH), a policyholder gives up the life insurance contract when she is facing a difficult personal economic/financial situation (unemployment, health/long term care problems).

The empirical literature can be classified according to the type of data employed in the analysis, either aggregate (at company or market level), see [7,9,19,21,27], or at policyholder level exploiting micro-data on features that are specific to each policyholder, see [6,10,17,12,13,20,22,24,26,28,31]. The choice of the data set and of the methodology affects the possibility to test the different hypotheses: aggregate data are more suitable to test the IRH and PRH, instead micro-data are more useful to test the EFH, see [11,12,22].

The empirical literature is mostly supportive of the EFH, see [4,7,9,12,13, 17,18,22,25–27,31]. There is also positive evidence on the PRH see [9,12,22, 27], while the evidence supportive of the IRH is more limited, we refer to [7, 20,21]. Differentiating with respect to product type, in [19] the author notices that there is no evidence for the IRH and the EFH in case of traditional policies in the German market while the evidence is more positive when unit-linked policies are taken into consideration.

To these interpretations we have to add recent research on a behavioral interpretation of lapse rates with results referring to cognitive ability and financial literacy of policyholders showing that being financially more literate leads to more rational decisions, see [20,25,26], and papers that investigate how qualitative features of policyholders (retirement, divorce, new house, death of a spouse, acquisition of a house) affect the lapse decision mostly confirming the EFH, see [6,10,12,13,17,22,25].

As concerns the Italian market, [6] considers a data set of a large Italian bancassurer. The authors concentrate their analysis on three factors (calendar year, time passed from the date of stipulation - named as contract age or anti-duration and product class). Our aim is to identify the determinants of lapses in the Italian market working on a database provided by a large Italian insurance company. We work both on data concerning the policyholder and on data aggregated at the regional level exploiting regional macroeconomic data to investigate the various hypotheses. Our paper adds to the literature investigating the phenomenon by considering the financial-Euro crisis period and exploiting the variability at the regional level which has not been investigated in previous studies. At the micro-level we rely on a full set of information about the contract including product type, policyholder age, contract age, professional employment, premium payment type.

Our analysis shows that the lapse rate significantly depends on the type of contract. The lapses of traditional products are weakly sensitive to financial turbulence, are more likely to occur in case of a small insured capital and of a short contract age. These results suggest that traditional life insurance contracts are likely to be interpreted as short term safe assets, instead unit-linked contracts can be assimilated to a financial product as their lapse rate is sensitive to financial turbulence, is higher in case of a large insured capital and of a long contract age. For both types of product we show that there is no evidence in favor of the IRH. Instead there is some evidence supportive of the EFH.

There is also some evidence that can be interpreted according to behavioral finance. The behavior of policyholders of traditional products significantly depends on the frequency of the premium payment. Traditional products with a single payment are assimilated to a short term safe investment and the policyholder withdraws from the contract if inflation is high and the return of the contract is high compared to a safe asset. Instead, in case of policies with a recurrent payment, the policyholder is only influenced by the time evolution of the yield of the policy: he is disappointed by a bad performance and this induces him to withdraw from the contract.

The behavioral interpretation provides some interesting insights also in case of withdraw from *young* contracts (early lapses, contracts not older than three years). We show that early lapses are not influenced by economic conditions and the performance of the life insurance product, the decision to lapse is only influenced by features of the policyholder and of the contract. Confirming this evidence, the volatility of the return on lapsed contracts is high for contracts stipulated few years before.

The paper is organized as follows. In Section 2 we provide a description of the data set and highlight the main descriptive features of the phenomenon. In Section 3 we briefly discuss the methodologies considered for the statistical analysis of the lapse phenomeon. In Section 4, we provide the results of our analysis on micro-data. In Section 5 we consider a panel study at the regional level. In Section 6 we investigate the lapse phenomenon for contracts having a short life. We conclude with a final discussion in Section 7.

2 Data set and preliminary results

The data set analyzed in this paper has been provided by an Italian large insurance company. The analysis covers the time period 2008-2017 and refers to existing and newly issued contracts. At the beginning of the observation period there are contracts that have been stipulated from 1985 onwards. The entire data set comprises 4.93×10^6 observations with more than 1.18×10^6 contracts. Capital insured by policyholders increased up to 2011 and then remained almost constant, instead the number of contracts decreased through time.

The data set consists of life insurance contracts: traditional contracts account for about 60.2% of the total number of contracts, while the remaining 39.8% are unit-linked contracts. In terms of insured capital, traditional contracts account for nearly 73.5% of the total capital. In particular, the relative importance of traditional contracts steadily increased till 2012, then the pace reduced. By the end of the period considered in our analysis, less than 25% of the insured capital comes from unit-linked products.

The descriptive analysis shows that the lapse rates of the two classes of products differ significantly, see Figures 1-4. The lapse rate of unit-linked contracts is much higher than that of traditional products. Moreover, the lapse rate of unit-linked contracts is much more volatile than that of traditional products: it goes up during the turbulence period in the Euro area (2010-2012) and in 2016-2017 (a period associated with high political uncertainty). Instead, the lapse rate of traditional products is more stable with a surge only in 2017.

The company provided a unique serial number in order to allow processing data by contract. As concerns policyholder's features, the available information is made up of: i) gender (56% men and 44% women), ii) age at contact entry (and therefore at the time of the observation), iii) region of residence and iv) profession. Information on the profession is available only for 40% of the contracts. Contract features include[1]: i) product type (traditional and unit-linked contracts), ii) date of contract inception, iii) insured capital, iv) contract age, iv) premium frequency - single premium contracts are separated from recurrent premium contracts that are characterized by the number of installments for the annual recurrent premium (once a year, twice a year, quarterly and monthly as well as three and six times a year) - and v) fund yield. The only contract feature that changes during the lifetime of a contract is the contract age/anti-duration, i.e., the years of the contract since its inception. Similarly, age of policyholders changes in each calendar year.

As partial requests of capital reimbursement are negligible in terms of insured capital, in the definition of the lapse event we only consider contract termination for the full amount of the contract.

[1] Information on the distribution channel is not fully available.

2.1 Descriptive analysis

Analyzing the information provided in the database on contracts and policyholders, the following results are obtained on explanatory variables.

1. `Calendar year`. In the Italian market a calendar year effect is observed in [6] with a steady increase in the lapse rate over the period 1991-2007. An increase over time is also observed for the German market in [10] but the lapse rate tends to decrease after the financial crisis. This observation is consistent with the EFH: in a crisis period policyholders opt to lapse a life contract because of liquidity needs. In our data set we observe that the lapse rate is almost stable in case of traditional products while unit-linked contracts are more volatile with an increase up to 2011-2012, followed by a decrease and again an increase in the last three years of the sample, see Figure 1. The result agrees with the EFH: while the first peak is associated with the Euro crisis, the most recent increase is associated with uncertainty in the Italian political environment.

2. `Age of policyholder`. It is not easy to predict the relation between age of the policyholder and the decision to lapse. Referring to the EFH, we observe that on one hand younger households are more exposed to liquidity shocks (unemployment) but, on the other hand, old households are more exposed to health and long term care shocks. In [10, 12, 13, 20, 24] the authors detect a positive relationship between the policyholder's age and the decision to lapse, with a decrease when the person is very old, no effect of age is observed in [25, 31]. In our data set the descriptive analysis shows a more complex relation with a two hump shape: young people and old policyholders tend to lapse more than very young and middle age policyholders, see Figure 2. The first hump is due to policyholders of traditional contracts while the second one is mostly associated with holders of unit-linked contracts (there is also a less pronounced hump for young policyholders of unit-linked contracts). The lapse rate of the latter contracts tends to increase with the age of policyholders. Notice that the lapse rate of traditional contracts shows a significant volatility for young and for old policyholders, instead in case of unit-linked contracts the volatility is high for old policyholders and is higher than in case of traditional contracts except for young policyholders.

3. `Contract age (anti-duration)`. Contract age is a relevant variable in all the empirical analysis looking for determinants of lapses. In [10, 15, 24, 28] it is observed that the lapse rate is at the highest level for contracts stipulated few years before and then steadily decreases with contract age. A reversed U-shape is observed for the Italian market in [6] with a peak at two and five years. In our data set we observe a similar shape with a peak for a contract age between four and ten years, see Figure 3. A deeper analysis based on contract type shows that traditional contracts are characterized by a lapse rate declining in the contract age. Instead, the lapse rate of unit-linked contracts is characterized by a more pronounced hump shape. The early lapse phenomenon (contract age between one and three years) seems to be a peculiarity of traditional products.

4. `Policy size`(insured capital) The size of the policy is an interesting explanatory variable because it can be related to the EFH. As a matter of fact, a person holding a contract for a large insured capital is likely to be wealthy and therefore he is less likely to lapse for liquidity reasons. According to this interpretation, the lapse rate should be higher in case of contracts with a small insured capital. No evidence confirming this hypothesis is obtained in [20], positive evidence is provided in [15]. A look at our data set confirms the hypothesis only for traditional contracts, see Figure 4. In case of unit-linked contracts, the relationship between lapse rate and insured capital is slightly increasing.

5. `Gender`. On the effect of the gender of the policyholder on the lapse rate, [6,31] find no evidence, instead [10,25] observe that women tend to lapse less than men. In our data set, the average lapse rate for women is 11.24% while is 12.34% for men.

6. `Product type`. The lapse rate may differ depending on the product type. In [10] the authors find that the lapse rate does not vary much across product categories while in [6,28] the authors find a strong variability, in particular the latter find that the lapse rate for unit-linked products is significantly higher than for other products. At first glance this result is confirmed in our data set: although traditional contracts represent the majority of contracts in our sample (60%), the number of lapses for unit-linked contracts (68%) is much higher than for traditional products (32%). As already noticed, we also observe a different lapse rate evolution through time for two classes of contracts.

7. `Premium frequency`. As far as premium payment is concerned, [10,24] observe that the lapse rate is higher in case of contracts with single premium than in case of recurrent premium contracts. In our dataset we observe that the average lapse rate for single premium contracts is 12.43%, while for the other contracts is 11.11%.

3 Methodology and model selection

Two main methodologies are used in the literature to study the determinants of lapses in life insurance contracts.[2] The first approach is based on survival analysis, while the second one is based on Generalized Linear Models (GLMs).

