

Reck, Lucas; Schupp, Johannes; Reuß, Andreas

**Article — Published Version**

## Identifying the determinants of lapse rates in life insurance: an automated Lasso approach

European Actuarial Journal

**Provided in Cooperation with:**

Springer Nature

*Suggested Citation:* Reck, Lucas; Schupp, Johannes; Reuß, Andreas (2022) : Identifying the determinants of lapse rates in life insurance: an automated Lasso approach, European Actuarial Journal, ISSN 2190-9741, Springer, Berlin, Heidelberg, Vol. 13, Iss. 2, pp. 541-569, <https://doi.org/10.1007/s13385-022-00325-1>

This Version is available at:

<https://hdl.handle.net/10419/311903>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>



# Identifying the determinants of lapse rates in life insurance: an automated Lasso approach

Lucas Reck<sup>1,2</sup> · Johannes Schupp<sup>1</sup> · Andreas Reuß<sup>1</sup>

Received: 14 February 2022 / Revised: 22 June 2022 / Accepted: 22 July 2022 /

Published online: 10 November 2022

© The Author(s) 2022

## Abstract

Lapse risk is a key risk driver for life and pensions business with a material impact on the cash flow profile and the profitability. The application of data science methods can replace the largely manual and time-consuming process of estimating a lapse model that reflects various contract characteristics and provides best estimate lapse rates, as needed for Solvency II valuations. In this paper, we use the Lasso method which is based on a multivariate model and can identify patterns in the data set automatically. To identify hidden structures within covariates, we adapt and combine recently developed extended versions of the Lasso that apply different subpenalties for individual covariates. In contrast to random forests or neural networks, the predictions of our lapse model remain fully explainable, and the coefficients can be used to interpret the lapse rate on an individual contract level. The advantages of the method are illustrated based on data from a European life insurer operating in four countries. We show how structures can be identified efficiently and fed into a highly competitive, automatically calibrated lapse model.

**Keywords** Lapse rate · Life insurance · Lasso · Fused Lasso · Trend filtering

---

✉ Lucas Reck  
l.reck@ifa-ulm.de

Johannes Schupp  
j.schupp@ifa-ulm.de

Andreas Reuß  
a.reuss@ifa-ulm.de

<sup>1</sup> ifa (Institut für Finanz- und Aktuarwissenschaften), Lise-Meitner-Straße 14, 89081 Ulm, Germany

<sup>2</sup> Institute of Insurance Science, Ulm University, Helmholtzstraße 20, 89081 Ulm, Germany

## 1 Introduction

Risk management in life insurance requires proper modelling, measuring, and managing of the key risk drivers of life and pension business. This includes lapse risk since the termination of a life insurance contract prior to maturity has a significant impact on the cash flow profile and the profitability of life insurance business. To reflect this risk, market consistent valuations are based on best estimate future lapse rates and the Solvency II standard formula assesses lapse risk in a specific risk module. Lapse in a broader sense includes a premature and full termination of contracts but also partial terminations as well as changes in the frequency and amount of premium payments. In what follows, we refer to lapse as a full and premature termination of the contract induced by the policyholder, as this is the most severe part of this risk. It has an important effect on an insurer's asset liability management, because the premium calculations assume that the contract lasts (and pays premium) until the end of the agreed term. Especially high acquisition costs are no longer covered if the contract is cancelled early on. Higher than anticipated lapse rates hence induce a severe risk for an insurance company. Another important reason is the liquidity risk: If a lot of policyholders lapse at the same time, the insurance company must ensure enough liquid capital to pay out their customers. This needs to be considered in the asset allocation as well. The supervisory authority also recognises the importance of lapse risk. Under Solvency II, lapse risk is the most material sub-module in the life underwriting risk, which itself is the second most material module (behind market risk); see EIOPA [9]. This shows the importance of an accurate lapse model which enables the insurance company to detect lapse risk in a portfolio and manage its cash flows accordingly.

The literature in the field of lapse models can be classified by data set and by model. The first class of data sets only contains macroeconomic variables, while the second class focuses on policy specific variables; of course, combinations are possible as well. Macroeconomic variables such as unemployment rate, internal or external rate of return, were used to analyse the interest rate hypothesis (high interest rates lead to higher lapse rates because policyholders are more likely to take the surrender value and reinvest it in more profitable products) and the emerging fund hypothesis (policyholders lapse if they face a financial crisis and need the surrender value from the life insurance contract). Two examples of such analyses are given in Kiesenbauer [18], where macroeconomic indicators are analysed for the German life insurance industry, and in Yu et al. [36], where lapse behaviour in life insurance is analysed for the Chinese market.

Of course, policy specific variables such as entry age, contract duration, gender or sum insured can be used to analyse individual and contract specific lapse behaviour. Overall, there are not many papers using policyholder specific data since the data is typically treated confidentially. Eling and Kochanski [11] give an overview of research on lapse in life insurance. They find 44 papers on theoretical lapse models and 12 empirical papers. From those 12 empirical papers, there are seven papers about macroeconomic variables and five papers about product and policyholder characteristics.

In the literature, only a few models have been used to analyse lapse. The main tools are survival analysis, e.g. Aleandri and Elettì [2] or Milhaud and Dutang [23], and Generalised Linear Models (GLM). For example, the Cox proportional hazard model (most popular representative of survival analysis models) can be used to analyse the aforementioned two hypotheses, i.e. whether the lapse risk is increased by macroeconomic variables like interest rates or by a policyholder's financial situation. The GLM models expected values, i.e. a lapse probability, either on a portfolio level or on a single contract level, and thus can be further subdivided into GLMs modelling the amount of lapsed policies with a Poisson distribution or on a single policy perspective using a Binomial distribution. The lapse probability as an output is required to quantify the impact on cash flows within an insurer's market consistent valuation model. We will therefore focus on these models in what follows.