These methodologies answer slightly different research questions. To investigate the probability that a policyholder decides to withdraw from a contract, we may use a GLM with binomial errors built as a generalization of a logistic regression widely applied in studies with binary response variables. Instead, a GLM with Poisson errors is useful to analyze the expected number of lapses. Finally, the survival analysis is well suited to understand how long will it take for a policyholder to lapse. The latter approach takes into account the fact that data are right censored as a policyholder may decide to terminate the contract beyond the calendar year 2017, when our sample ends.

[2] In this paper we do not investigate the effect of lapse/surrender rates in the premium computation as done for example in [1].

Classical linear models make a set of restrictive assumptions, most importantly, the dependent variable should be normally distributed conditioned on exogenous variables with a constant variance. GLMs accommodate dependent variables that violate these assumptions and allow for a nonlinear dependence between the dependent variable Y and the exogenous variables X, see [23].

The most well-known special case is the logistic regression, i.e., GLM with binomial errors, where the dependent variable takes only two values 0 and 1. In our setting, $Y = 1$ refers to *lapse* of the contract while $Y = 0$ to *non lapse*. In a logistic regression, the quantity to be estimated is the lapse *odds ratio* conditional to a set of exogenous variables X:

$$Odds(X) := \frac{\mathbb{P}(Y = 1|X)}{\mathbb{P}(Y = 0|X)} = \exp(\beta_0 + \beta X). \quad (1)$$

In the GLM with Poisson errors, the number of lapses Y is modeled through a Poisson distribution and a linear relation between the logarithm of the number of conditional expected lapses and the explanatory variables is considered:

$$\mathbb{E}(Y|X) = \exp(\beta_0 + \beta X).$$

Observe that in the GLM with binomial errors the coefficient β_0 identifies the *odd* of the reference class ($X = 0$), while in the GLM with Poisson errors the same parameter identifies the expected number of lapses for $X = 0$.

An example of survival models is the constant Proportional Hazard (PH) model proposed in [8] where the hazard function $h(t, X)$, defined as the instantaneous rate at which the lapse event occurs conditioned on the fact that lapse has not occurred up to time t, is modelled as:

$$h(t, X) = h_0(t)e^{\sum_{i=1}^{k} \beta_i X_i},$$

where $h_0(t)$ is the baseline hazard function. The model is *proportional* because the hazard ratio $\frac{h(t,X)}{h_0(t)}$ is time independent. In constant PH models, the output is provided not only by the lapse event but also by the time to event. For this reason we do not include contract age among the explanatory variables as in the case of GLMs.

Exploiting the information provided in the database on contracts and policyholders and referring to the literature on lapse decisions, we consider all the variables presented in Section 2.1 to estimate a binomial, a Poisson or a PH model. Therefore, the set of explanatory variables include: calendar year, age and gender of policyholder, product type, insured capital, premium frequency, and contract age (also known as anti-duration). Exploiting the information contained in our data set, we also consider as exogenous variable the macro-region where the policyholder lives (North-East, North-West, Center, South and Islands) and her profession.

In order to estimate the parameters of the regression equations, we need to represent the exogenous variables through a finite set of classes. Variables like product type and gender are already dichotomous, other variables, like for example premium frequency and profession, are easy to regroup in a finite number of classes. In particular, for premium frequency it is reasonable to distinguish only between single (at the inception of the contract) and recurrent payments: 53.4% of the contracts contained in the data set are characterized by a single premium, the remaining 46.6% present multiple payments. As far as profession is concerned, information is rather noisy and it only covers half of the data set. We distinguish between employee (fixed salary) and other jobs. It is reasonable to guess that the lapse decision by a policyholder endowed with a fixed salary should be less affected by the EFH. Variables like year and macro-region are left as in the original data set with respectively ten and four classes.

For continuous variables like insured capital, age and anti-duration a preliminary study is necessary in order to identify a restricted number of classes that are homogenous in terms of lapse rates. We start from the boxplot of lapse rates according to the variables. From Figure 2, it emerges that there are two peaks for the lapse rate of young people and of people in proximity of retirement age. We define four classes for the age as follows: Age 1= $\{0, \ldots, 34\}$, Age 2= $\{35, \ldots, 49\}$, Age 3= $\{50, \ldots, 69\}$ and Age 4 = $\{70, \ldots, 115\}$. Figure 3 contains the boxplot of lapse rates for contract age, we define the following classes AD1= $\{1, 2, 3\}$, AD2= $\{4, 5, 6, 7\}$, AD3= $\{8, 9, 10, 11\}$ and AD4= $\{12, \ldots, 32\}$. The lapse rate significantly varies depending on the insured capital, so we define six capital classes based on the empirical quantiles of the aggregated data set.

Lapse rates across regions do not display great differences. For all regions, we have that the lapse rate increases in 2011-2012 and in 2017.

The exogenous variables and the classes adopted in our analysis are defined in Table 1. To estimate the models we have to define a reference class. For each exogenous variable, we opt to define as reference class the one for which we have more observations in the data set.

Variables	Classes
Product type	**Traditional** (T) and Unit Linked (UL).
Gender	**Male** (M) and Female (F).
Age	Age 1= $\{0, \ldots, 34\}$, Age 2= $\{35, \ldots, 49\}$, **Age 3**= $\{50, \ldots, 69\}$ and Age 4 = $\{70, \ldots, 115\}$
Calendar Year	**2008**, ..., 2017
Region	North-East, **North-West**, Center, South and Islands.
Premium frequency	**Unique** (U) and Periodic (P)
Capital categories	cap1 = (0, 284], cap2 = (284, 4 875], cap3 = (4 875, 11 932], **cap4** = (11 932, 29 494], cap5 = (29 494, 109 522], cap6 = (109 522, 32 020 328]
Anti-duration	**AD1**=$\{1, 2, 3\}$, AD2=$\{4, 5, 6, 7\}$, AD3=$\{8, 9, 10, 11\}$ and AD4=$\{12, \ldots, 32\}$.
Profession	**Employee** (E) and Other jobs (O)

Table 1 Variables used in the analyses, classes and reference classes (in bold).

4 Micro data analysis

As the profession is available only for 60% of the observations, we start considering the full data set without including it among the exogenous variables. In Table 2 we provide the parameter estimates considering the full sample (traditional and unit-linked contracts) and the binomial, Poisson and constant PH model.

The estimated parameters change depending on the model but the sign is always preserved except for the effect of calendar year 2014. It is important to notice that with few exceptions almost all the variables are statistically significant at the 1% level. In the binomial model, the exception is provided by the anti-duration class AD4, in the Poisson model by the capital class cap1. In the constant PH model all variables are significant at 1% level.

The main findings of the analysis are the following.

`Product type`. The lapse rate of a unit-linked product is about 300% higher than that of traditional contracts. This result confirms the descriptive analysis and suggests that the two products are interpreted in a different way by policyholders.

`Gender`. Women tend to lapse less than men. This result is consistent with the literature showing that women are more risk averse than men and tend to make more consistent and conservative financial decisions, see [2,16,30].

`Age`. The highest lapse rate is observed in case of young people (age smaller than 34) and of very old people (age above 70). The models confirm the two hump shape detected from a pure descriptive point of view. This result can be interpreted through the EFH as young and old people are more likely to experience an emergency liquidity need.

`Calendar Year`. The two GLMs suggest that the lapse rate is higher compared to the reference level 2008, except for the calendar year 2014. In particular, we observe that the lapse rate steadily increases from 2008 to 2012. The trend reverses for two years (2013 and 2014) before a new upward trend period with maximum level attained in 2017. In the constant PH model, the estimated parameters are always positive but the time evolution confirms what established by the other two methods. This result confirms the hypothesis that policyholders withdraw from their contracts when they experience turbulence in financial markets or political uncertainty is high.

`Region`. The reference region (North-West) is characterized by the highest lapse rate. Note that the North-West region in Italy is the richest one and therefore the result is at odds with the EFH. While the GLMs suggest that lapse rates for South and Islands are lower compared to the reference region[3], the positive sign of the coefficient in the constant PH model implies that the hazard ratio increases for this region and the length of contract survival decreases. We recall that GLMs focus only on the lapse event differently from the constant PH model that aims at predicting the time to lapse.

[3] In both cases we have a negative coefficient.

`Premium frequency`. The lapse rate is higher in case of contracts with single premium (at the inception of the contract) than in case of recurrent premium contracts.

`Capital`. The lapse rate of contracts with a large insured capital is smaller than the rate of contracts with a small insured capital. This evidence supports the EFH as policyholders holding a contract for a large insured capital tend to be wealthy.

`Anti-duration`: Policyholders of a contract with age belonging to the second class (AD2, contract age between 4 and 7 years) are characterized by the highest lapse rate, while the lapse rate tends to decrease as contract age increases.

As noticed in the descriptive analysis, traditional and unit-linked contracts seem to behave in a quite different way. In what follows, we investigate the lapse phenomenon for the two classes of contracts. In Tables 3 and 4 we provide the parameter estimates of the models.

In both cases we observe that women tend to lapse less than men. As far as the age profile of policyholders is concerned, in case of traditional contracts we observe a two hump shape with the first hump (below 34 years) more pronounced than the second one (above 70 years), instead in case of unit-linked contracts we observe that the lapse rate is increasing in the age of the policyholder (with a maximum for policyholders above 70 years). As far as regions is concerned, policyholders of traditional products living in the South and Islands region tend to lapse more than the others, instead it is confirmed that policyholders of unit-linked contracts in the North-West tend to lapse more frequently (whereas those living in the South and Islands lapse less frequently compared to the average).

The calendar year effect differs depending on the type of policy. The 2017 year is characterized by the highest lapse rate for both types of contracts, the phenomenon is particularly relevant for traditional contracts. In case of traditional contracts, the lapse rate is always higher than the level observed in 2008 with weak evidence of a pattern associated with financial turbulence (higher values in 2011-2012, in 2015 and afterwards). Observe that the pattern is much more pronounced in case of unit-linked contracts with the lowest lapse rate observed in 2014 (even lower than in 2008).

The most striking result concerns the change of sign for the premium frequency coefficient with respect to the full sample. Performing separately the analysis on the two contract types, we conclude that policyholders of contracts with recurrent premium payments are characterized by a higher lapse rate than those holding contracts with single initial premium. This result is at odds with the results presented in Table 2 for the full sample. The counterintuitive result can be interpreted through the Simpson's paradox, see [14]: a positive effect is associated to a single premium payment contract type in the full sample, but then the effect is reversed when the two subsets are considered separately. The root cause for the paradox is that the subsamples are characterized by different levels of lapse rates. The result obtained considering the two subsamples can be interpreted as a behavioral evidence that the policyholder is more likely to lapse in case she has to think about paying in a premium (in case of the reference class there is a single premium at the inception of the contract). The change of sign for this variable also occurs switching from GLMs to survival models, in this case the root cause is the absence of anti-duration among exogenous variables.

The two types of contracts also differ with respect to the insured capital. In both cases the pattern is hump shaped, but in case of traditional products the maximum is observed for a relatively low level of insured capital (cap2), instead in case of unit-linked contracts the highest lapse rate is observed for a large insured capital (cap5).

As far as the anti-duration of a contract is concerned, traditional contracts are characterized by a high lapse rate in case of contracts stipulated few years before and then the rate decreases, instead old unit-linked contracts are characterized by a higher lapse rate than contracts stipulated few years before.

In Table 5 we report the results obtained including profession as an explanatory variable. As the data set is significantly smaller than the original one, we concentrate on the effect of the profession type. In agreement with the EFH, employed policyholders show a lower lapse rate than other policyholders (self employed, unemployed, retired people). The rationale of this result is that employed people tend to suffer in a less significant way of emergency liquidity needs and, therefore, according to the EFH they should lapse in a less significant way. The main results for the other variables are confirmed. If we compare the estimated parameters of the models with Table 2 we observe that the age variable becomes less significant, there is a greater impact of the variable premium frequency and contracts with small insured capital have lower lapse rates compared to the reference level.

5 An aggregate analysis of lapse rates at regional level

In this section, we analyze the lapse rates at the regional level. Analyzing the aggregate lapse rate through the time dimension (year) and the longitudinal dimension (region) we can investigate the IRH, PRH, EFH refining the analysis developed at the policyholder/contract level. In the analysis presented in the previous section, we were able to investigate a limited set of information on the contract/policyholder without considering financial and economic conditions. This was a strong limit of the analysis. The aggregate analysis at the regional level allows us to overcome it.

We partition Italy in 20 regions, defined as the second level administrative division in the NUTS (Nomenclature des Unités Territoriales Statistiques) regions of the European Commission definitions, denoted as *Regioni* in the Italian institutional system. We construct a longitudinal dataset $z_{i,t}$ of lapse rates of the company in region i ($i \in \{1, \dots, 20\}$) and year t ($t = 2008, \dots, 2017$) by dividing the total number of lapses in region i at time t by the total number of outstanding contracts in that region at time t (`lapse`).

Since the lapse rate at time t is likely to be affected by its lagged value $z_{i,t-1}$, we consider a dynamic panel regression of the form:

$$\begin{aligned} z_{it} &= \alpha_0 + \alpha_1 z_{i,t-1} + \beta x_{i,t} + \nu_{i,t} \\ \nu_{i,t} &= u_i + \varepsilon_{i,t} \end{aligned} \tag{2}$$

where $x_{i,t} = (x^1_{i,t}, \dots x^k_{i,t})$ is a vector of suitable exogenous regressors, $\{u_i\}$ are the individual fixed effects and $\{\varepsilon_{i,t}\}$ represent the set of identically and independently distributed gaussian disturbances.

The dynamic panel framework accounts for regional diversities, and allows us to exploit the longitudinal dimension of the data set so as to partially compensate the lack of data in the time dimension. All data come from the Eurostat database.

In order to investigate the EFH, we consider variables describing economic and social conditions at the regional level. As a basic set of exogenous regressors, we consider the inflation rate (`hcpi`) the growth rate of disposable income (`disp_inc`), the growth rate of the number of not employed people (`unem`) and the first variation of the percentage of people under the poverty threshold (`poverty`). According to the EFH, an increase in `unem`, `poverty` or `hcpi`, or a decrease in `disp_inc` are likely to positively affect the lapse rate.

Among exogenous variables we include the return of the European stock index (`stoxx`) so as to account for the economic and political instability caused by the latest financial crisis[4]. Notice that this variable also provides us with a proxy of an alternative investment opportunity for policyholders. The effect of this variable on the lapse rate is ambiguous. On one hand, a high return for the European stock index signals a good shape of financial markets and therefore a good confidence (risk appetite) of investors, this may lead policyholders to lapse to profit from other (risky) opportunities; on the other hand, a high stock index return may signal an overvalued market, inducing policyholders not to invest in risky assets, and provides a wealth effect inducing policyholders not to withdraw from the contract for liquidity needs (EFH).

In order to investigate the IRH and the PRH, we include among the set of exogenous variables the interest rate on sovereign bonds (`interest`), i.e., the yield of the 5-year Italian government bonds. Notice that the 5-year maturity is coherent with the average duration of life insurance policies. According to the IRH and the PRH, policyholders are likely to lapse when an alternative safe hypothesis, like government bonds and insurance contracts, provide an interesting alternative investment opportunity. This implies that the lapse rate should be increasing in `interest`.

We also test whether the lapse rate is affected by the performance of traditional contracts and of unit-linked contracts. To this end, we consider the yield of the contract (`yield`), which is defined as the yield associated to traditional contracts, and as the growth rate of the net asset value of unit-linked contracts, respectively. As an alternative performance indicator, we consider

[4] We introduced the European stock index and not the Italian one because of strong positive correlation between the contract yields and the Italian stock index return.

the first variation of the contract yield (Δyield) and the convenience yield (conv_yield), defined as the difference between yield and interest. Note that stoxx, interest, yield, conv_yield do not vary across regions[5].

The estimator is the two-step GMM estimator of Blundell and Bond with orthogonal deviations transform and standard errors corrected so as to account for regional heteroscedasticity of the disturbances, see [5][6].
As the analysis in the previous section showed that traditional contracts and unit-linked contracts have different features, in Tables 7 and 8 we report the regression results for the two products, separately. For the sake of completeness, we also provide the analysis considering both types of contracts together, see Table 9. In this case we have a panel of 40 "individuals", as each region defines two different individuals through the two different contract types.
We notice that there is no evidence supporting the IRH. In all the models the coefficient associated with interest is negative, and is statistically significant in case of unit-linked contracts: at odds with the IRH and the PRH, the lapse rate is negatively affected by the yield of government bonds. The interpretation is that a high return of the yield of Italian government bonds incorporates a credit spread which leads risk averse policyholders to hold on life insurance contracts. However, we notice that results do not change substituting the 5y Euribor rate to interest in the panel regression. This result agrees with the observation that the lapse rate of life insurance contracts, and in particular of unit-linked contracts, is high in the middle of the Euro crisis. We may also interpret the results as showing that in a low interest rate environment policyholders' decisions are not driven by the IRH.
There is also some slight evidence that an increase in the stock index reduces the lapse rate, which however depends on the specification of the model. This evidence can be interpreted according to the EFH, since an increase in the stock market discourages policyholders from withdrawing a life contract to face liquidity needs.
The most surprising result concerns the yield of life contracts. The decision to withdraw from a traditional contract does not depend on its yield or on its spread over the yield of government bonds but depends on the yield's variation. Namely, a policyholder tends to lapse if the yield of the contract decreased compared to the previous year. On the contrary, no relationship with the performance of the contract is obtained for unit-linked contracts. These results

[5] We tested the robustness of the results of this model specification in several directions. We substituted the growth rate of the regional GDP to the growth rate of the disposable income. As an alternative metric of social conditions, we also considered the average educational level and the average life expectancy of the population. The main results are not affected.

[6] We tested for the presence of unit root for both endogenous and exogenous variables by using a Fisher-type unit root test for panel data based on Phillip-Perron univariate tests. None of the variables in the panel displays unit root with confidence level of less than 1%. We address multicollinearity issues using the Variance Inflation Factor criterion (VIF); results show an average VIF hovering between 2 and 3, which is below the threshold of 10 that would roughly indicate possible multicollinearity issues. The estimates are conducted with the Stata13 routine *xtabond2*: for a detailed explanation on the use of the two step GMM estimator see [29].

suggest that the performance of the contract is not a significant determinant to induce the policyholder to lapse, we only observe a reaction of the policyholder to a bad intertemporal performance in case of traditional policies. Both the yield and the convenience yield of the contract negatively affect the lapse rate on the full sample, but this result may be spurious as observed previously because of the different features of the data sets.
As far as the EFH is concerned, the analysis provides mixed evidence. No effect is associated with unemployment and poverty rate at the regional level. Coherently with the EFH, the lapse rate of traditional products is positively affected by the inflation rate, instead there is a positive evidence that the growth rate of disposable income negatively affects the lapse rate of unit-linked contracts.
We further develop our analysis considering the aggregate lapse rate of contracts with different features: gender and premium frequency.
The analysis obtained considering only male or female policyholders does not change significantly with respect to the one showed above in Tables 7 and 8 for traditional and unit-linked contracts. The only exception is provided by the analysis on unit-linked for female policyholders, in this case no variable turns out to be statistically significant. For the sake of brevity we omit to present the results.
Interesting results emerge partitioning the data set depending on the premium frequency. Adopting a behavioral finance perspective, the decision to lapse on the contract may depend on the fact that the policyholder has to pay or not a premium during the life of the contract. In the case of recurrent premiums it is more likely that she will put more attention to the return of the contract and the performance of the contract may affect her decision. Considering the lapse rates of contracts providing for a single premium or a recurrent premium, we observe some differences with respect to the analysis obtained for the full data set, see Tables 10 and 11 for traditional products and Table 12 for unit-linked.
Concerning traditional contracts with single premium (at the inception of the contract), the positive impact of inflation on lapses is around 2.5 time the coefficient estimated for the full data set, see Table 7. Furthermore, Δ`yield` is no more significant, while `interest` becomes significant at 1% or 5% level. The effect of the convenience yield becomes positive and statistically significant, i.e., a policyholder tends to lapse a traditional product which yields a positive performance. In case of recurrent premium payment the inflation rate is statistically significant only in one of the regressions while `interest` is never statistically significant. The only variable that is statistically significant at 1% is the time variation of the yield of the contract (Δ`yield`).
The difference between the two tables is striking. Results suggest an interesting interpretation. Traditional products with a single payment are assimilated to a safe investment and the policyholder withdraws from the contract if inflation is high, the return of the contract is high compared to government bonds, and there is no turbulence in financial markets (credit risk associated with Italian government bonds is low). In the case of contracts with a recurrent payment, the policyholder is influenced by the time evolution of the yield of the contract. As the policyholder has to pay in a premium during the life of the contract he

is more likely to be influenced by the time evolution of its performance: if the performance of the contract deteriorates over time, then the policyholder will be disappointed and he will withdraw from the contract.
In case of unit-linked contracts with a single premium, the results are in line with those presented in the general case of Table 8, with few slightly minor changes appearing only in the model with `yield`. We do not present the results for unit-linked contracts with recurrent premium payment because the model is not reliable as unit root in the lapse rate emerges[7].