Two examples of the analysis of lapse in the European life insurance industry on an individual contract level are given by Barucci et al. [5] for the Italian market and Eling and Kiesenbauer [10] for the German market. Barucci et al. [5] use both policy specific variables for each contract and some macroeconomic variables. They analysed both GLMs (with Poisson and Binomial assumption) and the Cox proportional hazard model and derive separate models for unit linked and traditional business. Some of their findings are higher lapse rates for unit linked products compared to traditional contracts and high lapse rates within the first years after inception. A similar analysis for the German market can be found in Eling and Kiesenbauer [10]. They analyse more than one million contracts from a German life insurer, using the Cox proportional hazard model and GLMs (again with Poisson and Binomial assumption). They also extended their analysis by including an interaction term in the model. Some main findings are increasing lapse rates over time (calendar year), especially in phases of crisis, decreasing lapse rates with increasing contract age and no major difference between unit-linked and traditional business.

GLMs are multivariate, but still easy to implement and interpret. The main difficulty is that with an increasing number of covariates also the number of coefficients to be estimated rises rapidly, especially when interactions are added. This holds for both numeric and discrete covariates. Discrete covariates with many distinct levels, i.e. high cardinality, like tariff require one estimate per level. Here, similar levels may be grouped together manually. Also for numeric variables, the model itself is not capable of detecting any trends or structures within a covariate. Here, polynomials (or trends) have to be defined manually. Both attempts of including structures within covariates aim to build a parsimonious model that still generates good forecasts. By the multivariate nature of the model, a decision for one covariate always impacts or even breaks the structures for any other covariate. I.e. incorporating one additional covariate requires a complete readjustment of the structures for all other covariates. Thus, the process of identifying a parsimonious and comprehensive model is highly iterative and time-consuming so far.

Therefore, we propose to apply the Lasso, introduced by Tibshirani [29], as a purely data-driven and automated lapse model. We combine different versions of the Lasso, namely fused and trend filtering (see Tibshirani et al. [33], Kim et al. [19] and Tibshirani and Taylor [32]), in order to include different structures depending on the covariate. We also show how interactions, which are prevalent in most practical

applications, can be included straightforwardly. The resulting model combines both model fitting and variable selection, and hence is a big advantage for identifying a parsimonious and comprehensive model as the iterative and time-consuming task can be replaced by an automated process. It requires less expert judgement or manual adjustments which makes it more flexible and less prone to under—or overfitting. The impact of any policy characteristic (e.g. contract duration, sum insured, etc.) can be directly derived from the models coefficients of the GLM, and interpreted accordingly. Being able to understand the impact of different covariates is crucial for some applications, e.g. the best estimate assumptions under Solvency II. Other machine learning algorithms lack this property, and the regulator has not yet fully clarified under which circumstances the application of a Black Box model is possible. However, for applications like e.g. lapse prevention campaigns, the highly accurate predictions of a Black Box model can be advantageous and outweigh the drawback of a worse explainability. Thus, depending on the application, both types of models have their advantages.

Azzone et al. [4] use a random forest to analyse the lapse behaviour in the Italian life insurance market and show that this approach outperforms a classical logistic regression model. They also apply useful tools for the interpretability of the random forest. There are also other lapse models using random forest or bagging algorithms, see e.g. Milhaud et al. [24], or Aleandri [1], or neural networks, see e.g. Xong and Kang [35], support vector machines, see e.g. Loisel et al. [21], or Xong and Kang [35], or boosting algorithms, see e.g. Loisel et al. [21]. See also Kiermayer [17] for a comparison of different machine learning algorithms for simulated data. However, all these models are not explainable on a similar level as a GLM, with direct interpretations of the model coefficients. Overall, machine learning models expand the spectrum of possible models for lapse predictions and the model choice depends on the particular application at hand. For risk management applications, we prefer a parsimonious and fully explainable model.

The remainder of this paper is organised as follows: In Sect. 2, we introduce the data set of a European life insurer operating in different countries and the current approach to model lapse rates within that portfolio. We discuss how the portfolio is typically subdivided and the induced difficulties of this current approach. Section 3 discusses the methodology of the Lasso and shows how the different versions can be combined within one framework. In particular, we compare several variants and their application for different covariates. We also show how these variants can be applied in order to include interactions. In Sect. 4, we provide numerical results and compare several lapse models with respect to parsimony and goodness of fit. Finally, Sect. 5 concludes.

## 2 Data set and common actuarial practice

This section includes a brief description of the data set used in the numerical analysis in subsequent sections. We also describe how lapse rates in this data set can be modelled with a Whittaker–Henderson approach, a popular smoothing method in actuarial practice. This will serve as a reference model for the analysis in Sect. 4.

In this paper, ‘lapse’ is defined to comprise surrender (insured person cancels the contract and gets the surrender value), ‘pure’ lapse (insurance contract is terminated without a surrender value payment) and transfer (insured person cancels the contract and transfers the surrender value to another insurance company). This implies that lapse is defined as the termination of the life insurance contract, excluding death and expiry at the maturity date. The option of making a contract paid-up (i.e. reducing regular premium payments to zero) is not considered.

## 2.1 Data set

The data set comprises a set of insurance contracts written over a period of roughly 12 years which are observed for almost 10 additional run-off years (with no new contracts written after the first 12 years). Note that the proposed methodology can also be applied to situations where the portfolio is not in run-off. The data set is well structured and without missing values. There are  $n = 501,251$  observations (104,555 unique contracts) with  $J = 13$  covariates: contract duration (number of years between inception and observation time), insurance type (traditional or unit-linked), country (four European countries), gender, payment frequency (e.g. monthly or yearly), payment method (e.g. debit advice or depositor), nationality (whether or not the country, in which the insurance was sold equals the nationality of the policyholder), dynamic premium increase percentage,<sup>1</sup> entry age, original term of the contract, premium payment duration, sum insured and yearly premium. Note that the original data set contains even more covariates, but many of them are omitted because they do not contain predictive information (e.g. name of the insured person).

In order to automatically model structures and complex trends with the Lasso, one pre-processing step appears useful. In this step, we transform the continuous covariates into discrete covariates by partitioning them into a finite set of classes (bins). We may choose the bins based on expert knowledge, in order to get reasonable intervals, or based on quantiles. It turns out that a fully data driven model performs best, where the bins are based on a univariate decision tree. Each decision tree is grown with the restriction that the number of observations in a terminal leaf should be at least 5% of the overall data set. Hence we implicitly set the maximal number of bins for each covariate to 20. Obviously, deeper trees increase the number of category levels rapidly and thus are prone to overfitting. Hence, this seems to be a reasonable and robust restriction. With this transformation of the continuous covariates, we can model more complex trends, rather than modelling just the overall trend by estimating a single parameter for those numeric covariates. This pre-processing step implies that the subsequent model calibration takes less than 10 minutes on a standard computer.

Note that the data set (and other typical data sets in the context of lapse analysis) contains information over several observation years. In each observation year,

<sup>1</sup> For the products in the considered portfolio, this quantity specifies a predefined fixed value by which the premium increases each year (0–10%).

every contract may lapse. For modelling purposes, a separate row is created for each in force contract in each observation year. As a consequence, one single contract may occur in several rows in the data set—once for each observation year where the contract is still in force. Note that by this modification, the rows of the data set are no longer fully independent. It also causes a selection bias, since contracts who lapse early on end up having less observations in the data set. Given the size of the data set, this seems acceptable. A thorough analysis of the impact of this assumption seems advisable.

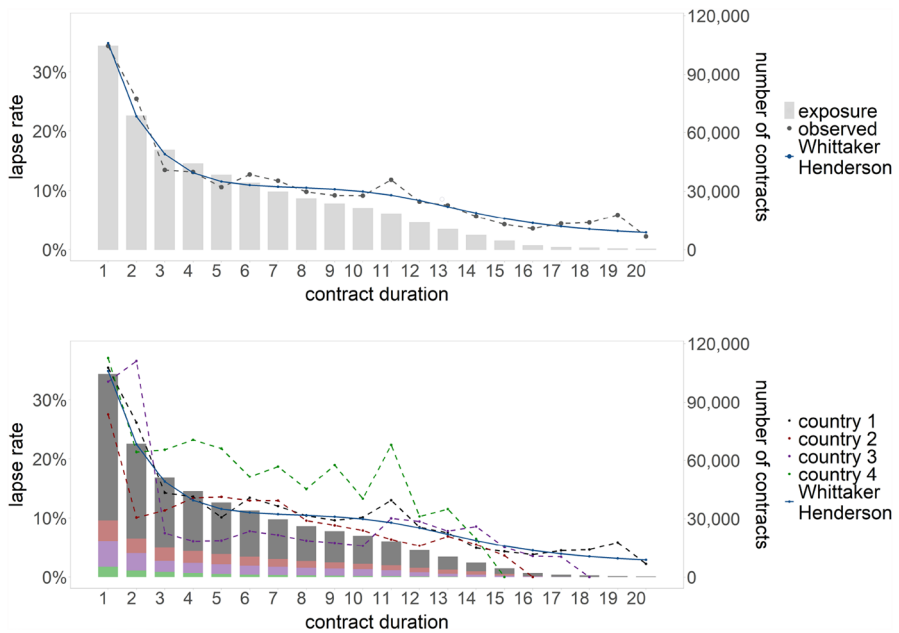
Unlike other papers, e.g. Eling and Kiesenbauer [10], we do not consider the calendar year as a covariate. Even though it may have an impact on lapse rates in specific situations (e.g. higher lapse rates during a financial crisis), it cannot be used as a covariate for future lapse rates. Therefore, for risk management, we explicitly exclude the calendar year from the model and focus on the contract duration to capture the development over time. However, to quantify the calendar year effect, we include the calendar year covariate and other macroeconomic covariates ((AMECO swap rate [3], ECB inflation rate [8] and ECB Eurostox performance [7]) in a sensitivity analysis in Sect. 4.

Overall, this analysis requires only limited pre-processing of the data set, since the Lasso itself performs the main variable selection.

## 2.2 Common actuarial practice

In practical business applications, it appears to be common practice to apply univariate smoothing algorithm for modelling biometric assumptions, e.g. the so-called Whittaker–Henderson approach. It was derived around 100 years ago, as described in Joseph [16]. Applications of this smoothing technique are manifold. For example, the German actuarial society (DAV) uses Whittaker–Henderson to smooth the crude death rates to derive the mortality rates (DAV2008T, without security loadings) used in most German life insurance companies. Another straight-forward and commonly used application of Whittaker–Henderson is the modelling and smoothing of lapse rates over time, see for example SAV [28]. Here, the raw lapse rates for the different values of the covariate contract duration are used to fit a smooth curve, resulting in estimates for the lapse rate depending on the contract duration. The upper panel of Fig. 1 shows the smoothed lapse rates for the analysed insurance portfolio. The bars show the exposure (right y-axis), and the points (connected with the dashed line) show the observed lapse rates for each contract duration (left y-axis). There is a decreasing trend in lapse rates, meaning that more contracts are terminated at the beginning of the contract, which appears intuitive. The blue line is the result of the Whittaker–Henderson smoothing approach. At first glance, the approach works quite well, since it captures the main trend, does not seem to overfit, and is very simple.