6 The early lapse phenomenon

One of the striking results of our analysis is the high rate of lapses for policyholders detaining a contract stipulated few years before. In what follows we concentrate our attention on *early lapses*, i.e., contracts lapsed in the first three years of life. A policyholder lapsing so quickly is likely to do a choice either because of urgent needs (EFH) or because he has done a wrong choice. We expect the determinants of early lapses to be different from those highlighted in the full sample analysis in Table 2.

In Table 6 we present the results on lapses in the first three years of the contract. Comparing the results with respect to those obtained for the full sample, we observe that there is not a great difference in terms of lapse rates between product types while the gender effect is more evident in early lapses. It emerges that younger policyholders are more prone to withdraw from a contract in the first years than in the case of all contracts. As far as the calendar effect is concerned, early lapses seem to be less frequent during the turbulence of the financial crisis and afterwards. As far as the geographic region is concerned, the most relevant difference is that early lapses are more frequent in the South and in Islands. As concerns payment frequency, the two GLMs agree that it is less likely to observe a lapse in the first years of contract's life for contracts with single payment. The survival based analysis suggests the opposite. The most important difference with the full sample analysis concerns the insured capital: the probability of lapse on contracts with small and medium capital is much higher than in the full sample.

When considering early lapses, we notice that the panel regression models at the regional level poorly perform in explaining lapse rates, both in the case of traditional and unit-linked contracts, see Tables 13 and 14. In the case of traditional contracts, the F-statistic does not reject the null hypothesis that all regression coefficients are equal to zero at 1% level. This means that macroeconomic and financial variables are not able to explain the early lapse phenomenon. We retain that early lapses are mostly explained by individual characteristics, as they do not appear to be affected by financial factors. In case of unit-linked contracts, no economic/financial variable turns out to be

[7] We test the presence of unit roots using a Fisher-type unit root test based on Phillip-Perron tests; the test is robust to an increase in the number of lags.

statistically significant. Confirming that the financial performance of the contract does not affect the lapse decision, the return on lapsed contracts shows a significant volatility for early lapses, see Figure 5. Instead, results on lapses for contracts older than three years are in line with the results obtained for the full sample displayed in Table 7 and 8. We omit the results for the sake of brevity.

7 Conclusions

We have analyzed the lapse rate phenomenon considering a large data set of life insurance contracts of an Italian insurance company. We have provided both a micro level analysis and an aggregate analysis considering the lapse rate at the regional level.

First of all, we have noticed that the lapse rates of traditional (participating) contracts and the lapse rates of unit-linked contracts have very different features: the lapse rate of a unit-linked product is much higher than the lapse rate of traditional contracts; in case of traditional contracts we observe a higher lapse rate for young policyholders, instead for unit-linked we observe that the lapse rate is increasing in the age of the policyholder. In the case of traditional contracts, the lapse rate is always higher than the level observed in 2008 with weak evidence of a pattern associated with financial turbulence, instead the intertemporal pattern is much more prominent in case of unit-linked contracts. As far as the insured capital is concerned, in case of traditional products the highest lapse rate is observed for a relatively low level of insured capital, instead in case of unit-linked contracts the highest lapse rate is observed for a large insured capital; traditional contracts are characterized by a high lapse rate in case of contracts stipulated few years before and then the lapse rate tends to decrease, instead old unit-linked contracts tend to experience a higher lapse rate than contracts stipulated few years before.

Acknowledgements The work in the paper has been partially supported from the European Union's Horizon 2020 training and innovation program "FIN-TECH", under the grant agreement No. 825215 (Topic ICT-35-2018, Type of actions: CSA)

References

1. Bacinello, A.R.: Fair valuation of a guaranteed life insurance participating contract embedding a surrender option. Journal of Risk and Insurance **70**(3), 461–487 (2003)
2. Barber, B.M., Odean, T.: Boys will be boys: Gender, overconfidence and common stock investment. Quarterly Journal of Economics **116**, 261–292 (2001)
3. Bauer, D., Gao, J., Moenig, T., Ulm, E., Zhu, N.: Policyholder exercise behavior in life insurance: The state of affairs. North American Actuarial Journal **21**(4), 485–501 (2017)
4. Belaygorod, A., Zardetto, A., Liu, Y.: Bayesian modeling of shock lapse rates provides new evidence for emergency fund hypothesis. North American Actuarial Journal **18**(4), 501–514 (2014)

5. Blundell, R., Bond, S.: Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics **87**, 115–143 (1998)
6. Cerchiara, R.R., Gambini, A., Matthew, E.: Generalized linear models in life insurance: Decrements and risk factor analysis under Solvency II. Giornale dell'Istituto degli Attuari **72**, 100–122 (2009)
7. Changki, K.: Modeling surrender and lapse rates with economic variables. North American Actuarial Journal **9**, 56–70 (2005)
8. Cox, D.: Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological) **34**(2), 187–202 (1972)
9. Dar, A., Dodds, C.: Interest rates, the emergency fund hypothesis and saving through endowment policies: some empirical evidence for the U.K. The Journal of Risk and Insurance **56**(3), 415–433 (1989)
10. Eling, M., Kiesenbauer, D.: What policy features determine life insurance lapse? An analysis of the German market. The Journal of Risk and Insurance **81**(2), 241–269 (2014)
11. Eling, M., Kochanski, M.: Research on lapse in life insurance-what has been done and what needs to be done? The Journal of Risk Finance **14**(4), 392–413 (2013)
12. Fier, S., Liebenberg, A.: Life insurance lapse behavior. North American Actuarial Journal **17**(2), 153–167 (2013)
13. Gemmo, I., Gotz, M.: Life insurance and demographic change: an empirical analysis of surrender decisions based on panel data. ICIR working paper (2016)
14. Good, I.J., Mittal, Y.: The amalgamation and geometry of two-by-two contingency tables. The Annals of Statistics **15**(2), 694–711 (1987)
15. Gottlieb, D., Smetters, K.: Lapse-based insurance. Working Paper (2014)
16. Guiso, L., Sodini, P.: Household Finance: An emerging field. pp. 1397 – 1532. Elsevier (2013)
17. Hanming, F., Kung, E.: Why do life insurance policyholders lapse? the roles of income health and bequest motive shocks. NBER workign paper nr.17899 (2012)
18. Jiang, S.J.: Voluntary termination of life insurance policies: evidence from the U.S. market. North American Actuarial Journal **14**(4), 369–380 (2010)
19. Kiesenbauer, D.: Main determinants of lapse in the German life insurance industry. North American Actuarial Journal **16**(1), 52–73 (2011)
20. Knoller, C., Gunther, K., Schoenmaekers, P.: On the propensity to surrender a variable annuity contract: an empirical analysis of dynamics policyholder behavior. The Journal of Risk and Insurance **83**(4), 979–1006 (2016)
21. Kuo, W., Tsai, C., Chen, W.K.: An empirical study on the lapse rate: the cointegration approach. The Journal of Risk and Insurance **70**(3), 489–508 (2003)
22. Liebenberg, A., Carson, J., Dumm, R.: A dynamic analysis of the demand for life insurance. The Journal of Risk and Insurance **79**(3), 619–644 (2012)
23. McCullagh, P., Nelder, J.: Generalized Linear Models. Chapman and Hall / CRC, London (1989)
24. Milhaud, X., Loisel, S., Maume-Deschamps, V.: Surrender triggers in life insurance: what main features affect the surrender behavior in a classical economic context? Bulletin Français d'Actuariat **11**(22), 5–48 (2011)
25. Mulholland, B., Finke, M.: Does cognitive ability impact life insurance policy lapsation? Working paper (2014)
26. Nolte, S., Schneider, J.: Don't lapse into temptation: a behavioral explanation for policy surrender. Journal of Banking and Finance **79**, 12–27 (2017)
27. Outreville, J.F.: Whole-life insurance lapse rates and the emergency fund hypothesis. Insurance Mathematics and Economics **9**(4), 249–255 (1990)
28. Renshaw, A., Haberman, S.: Statistical analysis of life insurance lapses. Journal of the Institute of Actuaries **113**, 459–497 (1986)
29. Roodman, D.: How to do xtabond2: An introduction to difference and system gmm in stata. The Stata Journal **9**(1), 86–136 (2009)
30. Sapienza, P., Zingales, L., Maestripieri, D.: Gender differences in financial risk aversion and career choices are affected by testosterone. Proceedings of the National Academy of Sciences (2009). DOI 10.1073/pnas.0907352106
31. Sirak, A.: Income and unemployment effects on life insurance lapse. Working paper (2015)

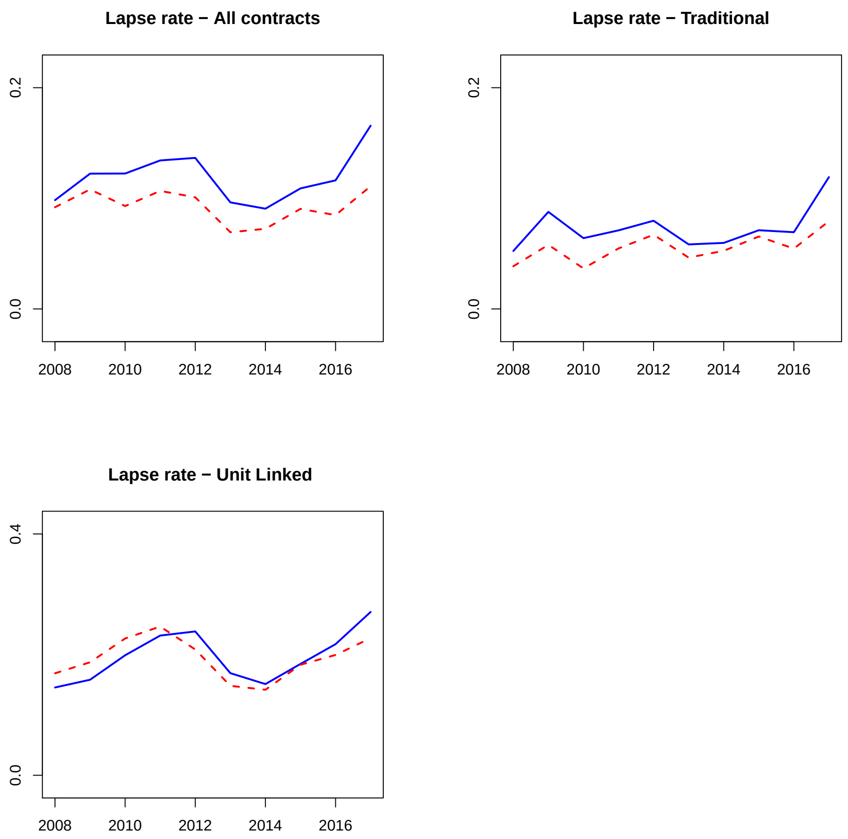

Fig. 1 Lapse rates of all contracts, traditional contracts and unit-linked contracts by year. We report the lapse rate computed as number of lapses divided by total outstanding contracts (blue solid line) and the lapse rate computed as total lapsed capital divided by total capital of outstanding contracts (red dashed line).

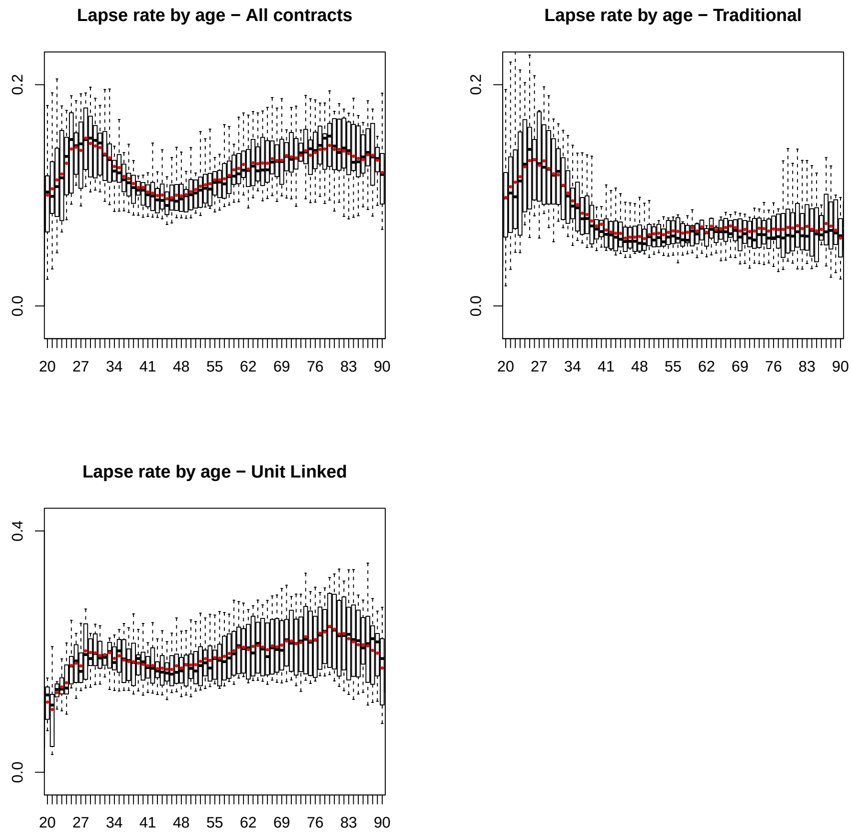

Fig. 2 Lapse rates of all contracts, traditional contracts and unit-linked contracts by age of policyholder. We consider age between 20 and 90 to cut off the outliers. Fixed an age, the vertical whiskers show the distribution of the lapse rate across 2008-2017: the black box contains 75% of data, while the whiskers extend to 95% of data. The red dots show the mean lapse rate for a given age across 2008-2017.

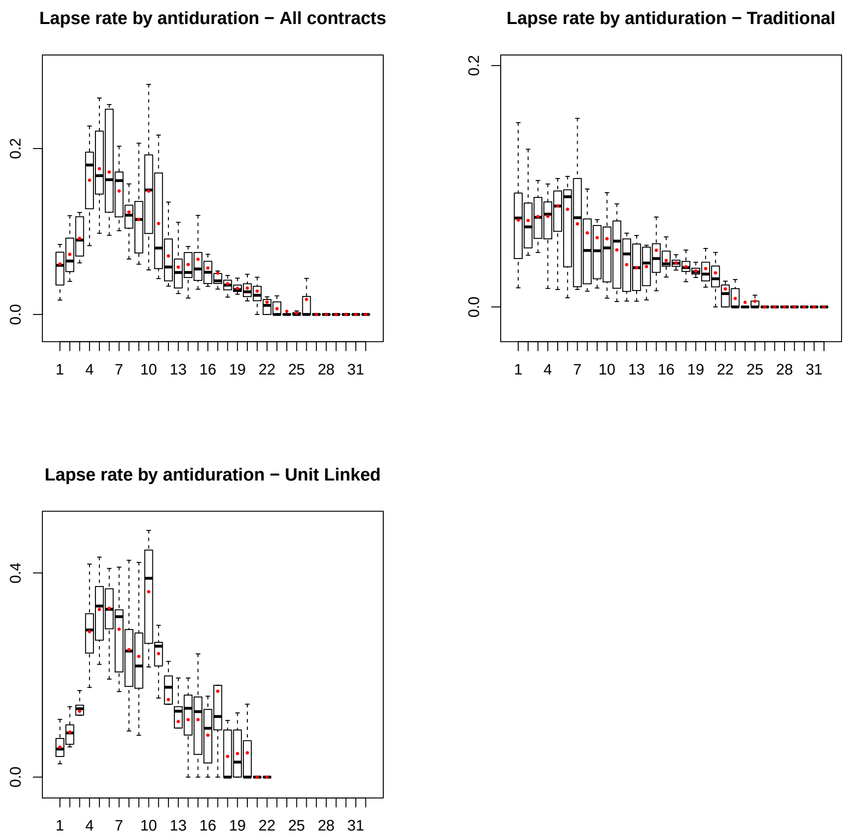

Fig. 3 Lapse rates of all contracts, traditional contracts and unit-linked contracts by contract anti-duration. We consider anti-duration between 1 and 32 years. Fixed an anti-duration, the vertical whiskers show the distribution of the lapse rate across 2008-2017: the black box contains 75% of data, while the whiskers extend to 95% of data. The red dots show the mean lapse rate for a given anti-duration across 2008-2017.

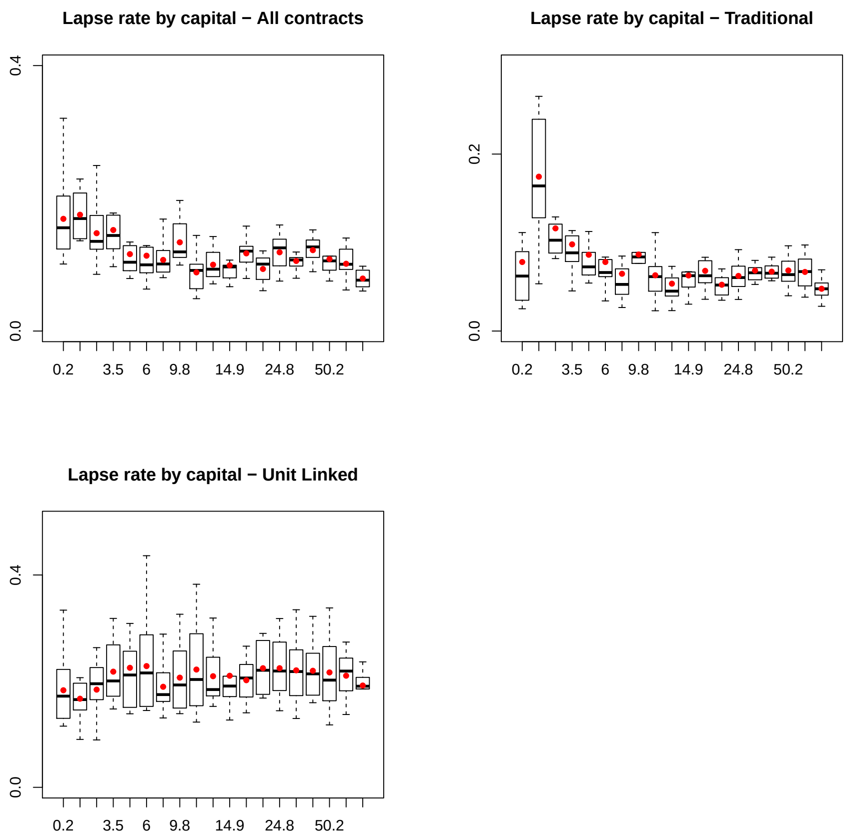

Fig. 4 Lapse rates of all contracts, traditional contracts and unit-linked contracts by capital insured by policyholders. We divide capital in buckets using the sample quantile of the distribution of capital insured across all contracts in the sample (in thousand euros). The quantiles are reported on the x-axis. Fixed a capital, the vertical whiskers show the distribution of the lapse rate across 2008-2017: the black box contains 75% of data, while the whiskers extend to 95% of data. The red dots show the mean lapse rate for a given anti-duration across 2008-2017.