However, the Whittaker–Henderson approach is a univariate approach and therefore can only capture the effect of one covariate on lapse rates. Hence, it does not predict lapse rates on an individual policy level, as simply all insured with identical contract duration get equal predictions. To illustrate this, we add



**Fig. 1** Whittaker–Henderson smoothed lapse rates for one covariate (contract duration—blue line) visualised for one covariate (contract duration—upper panel) and for two covariates (contract duration and country—lower panel) (color figure online)

the country as second covariate. The lower panel of Fig. 1 shows the result of smoothing the covariate contract duration with the Whittaker–Henderson approach (blue line) visualised for two covariates (contract duration and country). The points (connected with a dashed line) show the observed lapse rates for each contract duration for the different countries and the blue line—which has not changed—shows the result of the Whittaker–Henderson approach. Clearly, the countries show significantly different lapse rates—in both magnitude and shape. In order to model structures with several covariates, perhaps even on an individual policy level, sub-groups have to be analysed independently. In this example of adding the covariate country, we would have to build four individual Whittaker–Henderson curves for each of the sub-groups (countries).

First of all, these sub-portfolios are subjective and require expert judgement. Secondly, the approach still ignores all other covariates. The problem is not only ignoring some information, but also the risk of misinterpreting the information. One might see and interpret trends in the univariate perspective that actually come from other covariates which are not included. For instance, it is possible that the contracts in country four show higher lapse rates because they belong to a specific entry age group or insurance type that has a higher lapse risk. Finally, another problem in building sub-groups is that the segments may become very small and hence the number of observations falling in a specific segment is not sufficient to derive a reliable estimate.



### 3 Methodology

In this section, we introduce a novel way of modelling the lapse rates of an insurance portfolio on an individual contract level. Our method is based on GLMs, i.e. we apply a multivariate approach to analyse lapse rates depending on several covariates. Introduced by Nelder and Wedderburn [25], this class of models is widely used in actuarial research (see Haberman and Renshaw [13]) and also in actuarial practice. We omit a detailed description here and focus on aspects that are required for our specific method. See for instance McCullagh and Nelder [22] for a more detailed description of GLMs.

#### 3.1 The GLM for actuarial modelling

The GLM models the expected value of a target variable  $Y$ , a random outcome according to a distribution of the exponential family, as a linear combination of the explanatory covariates  $X_i$ :

$$E(Y) = g^{-1}(\beta_1 + \beta_2 \cdot X_2 + \dots + \beta_p \cdot X_p) = g^{-1}(X\beta).$$

The link function  $g$  connects the linear combination of the explanatory covariates to the target variable—hereafter lapse. The choice of the link function must reflect possible entry values of the target variable, in this case lapse ( $Y = 1$ ) or no lapse ( $Y = 0$ ). This can be achieved with the logit link,  $g(x) = \log(x/(1-x))$ , which is also the natural link function implied by the exponential family. Other link functions are possible, e.g. a probit link or a cloglog link function. However, the natural link function has some advantages, e.g. unbiased estimates, see Wuthrich [34]. The parameters of the model ( $\beta$ ) can be estimated efficiently by optimising the likelihood, assuming a Binomial distribution for the target variable lapse. Maximising the likelihood is equivalent to minimising the negative of the log-likelihood:

$$-\log L(\beta|X, Y) = -\sum_{i=1}^n Y_i \cdot \log g^{-1}(X\beta) + (1 - Y_i) \cdot \log(1 - g^{-1}(X\beta)). \quad (1)$$

Based on this likelihood function, we can use the residual deviance as a performance measure. It is a generalisation of the mean squared error (MSE) for normally distributed data to other distributions. The residual deviance is two times the difference of the log likelihood of a fully saturated model that fits the data perfectly and the model under consideration, i.e. a measure for residuals for arbitrary distributions. Starting from a model that only contains an intercept (so-called intercept only model with the

null deviance), one can analyse how much better a model is compared to other models and also compared to the intercept only model.

The calibration of a competitive GLM with best possible residual deviance requires further modifications of the covariates.<sup>2</sup> For instance, a GLM that models the covariate entry age with only one coefficient typically shows a poor performance (underfitting). The model can then only detect one general trend for that covariate but cannot incorporate the different effects on lapse for different entry age groups (marginals), say if the lapse rate is increasing up to a certain entry age and decreasing afterwards. Conversely, a GLM that determines an estimate for each entry age individually will presumably overfit, i.e. the performance on new test data will be considerably worse than on training data. Therefore, a covariate like entry age is typically divided into bins, in order to achieve a more accurate model using pre-specified entry age groups:  $\beta_{\text{entry age}} \cdot X_{\text{entry age}} = \beta_{\text{entry age}}^1 \cdot X_{\text{entry age}}^1 + \dots + \beta_{\text{entry age}}^{p_{\text{entry age}}} \cdot X_{\text{entry age}}^{p_{\text{entry age}}}$ , where  $p_j$  is the number of categories for covariate  $j$ , e.g. the number of entry age intervals for the covariate entry age. Here,  $X_{\text{entry age}}^j = 1$  if the entry age of an observation lies within the  $j$ -th entry age interval and zero otherwise. Alternatively, Generalised Additive Models (GAM) can be used to incorporate functional transformations using splines or polynomials of the covariates.