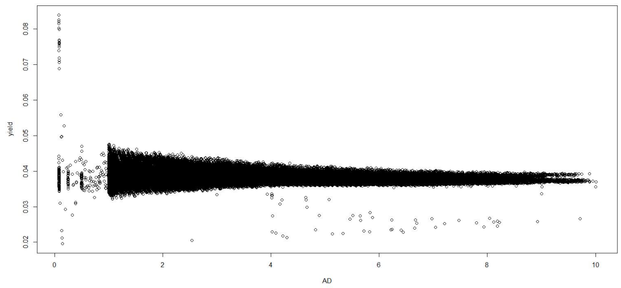

Fig. 5 Yield of lapsed policies and contract anti-duration (AD). Each dot of the plot refers to the pair (anti-duration, yield) of lapsed contracts.

Variable	*Level*	*Binomial*	*Poisson*	*PH*
Product type (ref. level: T)	UL	1.076***	1.112***	0.945***
		(0.003)	(0.003)	(0.003)
Gender (ref. level: M)	F	−0.093***	−0.085***	−0.060***
		(0.003)	(0.003)	(0.003)
Age (ref. level: Age3)	Age1	0.253***	0.201***	0.287***
		(0.006)	(0.005)	(0.005)
	Age2	−0.073***	−0.085***	−0.156***
		(0.004)	(0.003)	(0.003)
	Age4	0.120***	0.122***	0.212***
		(0.004)	(0.003)	(0.003)
Calendar year (ref. level: 2008)	2009	0.265***	0.244***	0.218***
		(0.006)	(0.006)	(0.006)
	2010	0.273***	0.265***	0.245***
		(0.006)	(0.006)	(0.006)
	2011	0.406***	0.401***	0.411***
		(0.006)	(0.006)	(0.006)
	2012	0.521***	0.529***	0.487***
		(0.006)	(0.006)	(0.006)
	2013	0.116***	0.124***	0.181***
		(0.007)	(0.006)	(0.006)
	2014	−0.017**	−0.028***	0.178***
		(0.007)	(0.007)	(0.007)
	2015	0.192***	0.184***	0.472***
		(0.007)	(0.006)	(0.006)
	2016	0.273***	0.276***	0.676***
		(0.007)	(0.007)	(0.006)
	2017	0.759***	0.749***	1.255***
		(0.007)	(0.006)	(0.006)
Region (ref. level: North-West)	Center	−0.025***	−0.023***	−0.013***
		(0.004)	(0.004)	(0.004)
	North-East	−0.044***	−0.044***	−0.057***
		(0.004)	(0.004)	(0.004)
	South and Islands	−0.032***	−0.045***	0.015***
		(0.005)	(0.004)	(0.004)
Premium freq. (ref. level: U)	P	−0.017***	−0.017***	−0.541***
		(0.004)	(0.003)	(0.003)
Capital (ref. level: cap4)	cap1	0.058***	0.002	0.197***
		(0.007)	(0.006)	(0.006)
	cap2	0.193***	0.124***	0.068***
		(0.005)	(0.004)	(0.004)
	cap3	0.036***	0.012***	0.112***
		(0.004)	(0.004)	(0.004)
	cap5	0.031***	0.031***	0.144***
		(0.005)	(0.004)	(0.004)
	cap6	−0.240***	−0.252***	−0.249***
		(0.008)	(0.008)	(0.008)
Antiduration (ref. level: AD1)	AD2	0.835***	0.877***	
		(0.003)	(0.003)	
	AD3	0.490***	0.536***	
		(0.005)	(0.004)	
	AD4	0.013*	0.050***	
		(0.007)	(0.006)	

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 2 Lapse rates and contract features - all contracts: Generalized linear regression results for lapse rates of the total sample. Regressors are the product type, the policyholder's gender, the policyholder's age, the year of observation, the macro-region of residence of the policyholder, the premium frequency, the insured capital, the anti-duration of the policy. All regressors and the correspondent reference levels are described in Table 1. The first column reports the results using the Binomial model, the second column reports the results using the Poisson model and the third column reports the results using the PH model. The constant in the regression has been removed to maintain confidentiality.

Variable	*Level*	*Binomial*	*Poisson*	*PH*
Gender (ref. level: M)	F	−0.179***	−0.180***	−0.152***
		(0.005)	(0.004)	(0.004)
Age (ref. level: Age3)	Age1	0.490***	0.501***	0.643***
		(0.008)	(0.007)	(0.007)
	Age2	−0.014**	−0.016***	−0.044***
		(0.006)	(0.005)	(0.005)
	Age4	0.103***	0.105***	0.179***
		(0.006)	(0.006)	(0.006)
Calendar year (ref. level: 2008)	2009	0.611***	0.616***	0.557***
		(0.010)	(0.010)	(0.010)
	2010	0.343***	0.353***	0.324***
		(0.011)	(0.011)	(0.011)
	2011	0.511***	0.517***	0.505***
		(0.011)	(0.010)	(0.010)
	2012	0.665***	0.667***	0.669***
		(0.011)	(0.010)	(0.010)
	2013	0.319***	0.327***	0.396***
		(0.012)	(0.011)	(0.011)
	2014	0.354***	0.362***	0.485***
		(0.012)	(0.011)	(0.011)
	2015	0.559***	0.566***	0.771***
		(0.012)	(0.011)	(0.011)
	2016	0.556***	0.565***	0.928***
		(0.012)	(0.011)	(0.011)
	2017	1.184***	1.201***	1.690***
		(0.011)	(0.010)	(0.010)
Region (ref. level: North-West)	Center	−0.009	−0.011	−0.012*
		(0.007)	(0.007)	(0.007)
	North-East	−0.013**	−0.011*	−0.010*
		(0.006)	(0.006)	(0.006)
	South and Islands	0.044***	0.047***	0.086***
		(0.007)	(0.006)	(0.006)
Premium freq. (ref. level: U)	P	0.114***	0.123***	−0.787***
		(0.006)	(0.005)	(0.005)
Capital (ref. level: cap4)	cap1	−0.399***	−0.468***	−0.677***
		(0.060)	(0.061)	(0.059)
	cap2	0.832***	0.837***	1.073***
		(0.007)	(0.007)	(0.007)
	cap3	0.236***	0.236***	0.397***
		(0.006)	(0.006)	(0.006)
	cap5	0.042***	0.036***	0.168***
		(0.007)	(0.007)	(0.007)
	cap6	−0.250***	−0.259***	−0.244***
		(0.011)	(0.011)	(0.011)
Antiduration (ref. level: AD1)	AD2	0.150***	0.135***	
		(0.005)	(0.005)	
	AD3	−0.046***	−0.055***	
		(0.008)	(0.007)	
	AD4	−0.663***	−0.688***	
		(0.009)	(0.009)	

**p<0.1; **p<0.05; ***p<0.01*

Table 3 Lapse rates and contract features - Traditional: Generalized linear regression results for lapse rates of traditional policies, when the regressors are the policyholder's gender, the policyholder's age, the year of observation, the macro-region of residence of the policyholder, the premium frequency, the insured capital, the anti-duration of the policy. All regressors and the correspondent reference levels are described in Table 1. The first column reports the results using the Binomial model, the second column reports the results using the Poisson model and the third column reports the results using the PH model. The constant in the regression has been removed to maintain confidentiality.

Variable	*Level*	*Binomial*	*Poisson*	*PH*
Gender (ref. level: M)	F	−0.029***	−0.025***	−0.011***
		(0.004)	(0.003)	(0.003)
Age (ref. level: Age3)	Age1	−0.045***	−0.084***	0.004
		(0.009)	(0.008)	(0.008)
	Age2	−0.118***	−0.132***	−0.166***
		(0.005)	(0.004)	(0.004)
	Age4	0.092***	0.091***	0.150***
		(0.005)	(0.004)	(0.004)
Calendar year (ref. level: 2008)	2009	0.102***	0.092***	0.064***
		(0.008)	(0.007)	(0.007)
	2010	0.210***	0.196***	0.223***
		(0.008)	(0.007)	(0.007)
	2011	0.278***	0.289***	0.374***
		(0.008)	(0.007)	(0.007)
	2012	0.411***	0.447***	0.406***
		(0.008)	(0.007)	(0.007)
	2013	0.110***	0.137***	0.079***
		(0.009)	(0.008)	(0.008)
	2014	−0.143***	−0.150***	0.015*
		(0.010)	(0.009)	(0.008)
	2015	0.085***	0.075***	0.327***
		(0.009)	(0.008)	(0.008)
	2016	0.220***	0.227***	0.570***
		(0.009)	(0.008)	(0.008)
	2017	0.544***	0.551***	0.982***
		(0.009)	(0.008)	(0.008)
Region (ref. level: North-West)	Center	−0.015***	−0.012**	−0.005
		(0.006)	(0.005)	(0.005)
	North-East	−0.020***	−0.019***	−0.035***
		(0.005)	(0.005)	(0.005)
	South and Islands	−0.134***	−0.143***	−0.070***
		(0.006)	(0.006)	(0.006)
Premium freq. (ref. level: U)	P	0.190***	0.178***	−0.0005
		(0.005)	(0.005)	(0.004)
Capital (ref. level: cap4)	cap1	−0.293***	−0.336***	−0.409***
		(0.008)	(0.007)	(0.007)
	cap2	−0.481***	−0.492***	−0.821***
		(0.007)	(0.006)	(0.006)
	cap3	−0.236***	−0.243***	−0.270***
		(0.006)	(0.005)	(0.005)
	cap5	0.107***	0.107***	0.153***
		(0.007)	(0.006)	(0.006)
	cap6	−0.025**	−0.049***	−0.067***
		(0.012)	(0.011)	(0.011)
Antiduration (ref. level: AD1)	AD2	1.405***	1.422***	
		(0.005)	(0.004)	
	AD3	1.104***	1.133***	
		(0.007)	(0.006)	
	AD4	0.893***	0.884***	
		(0.010)	(0.009)	

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 4 Lapse rates and contract features - Unit Linked: Generalized linear regression results for lapse rates of unit-linked policies, when the regressors are the policyholder's gender, the policyholder's age, the year of observation, the macro-region of residence of the policyholder, the premium frequency, the insured capital, the anti-duration of the policy. All regressors and the correspondent reference levels are described in Table 1. The first column reports the results using the Binomial model, the second column reports the results using the Poisson model and the third column reports the results using the PH model. The constant in the regression has been removed to maintain confidentiality.

Variable	*Level*	*Binomial*	*Poisson*	*PH*
Level (ref. level: T)	UL	1.248***	1.268***	1.189***
		(0.005)	(0.005)	(0.005)
Gender (ref. level: M)	F	−0.056***	−0.049***	−0.015***
		(0.004)	(0.004)	(0.004)
Age (ref. level: Age3)	Age1	0.015	−0.063***	0.044***
		(0.010)	(0.009)	(0.009)
	Age2	−0.095***	−0.118***	−0.166***
		(0.006)	(0.005)	(0.005)
	Age4	0.009*	0.008*	0.104***
		(0.005)	(0.004)	(0.004)
Calendar year (ref. level: 2008)	2009	0.238***	0.223***	0.194***
		(0.009)	(0.009)	(0.009)
	2010	0.253***	0.229***	0.333***
		(0.009)	(0.008)	(0.008)
	2011	0.428***	0.401***	0.556***
		(0.009)	(0.008)	(0.008)
	2012	0.540***	0.528***	0.574***
		(0.009)	(0.008)	(0.008)
	2013	0.117***	0.094***	0.262***
		(0.010)	(0.009)	(0.009)
	2014	−0.125***	−0.183***	0.250***
		(0.010)	(0.010)	(0.009)
	2015	0.204***	0.136***	0.650***
		(0.010)	(0.009)	(0.009)
	2016	0.395***	0.341***	0.896***
		(0.010)	(0.009)	(0.009)
	2017	0.847***	0.787***	1.347***
		(0.010)	(0.009)	(0.009)
Region (ref. level: North-West)	Center	−0.045***	−0.047***	−0.069***
		(0.006)	(0.005)	(0.005)
	North-East	−0.045***	−0.055***	−0.030***
		(0.006)	(0.005)	(0.005)
	South and Islands	−0.011	−0.022***	−0.033***
		(0.007)	(0.006)	(0.006)
Premium freq. (ref. level: U)	P	−0.320***	−0.296***	−0.665***
		(0.006)	(0.005)	(0.005)
Capital (ref. level: cap4)	cap1	−0.037***	−0.105***	0.028***
		(0.009)	(0.008)	(0.008)
	cap2	−0.163***	−0.200***	−0.248***
		(0.007)	(0.006)	(0.006)
	cap3	−0.046***	−0.064***	−0.002
		(0.006)	(0.005)	(0.005)
	cap5	0.011*	0.013**	0.080***
		(0.006)	(0.005)	(0.005)
	cap6	−0.192***	−0.198***	−0.092***
		(0.011)	(0.010)	(0.010)
Antiduration (ref. level: AD1)	AD2	1.321***	1.365***	
		(0.005)	(0.004)	
	AD3	1.103***	1.128***	
		(0.008)	(0.007)	
	AD4	0.588***	0.567***	
		(0.011)	(0.011)	
Profession (ref. level: O)	E	−0.116***	−0.111***	−0.140***
		(0.005)	(0.005)	(0.005)

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 5 Lapse rates and contract features including profession - all contracts: Generalized linear regression results for lapse rates of total policies, when the regressors are the policy type, the policyholder's gender, the policyholder's age, the year of observation, the macro-region of residence of the policyholder, the premium frequency, the insured capital, the anti-duration of the policy and the profession of the policyholder. All regressors and the correspondent reference levels are described in Table 1. The first column reports the results using the Binomial model, the second column reports the results using the Poisson model and the third column reports the results using the PH model. The constant in the regression has been removed to maintain confidentiality.

Variable	*Level*	*Binomial*	*Poisson*	*PH*
Product type (red. level: T)	UL	0.198***	0.165***	−0.170***
		(0.006)	(0.005)	(0.006)
Gender (ref. level: M)	F	−0.212***	−0.219***	−0.199***
		(0.005)	(0.005)	(0.005)
Age (ref. level: Age 3)	Age1	0.537***	0.547***	0.520***
		(0.009)	(0.008)	(0.008)
	Age2	0.146***	0.148***	0.140***
		(0.007)	(0.006)	(0.006)
	Age4	0.084***	0.082***	0.076***
		(0.007)	(0.007)	(0.007)
Calendar year (ref.level: 2008)	2009	0.364***	0.364***	0.323***
		(0.009)	(0.008)	(0.008)
	2010	0.207***	0.211***	0.173***
		(0.010)	(0.009)	(0.009)
	2011	0.131***	0.133***	0.134***
		(0.010)	(0.010)	(0.010)
	2012	0.156***	0.155***	0.175***
		(0.010)	(0.010)	(0.010)
	2013	−0.225***	−0.222***	−0.146***
		(0.012)	(0.011)	(0.011)
	2014	−0.307***	−0.304***	−0.183***
		(0.013)	(0.013)	(0.013)
	2015	−0.230***	−0.228***	0.041***
		(0.014)	(0.013)	(0.013)
	2016	−0.328***	−0.330***	0.279***
		(0.016)	(0.015)	(0.015)
	2017	0.131***	0.111***	1.345***
		(0.014)	(0.013)	(0.013)
Region (ref. level: North-West)	Center	−0.046***	−0.048***	−0.053***
		(0.008)	(0.008)	(0.008)
	North-East	0.001	0.003	0.013*
		(0.007)	(0.007)	(0.007)
	South an Islands	0.029***	0.025***	0.019**
		(0.008)	(0.007)	(0.007)
Premium freq. (ref. level: U)	P	0.149***	0.156***	−0.050***
		(0.007)	(0.006)	(0.006)
Capital (ref. level: cap4)	cap1	0.520***	0.525***	0.656***
		(0.012)	(0.012)	(0.012)
	cap2	0.598***	0.593***	0.708***
		(0.009)	(0.008)	(0.008)
	cap3	0.314***	0.314***	0.318***
		(0.008)	(0.008)	(0.008)
	cap5	0.037***	0.039***	0.017**
		(0.009)	(0.008)	(0.008)
	cap6	−0.055***	−0.052***	−0.094***
		(0.016)	(0.015)	(0.015)

$^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 6 Lapse rates and contract features for early lapses - all contracts: Generalized linear regression results for lapse rates of total policies with anti-duration of one year or less, when the regressors are the policy type, the policyholder's gender, the policyholder's age, the year of observation, the macro-region of residence of the policyholder, the premium frequency and the insured capital. All regressors and the correspondent reference levels are described in Table 1. The first column reports the results using the Binomial model, the second column reports the results using the Poisson model and the third column reports the results using the PH model. The constant in the regression has been removed to maintain confidentiality.

	(i)	(ii)	(iii)	(iv)
$\texttt{lapse}_{-1}$	0.526***	0.454*	0.525***	0.473**
	[0.108]	[0.237]	[0.149]	[0.216]
hcpi	0.818***	0.731*	0.968***	0.880***
	[0.270]	[0.435]	[0.320]	[0.298]
disp_inc	-0.044	-0.216	-0.026	-0.049
	[0.193]	[0.342]	[0.273]	[0.240]
unem	-0.027	-0.032	-0.018	-0.022
	[0.032]	[0.035]	[0.032]	[0.027]
stoxx	-0.016	-0.027	-0.048**	-0.016
	[0.020]	[0.022]	[0.022]	[0.017]
poverty	-0.064	-0.036	-0.052	-0.087
	[0.122]	[0.143]	[0.183]	[0.135]
interest	-0.006*	-0.005	-0.008*	[-]
	[0.004]	[0.004]	[0.004]	[-]
yield	[-]	-0.010	[-]	[-]
	[-]	[0.037]	[-]	[-]
Δyield	[-]	[-]	-0.060***	[-]
	[-]	[-]	[0.022]	[-]
conv_yield	[-]	[-]	[-]	0.006
	[-]	[-]	[-]	[0.004]
Model Diagnostics				
ar2p	0.365	0.632	0.620	0.469
hansp	0.246	0.955	0.948	0.593
$\texttt{F}_p$	0.000	0.000	0.000	0.000
$\texttt{N}_g$	20	20	20	20
N	157	157	157	157

Table 7 Panel regression results - Traditional: Dynamic panel regression results for the endogenous variable lapse computed for traditional contracts. Statistics: p-value for the Arellano and Bond **ar2** test for residual autocorrelation, p-value for the Hansen test **hansp** for model overidentification (if the null is rejected, the model presents autocorrelation of residuals in the first case and too many instrumental variables in the GMM estimator in the second case); p-value from the F-statistic $\mathbf{F}_p$ testing the null of overall parameters significance; $\mathbf{N}_g$ is the number of individuals and $\mathbf{N}$ is the sample size. The constant in the regression has been removed to maintain confidentiality.

	(i)	(ii)	(iii)	(iv)
lapse$_{-1}$	0.513***	0.393***	0.379***	0.343***
	[0.140]	[0.121]	[0.108]	[0.093]
hcpi	1.17	1.478	1.722	-0.041
	[0.779]	[1.067]	[1.185]	[0.606]
disp_inc	-0.774**	-0.661*	-0.468*	-0.788**
	[0.335]	[0.370]	[0.343]	[0.286]
unem	0.046	0.043	0.04	0.014
	[0.044]	[0.047]	[0.047]	[0.057]
stoxx	-0.173**	-0.101	-0.107	-0.113
	[0.070]	[0.093]	[0.066]	[0.080]
poverty	0.044	0.326	0.192	0.412
	[0.315]	[0.358]	[0.309]	[0.347]
interest	-0.016**	-0.015**	-0.017**	[-]
	[0.006]	[0.007]	[0.008]	[-]
yield	[-]	0.028	[-]	[-]
	[-]	[0.040]	[-]	[-]
Δyield	[-]	[-]	0.019	[-]
	[-]	[-]	[0.018]	[-]
conv_yield	[-]	[-]	[-]	0.034
	[-]	[-]	[-]	[0.036]
Model Diagnostics				
ar2p	0.166	0.24	0.23	0.314
hansp	0.999	0.995	0.998	0.991
F$_p$	0.000	0.000	0.000	0.000
N$_g$	20	20	20	20
N	157	157	157	157

Table 8 Panel regression results - Unit Linked: Dynamic panel regression results for the endogenous variable `lapse` computed for unit-linked contracts. The models and diagnostics are the same as in Table 7. The constant in the regression has been removed to maintain confidentiality.