However, in both approaches the selection of the number of entry age groups, or the number of splines/polynomials is crucial and needs to be specified priorly. Without deeper knowledge of the structures of the marginals, it is hard to identify the ‘correct’ number and comes along with a considerable effort and model risk. In a multivariate model, the ‘correct’ number of groups also depends on the other covariates in the model. Therefore, the process of identifying a good and robust multivariate model is even more complex and only possible with a high effort by iteratively refitting and readjusting the GLM. Hereafter, we show how an extension of the GLM can solve this problem. We will see that the predictions of our lapse model remain fully explainable, and the coefficients can be used to interpret the lapse rate on an individual contract level, as in a standard GLM. Note that most other machine learning techniques lack one or even more of these properties. Often, higher accuracy in forecasts comes along with a loss of interpretability or explainability. For many applications in the insurance sector, however, the ability to explain and interpret the models outcome is of decisive importance, e.g. in risk modelling or pricing.

One of the major challenges in training a competitive model is to avoid overfitting. Most popular in statistical learning is the use of cross-validation, where several models are built and compared with respect to their accuracy in order to choose the best model among them. The  $k$ -fold cross-validation (see James et al. [15]) splits the data set randomly in  $k$  distinct groups, where  $k = 5$  or  $k = 10$  are typical choices. Afterwards, each model is built on  $k - 1$  data sets (training data) and validated on the remaining data set (validation data), by using the deviance of the model on the validation data set. This step can be repeated  $k$  times, and each time, a different group is treated as the validation data set. For each model, the  $k$ -fold cross-validation gives  $k$  estimates of a performance

<sup>2</sup> The design matrix  $X$  is typically standardised, otherwise the magnitude of  $X_j$  can affect the estimate for  $\beta_j$ .

measure on a changing validation data set. Then one can decide which configuration of the model performs best on average.

In addition to the optimum, the  $k$ -fold cross-validation gives an estimate for the variation of the performance measure. Similar to the optimum, this can be estimated directly from the  $k$ -fold repetitions. When choosing a model, the optimum is clearly a plausible choice. But sometimes it can also be advantageous to choose a model that is less complex, e.g. a GLM with fewer coefficients. As long as this model is still within the expected variation of the performance measure, this is also statistically justified. A frequent choice here is the one-standard-error (1-SE) rule, so choosing the least complex model which is still within one standard error from the optimum. A model with fewer coefficients and a similar performance may be preferable, since the model is likely to be more robust and to recognise the patterns in the data well without overfitting. The selection of a robust model also implies that a similar performance can be expected on new data.

In contrast to the beginnings of statistical learning applications, nowadays there is usually a lot of data available. As a result, it can often happen that, through the application of cross-validation alone, models are selected that still have a large number of parameters, e.g. coefficients in the GLM. In addition to a good cross-validation performance it is therefore often important to select a parsimonious model with fewer parameters.

The Lasso is an extension of the GLM using regularisation and is able to handle overfitting and generating a parsimonious model simultaneously. Here, the number of parameters is reduced by a penalty term and a more robust and stable model is developed through the parameter estimation itself.

### 3.2 The Lasso

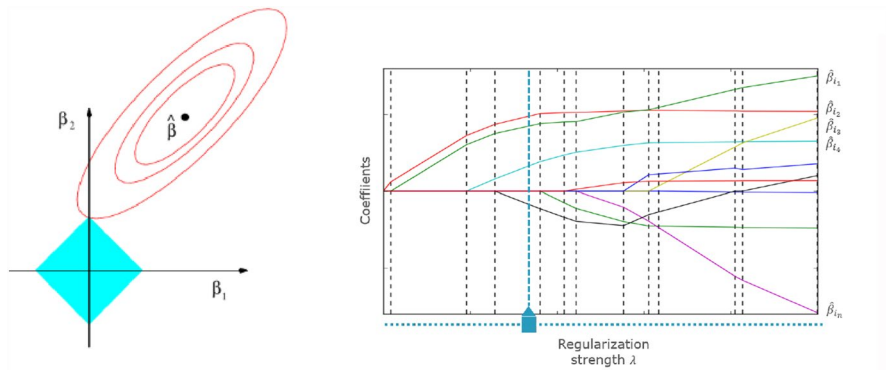
Introduced by Tibshirani [29], the least absolute shrinkage and selection operator (Lasso) is a GLM with regularisation. Lasso adjusts the maximum likelihood estimation of the GLM by adding a penalty term. The optimisation function extends to:

$$-\log L(\beta|X, Y)_{Lasso} = -\log L(\beta|X, Y) + \lambda \sum_{j=1}^J g_{L_j}(\beta_j). \quad (2)$$

We use the index  $L_j$  to specify different version of the Lasso for each covariate.  $R$  stands for regular Lasso where the penalty term is defined as:

$$g_R(\beta_j) = \sum_{i=1}^{p_j} |\beta_{j,i}| = \|\beta_j\|_1.$$

Note that this penalty term equals the  $L_1$  norm of the coefficient vector  $\beta_j$ . Other norms are also possible, e.g. the  $L_2$  norm, which corresponds to the Ridge Regression, see Hoerl et al. [14]. Optimising Eq. (2) will result in a similar GLM as in Eq. (1) but with different values for  $\beta$ . The Lasso has advantageous properties: For some values of  $\lambda > 0$ , selected coefficients will be set to zero, and thus they will be



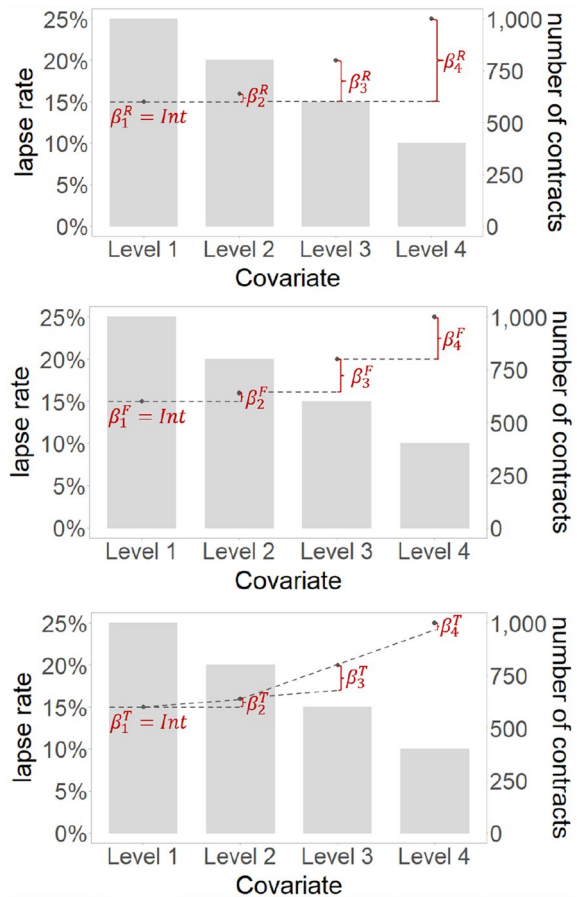
**Fig. 2** Lasso shrinking coefficients to zero with regularisation. Left image from James et al. [15, page 224]

effectively excluded from the model. Therefore, Lasso reduces the complexity of a model systematically. Here, the regularisation shrinks the  $\beta_j$  of different covariates simultaneously towards zero, as the left panel of Fig. 2 shows for two covariates with parameters  $\beta_1$  and  $\beta_2$ . The  $L_1$  norm as a penalty term implies that the optimum for different values of  $\lambda$  (visualised by the ellipses) intersects the penalty term (diamond) where some of the coefficients are zero, here  $\beta_1$ . The amount of shrinkage is controlled by the tuning parameter  $\lambda$ . Set to zero, the resulting GLMs of Eqs. (1) and (2) coincide. Set very high, the resulting GLM of Eq. (2) will be an intercept only model, see right panel of Fig. 2. The specification of a parameter  $\lambda$  that produces a robust GLM in-between those extremes can be done with cross-validation. Note that compared to cross-validating several different GLMs with various parameters, the Lasso requires a cross-validation for only one tuning parameter. Cross-validation can show which values of the tuning parameter  $\lambda$  perform well, measured by the residual deviance of the model. The 1-SE rule gives the range of  $\lambda$  where the performance is within one standard error of the cross-validated optimum. Within this range, higher values of  $\lambda$  shrink even more coefficients to zero and produce easier and probably less complex models without a significant loss of performance. In what follows, we will use a 5-fold cross-validation with this 1-SE rule to determine the tuning parameter  $\lambda$ .

### 3.3 The fused Lasso and trend filtering

The coefficients in the so-called fused Lasso, see Devriendt et al. [6] and Tibshirani et al. [33], do not reflect the difference of a category level to the intercept (upper panel of Fig. 3) but to the adjacent category level, cf. middle panel of Fig. 3. Without regularisation, this formulation results in identical predictions. After regularisation, however, some of the  $\beta_j$  will again be zero, thus adjacent category levels, e.g. entry age groups are fused together forming larger entry age groups. This means that the originally very time-consuming task of identifying the ‘correct’ number of entry age groups can be replaced by an automated and fully data-driven process.

**Fig. 3** Comparison of the effect of different versions of the Lasso (regular, fused, trend filtering)



The number of observations in the category levels has an important impact on the estimation. Since category levels with many observations have more weight in the (penalised) maximum likelihood estimator, the model will particularly focus on these areas. Adjacent category levels with fewer observations are combined more frequently by the fused Lasso and receive a common estimator. Hence, the fused Lasso is also well suited for smoothing the oscillating edges of covariates with fewer observations which is often a challenge in practical applications.

The fused Lasso ( $L_j = F$ ) applies the  $L_1$ -penalty to the differences of subsequent coefficients:

$$g_F(\beta_j) = |\beta_{j,1}| + \sum_{i=2}^{p_j} |\beta_{j,i} - \beta_{j,i-1}| = \|D^{(1)}\beta_j\|_1.$$

Similar to the notation in Tibshirani [31], the  $p_j \times p_j$  dimensional transformation matrix  $D^{(1)}$  can be expressed as<sup>3</sup>:

$$D^{(1)} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & & & \ddots & & \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix}.$$

Note that the coefficient  $\beta_{j,1}$  (corresponds to the first row of the matrix) is typically not penalised, since it is part of the intercept. For the fused Lasso, the choice of the category level which serves as the intercept does not change the model results.

An equivalent formulation uses a contrast matrix, which is defined as the inverse of the transformation matrix:

$$C^{(1)} = D^{(1)-1} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 1 & 0 & \dots & 0 & 0 \\ 1 & 1 & 1 & \dots & 0 & 0 \\ \vdots & & & \ddots & & \\ 1 & 1 & 1 & \dots & 1 & 0 \\ 1 & 1 & 1 & \dots & 1 & 1 \end{bmatrix}. \quad (3)$$

Using the fused Lasso penalty for covariate  $j$  then leads to the transformation of the design matrix  $X_j^{new} = X_j C^{(1)}$ .