	(i)	(ii)	(iii)	(iv)
$\texttt{lapse}_{-1}$	0.435***	0.294*	0.403**	0.536***
	[0.145]	[0.169]	[0.175]	[0.112]
hcpi	0.9	0.809	0.892	-0.291
	[0.606]	[0.499]	[0.641]	[0.439]
disp_inc	-0.574**	-0.590**	-0.637**	-0.558**
	[0.241]	[0.238]	[0.248]	[0.262]
unem	0.008	-0.016	0.005	-0.007
	[0.031]	[0.034]	[0.037]	[0.033]
stoxx	-0.087*	-0.060	-0.083	-0.060
	[0.047]	[0.043]	[0.051]	[0.046]
poverty	0.068	0.075	0.135	0.182
	[0.201]	[0.195]	[0.192]	[0.226]
interest	-0.008	-0.007	-0.007	[-]
	[0.005]	[0.005]	[0.006]	[-]
yield	[-]	-0.024***	[-]	[-]
	[-]	[0.005]	[-]	[-]
Δyield	[-]	[-]	0.043	[-]
	[-]	[-]	[0.037]	[-]
conv_yield	[-]	[-]	[-]	-0.010**
	[-]	[-]	[-]	[0.005]
Model Diagnostics				
ar2p	0.153	0.232	0.172	0.203
hansp	0.133	0.181	0.19	0.123
$\texttt{F}_p$	0.000	0.001	0.002	0.003
$\texttt{N}_g$	40	40	40	40
N	314	314	314	314

Table 9 Panel regression results - all contracts: Dynamic panel regression results for the endogenous variable `lapse` computed for all contracts. The models and diagnostics are the same as in Table 7. The constant in the regression has been removed to maintain confidentiality.

lapse$_{-1}$	0.327**	0.339**	0.314**	0.330**
	[0.156]	[0.159]	[0.158]	[0.154]
hcpi	2.026***	2.450**	2.056***	2.200***
	[0.668]	[1.027]	[0.670]	[0.722]
disp_inc	0.004	0.249	0.014	0.087
	[0.329]	[0.589]	[0.339]	[0.350]
unem	-0.045	-0.018	-0.054	-0.039
	[0.049]	[0.055]	[0.051]	[0.051]
stoxx	-0.034	-0.023	-0.032	-0.028
	[0.034]	[0.051]	[0.028]	[0.034]
poverty	0.047	-0.006	0.015	0.048
	[0.338]	[0.295]	[0.316]	[0.333]
interest	-0.016***	-0.021**	-0.015***	[-]
	[0.006]	[0.008]	[0.006]	[-]
yield	[-]	0.041	[-]	[-]
	[-]	[0.069]	[-]	[-]
Δyield	[-]	[-]	-0.029	[-]
	[-]	[-]	[0.030]	[-]
conv_yield	[-]	[-]	[-]	0.017***
	[-]	[-]	[-]	[0.006]
Model Diagnostics				
ar2p	0.498	0.592	0.391	0.488
hansp	0.98	0.993	0.991	0.982
F$_p$	0.000	0.000	0.000	0.000
N$_g$	20	20	20	20
N	155	155	155	155

Table 10 Panel regression results - Traditional policies with unique premium payment: Dynamic panel regression results for the endogenous variable lapse computed for traditional policies with unique premium payment. The models and diagnostics are the same as in Table 7. The constant in the regression has been removed to maintain confidentiality.

lapse$_{-1}$	0.292**	0.325**	0.560***	0.281**
	[0.140]	[0.151]	[0.102]	[0.140]
hcpi	0.239	0.057	0.615***	0.168
	[0.263]	[0.438]	[0.216]	[0.296]
disp_inc	-0.299	-0.421*	-0.181	-0.309
	[0.197]	[0.225]	[0.205]	[0.199]
unem	-0.016	-0.026	-0.006	-0.017
	[0.025]	[0.029]	[0.020]	[0.025]
stoxx	0.008	-0.011	-0.039	0.008
	[0.020]	[0.027]	[0.026]	[0.019]
poverty	0.161	0.117	0.133	0.163
	[0.116]	[0.122]	[0.160]	[0.117]
interest	0.003	0.004	-0.004	[-]
	[0.003]	[0.005]	[0.003]	[-]
yield	[-]	-0.031	[-]	[-]
	[-]	[0.032]	[-]	[-]
Δyield	[-]	[-]	-0.060***	[-]
	[-]	[-]	[0.011]	[-]
conv_yield	[-]	[-]	[-]	-0.004
	[-]	[-]	[-]	[0.004]
Model Diagnostics				
ar2p	0.793	0.935	0.31	0.747
hansp	0.9	0.912	0.981	0.901
F$_p$	0.000	0.000	0.000	0.000
N$_g$	20	20	20	20
N	157	157	157	157

Table 11 Panel regression results - Traditional policies with periodic premium payment: Dynamic panel regression results for the endogenous variable lapse computed for traditional policies with periodic premium payment. The models and diagnostics are the same as in Table 7. The constant in the regression has been removed to maintain confidentiality.

lapse$_{-1}$	0.525***	0.516***	0.452***	0.369**
	[0.126]	[0.142]	[0.116]	[0.168]
hcpi	1.086	1.989	1.352	-0.038
	[0.930]	[2.622]	[1.228]	[0.896]
disp_inc	-0.984***	-0.767	-0.803**	-0.944**
	[0.329]	[0.812]	[0.373]	[0.382]
unem	0.054	0.056	0.038	0.011
	[0.043]	[0.047]	[0.038]	[0.058]
stoxx	-0.208***	-0.167	-0.138	-0.144
	[0.067]	[0.135]	[0.095]	[0.105]
poverty	0.109	0.164	0.272	0.379
	[0.295]	[0.291]	[0.368]	[0.349]
interest	-0.016**	-0.021	-0.015*	[-]
	[0.007]	[0.019]	[0.008]	[-]
yield	[-]	0.02	[-]	[-]
	[-]	[0.142]	[-]	[-]
Δyield	[-]	[-]	0.016	[-]
	[-]	[-]	[0.039]	[-]
conv_yield	[-]	[-]	[-]	0.02
	[-]	[-]	[-]	[0.048]
Model Diagnostics				
ar2p	0.235	0.235	0.294	0.428
hansp	0.999	0.999	0.997	0.982
F$_p$	0.000	0.000	0.000	0.001
N$_g$	20	20	20	20
N	157	157	157	157

Table 12 Panel regression results - Unit Linked policies with unique premium payment: Dynamic panel regression results for the endogenous variable lapse computed for unit linked policies with unique premium payment. The models and diagnostics are the same as in Table 7. The constant in the regression has been removed to maintain confidentiality.

lapse$_{-1}$	0.589*	0.514	0.587*	0.598*
	[0.345]	[0.346]	[0.349]	[0.361]
hcpi	0.461	1.052	0.689	0.478
	[0.655]	[1.184]	[0.690]	[0.668]
disp_inc	-0.574	-0.56	-0.521	-0.466
	[0.544]	[0.558]	[0.506]	[0.593]
unem	0.008	-0.004	-0.009	0.022
	[0.065]	[0.057]	[0.058]	[0.063]
stoxx	-0.094*	-0.078	-0.092**	-0.110**
	[0.053]	[0.055]	[0.045]	[0.054]
poverty	1.109*	0.64	0.888	1.076**
	[0.574]	[0.577]	[0.631]	[0.467]
interest	-0.006*	-0.01	-0.007	[-]
	[0.004]	[0.011]	[0.005]	[-]
yield	[-]	0.023	[-]	[-]
	[-]	[0.068]	[-]	[-]
Δyield	[-]	[-]	-0.03	[-]
	[-]	[-]	[0.042]	[-]
conv_yield	[-]	[-]	[-]	0.007
	[-]	[-]	[-]	[0.006]
Model Diagnostics				
ar2p	0.318	0.465	0.358	0.422
hansp	0.99	0.996	0.989	0.983
F$_p$	0.008	0.058	0.075	0.043
N$_g$	20	20	20	20
N	153	153	153	153

Table 13 Panel regression results - Traditional policies and early lapses: Dynamic panel regression results for the endogenous variable lapse computed for traditional policies with anti-duration less or equal than 3 years. The models and diagnostics are the same as in Table 7. The constant in the regression has been removed to maintain confidentiality.

`lapse`$_{-1}$	0.372**	0.262*	0.323*	0.300
	[0.165]	[0.140]	[0.185]	[0.200]
`hcpi`	-0.095	4.035	-0.673	0.344
	[1.886]	[2.998]	[1.572]	[0.701]
`disp_inc`	-1.240	0.295	-0.866	-0.809
	[0.960]	[0.706]	[0.900]	[0.671]
`unem`	-0.003	0.066	0.006	0.007
	[0.053]	[0.062]	[0.052]	[0.041]
`stoxx`	-0.041	0.129	0.031	0.035
	[0.111]	[0.093]	[0.114]	[0.108]
`poverty`	-0.079	-0.019	0.088	0.093
	[0.340]	[0.303]	[0.335]	[0.340]
`interest`	0.000	-0.032	0.008	[-]
	[0.014]	[0.024]	[0.010]	[-]
`yield`	[-]	0.239*	[-]	[-]
	[-]	[0.141]	[-]	[-]
Δ`yield`	[-]	[-]	0.063	[-]
	[-]	[-]	[0.042]	[-]
`conv_yield`	[-]	[-]	[-]	0.058
	[-]	[-]	[-]	[0.044]
Model Diagnostics				
`ar2p`	0.28	0.411	0.354	0.401
`hansp`	0.929	0.948	0.946	0.899
`F`$_p$	0.000	0.000	0.000	0.001
`N`$_g$	20	20	20	20
`N`	157	157	157	157

Table 14 Panel regression results - Unit linked policies and early lapses: Dynamic panel regression results for the endogenous variable `lapse` computed unit linked policies with anti-duration less or equal than 3 years. The models and diagnostics are the same as in Table 7. The constant in the regression has been removed to maintain confidentiality.