In practical applications, there are often covariates that show a monotonically increasing or decreasing behaviour with respect to the target variable. Fusing adjacent category levels together can be insufficient to model this monotone behaviour precisely. In this case, common actuarial practice is to use GAMs, i.e. to model these effects with splines or polynomials. As mentioned above, this comes along—again—with the crucial and complex choice of the ‘correct’ degree of polynomials or the knots of splines. The trend filtering, see Kim et al. [19], was developed for these applications and is another extension of the Lasso. It works similar as the fused Lasso, but also allows for non-zero slopes. So rather than penalising the change between adjacent levels, this approach penalises the change of the slope between category levels, cf. lower panel of Fig. 3. Hereby, changing linear trends and monotone structures within a covariate can be identified efficiently. The trend penalty can be expressed as:

<sup>3</sup> In our notation, we use square matrices including the intercept. There are other papers in the literature that omit the first row.

$$g_T(\beta_j) = |\beta_{j,1}| + |\beta_{j,2} - 2\beta_{j,1}| + \sum_{i=3}^{p_j} |\beta_{j,i} - 2\beta_{j,i-1} + \beta_{j,i-2}| = \|D^{(2)}\beta_j\|_1.$$

In the formula,  $|\beta_{j,1}|$  is treated as the intercept (which is not penalised), while  $|\beta_{j,2} - 2\beta_{j,1}|$  penalises the change in trend compared to the initial trend (with slope 0)—again see lower panel of Fig. 3. The intercept is rather arbitrary and could also be set to any other category level of the covariate, e.g. the middle category level or the last category level. Since the term  $|\beta_{j,2} - 2\beta_{j,1}|$  is penalised, the choice of the initial trend does affect the result of the model.

Here, the  $p_j \times p_j$  dimensional transformation matrix  $D^{(2)}$  can be expressed as:

$$D^{(2)} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ -2 & 1 & 0 & \dots & 0 & 0 & 0 \\ 1 & -2 & 1 & \dots & 0 & 0 & 0 \\ \vdots & & & \ddots & & & \\ 0 & 0 & 0 & \dots & -2 & 1 & 0 \\ 0 & 0 & 0 & \dots & 1 & -2 & 1 \end{bmatrix}.$$

In terms of the contrast matrix, we can write:

$$C^{(2)} = D^{(2)-1} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 2 & 1 & 0 & \dots & 0 & 0 \\ 3 & 2 & 1 & \dots & 0 & 0 \\ \vdots & & & \ddots & & \\ p_j - 1 & p_j - 2 & p_j - 3 & \dots & 1 & 0 \\ p_j & p_j - 1 & p_j - 2 & \dots & 2 & 1 \end{bmatrix}.$$

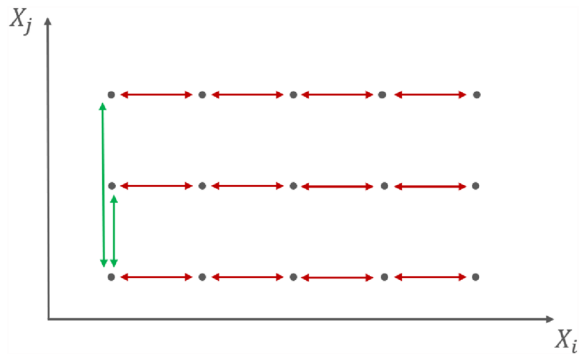
Using the trend filtering penalty for covariate  $j$  then leads to the transformation of the design matrix  $X_j^{new} = X_j C^{(2)}$ .

We could also consider other functions in a similar setting. Note that the transformation matrix for higher orders can be defined recursively, see Tibshirani [31], i.e.  $D^{(k)} = D^{(k-1)}D^{(1)}$ . Thus, also higher orders can be included straightforwardly, e.g. with cubic structures. Again, the time-consuming task of identifying the ‘correct’ degree or knots can be replaced adequately by an automated and fully data-driven process.

### 3.4 Modelling interactions with Lasso

Multivariate models like a GLM can jointly assess the impact of covariates on the target variable. This allows to identify the factors triggering a high or low estimate of the target variable. Sometimes, the influence of a covariate on the target variable also depends on the value of another covariate. In that case, common actuarial practice is to include an interaction term in the model. Including an interaction between covariate  $X_i$  and  $X_j$ , the GLM formula extends to:

**Fig. 4** Combination of two covariates with an interaction (regular in  $X_j$ , fused in  $X_i$ )



$$E(Y) = g^{-1}(\beta_1 + \beta_2 \cdot X_2 + \cdots + \beta_{i,j} X_i \cdot X_j + \beta_p \cdot X_p) = g^{-1}(X\beta).$$

By this definition, the model gets  $p_i \cdot p_j$  additional coefficients. Through the interaction of  $X_i$  with  $X_j$ , further structures can now be included in the model. For example, if adjacent category levels of  $X_i$  show differences depending on their value of  $X_j$ . E.g. for certain values of  $X_j$  the observed lapse rates may be increasing in  $X_i$ , whereas for other values of  $X_j$  the observed lapse rates may be decreasing in  $X_i$ . This can be reflected in the model with an interaction term.

Again, the choice of which interaction to include in a model is crucial and comes along with a considerable effort. Here, especially overfitting is a potential risk, as interactions incorporate a quadratic number of additional coefficients in the model. The Lasso proves to be a helpful tool for including only the most important coefficients (here interactions) in the model. For example, we may apply the regular Lasso to these coefficients, i.e.

$$g_{R,R}(\beta_{i,j}) = \sum_{k=1}^{p_i} \sum_{l=1}^{p_j} |\beta_{i,j,k,l}| = \|\beta_{i,j}\|_1, \quad (4)$$

where  $p_i$  ( $p_j$ ) is the number of different category levels of covariate  $X_i$  ( $X_j$ ). After regularisation, only the most important coefficients remain in the model. This can significantly reduce the number of coefficients (here interactions) in the model. Yuan and Lin [37] apply a grouped Lasso, where all or no coefficient of a covariate remain in the model. For our data set, however, the grouped Lasso leads to overfitting as too many parameters remain in the model.

Another example of penalising the interaction term is in case of possible levels or trends in  $X_i$  or  $X_j$ . We can then apply a fused Lasso or a trend filtering approach for one direction (or both). This is illustrated in Fig. 4 for an interaction of  $X_i$  and  $X_j$ . Say that in the data,  $X_j$  is not ordinal scaled and hence the regular Lasso is used for that direction (y-direction). Hence, we compare each coefficient with the intercept—visualised by the green arrows.  $X_i$  however is ordinal scaled and for the interaction term, a fused Lasso penalty is chosen for that direction (x-direction). Hence, we compare each coefficient with its subsequent coefficient—visualised by the red arrows.



Similar to Eq. (4), we can formalise the penalty term for this example:

$$g_{F,R}(\beta_{ij}) = \sum_{l=1}^3 |\beta_{ij,1,l}| + \sum_{k=2}^5 \sum_{l=1}^3 |\beta_{ij,k,l} - \beta_{ij,k-1,l}|.$$

Note that for this example, we assumed that  $\beta_{ij,1,1}$  is included in the intercept and is therefore not penalised.

Fusing in both directions  $X_i$  and  $X_j$  leads to the 2d-fused Lasso, see again Devriendt et al. [6] and Tibshirani et al. [33]. This approach can be generalised using an arbitrary graph indicating which two points in the plane should be considered for the fusing. One example is given in Tibshirani et al. [33], where they use a graph with the US (mainland) states as nodes and edges whenever two states share a border. One difficulty, however, is that in general, the number of edges is greater than the number of nodes. Consequently, we need additional second-order conditions to obtain a unique optimum. In order to avoid this case and the more complex optimisation associated with it, we have chosen a regular Lasso for one dimension in the numerical example in the next section.

## 4 Results

### 4.1 Implementation

The implementation uses the R [26] interface for **h2o**, see LeDell et al. [20]. Note that the R package **smurf**, see Reynkens et al. [27], which is based on the analysis of Devriendt et al. [6] already introduces different regularisation types within a Lasso setting (SMURF algorithm). However, it does not contain the trend filtering, so that we ended up developing a new implementation within the **h2o** framework.

For covariates with fused Lasso or trend filtering penalty, we use the contrast matrices  $C^{(1)}$  (fused Lasso) and  $C^{(2)}$  (trend filtering) to adjust the structure in the design matrix, as described in Sect. 3. To be more precise: We apply the contrast matrices to the original design matrix  $X$  using the `contrast` argument in the R built-in `model.matrix` function. By that, the corresponding columns for each covariate are changed accordingly. This is equivalent to:

$$X^{new} = XC, \quad C = \begin{bmatrix} C_1 & 0 & 0 & \dots & 0 \\ 0 & C_2 & 0 & \dots & 0 \\ 0 & 0 & C_3 & \dots & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \dots & C_J \end{bmatrix}, \quad (5)$$

where  $C_i$  is the  $p_i \times p_i$  contrast matrix for the  $i$ -th covariate and 0 is a matrix of zeros of corresponding size.

The **h2o** command `h2o.glm` for building the actual model is then straightforward. The necessary arguments are the response (`lapse`), the design matrix ( $X^{new}$ ), the family (binomial), the penalty type (Lasso), the number for the cross-validation

(=5) and the lambda search (true), which automatically screens a range of possible values for  $\lambda$ . By default, the command standardises the data set. Note that the `h2o.glm` command itself uses the regular Lasso and does not account for different penalty types. This has already been done by adjusting the design matrix accordingly. There are also other implementations for Lasso in R. We chose **h2o**, since it outperforms the **smurf** package with respect to computing time. Additionally, it also satisfies the consistency condition, that the fit for  $\lambda = 0$  coincides with a GLM fit. Other packages like **glmnet**, see Friedman et al. [12] do not satisfy this condition.

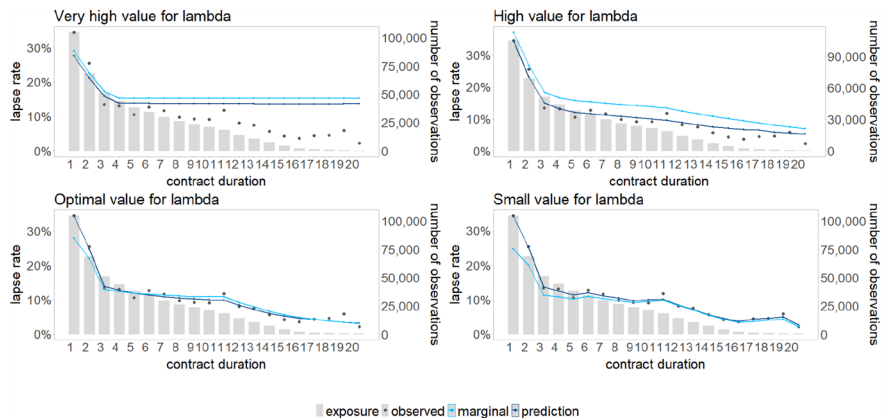
## 4.2 Model selection

We apply the different versions of the Lasso to model the lapse rates of a European insurance portfolio, see Sect. 2. For each covariate, one of the regularisation types regular Lasso, fused Lasso or trend filtering needs to be specified. There are some recommendations for the choice of the regularisation type but no strict rules. For example, binary or unordered nominal covariates are typically modelled with a regular Lasso. Otherwise, the arbitrary selection of the order of the category levels would impact the predictions. Ordinal covariates with many category levels often perform best with a fused Lasso that forms groups with joint coefficients. Note that a certain degree of expert judgement is required for ordinal covariates, especially when there are only a few category levels. It is difficult to specify whether or not an effect of a covariate follows monotone trends with slopes not equal to zero before analysing the data and building a model. The decision between fused Lasso and trend filtering is essentially a choice whether adjacent categories should be fused together to one level or one common trend. Hence, the model specification should always be checked carefully for effects in a multivariate setting. An example for this analysis is the covariate entry age, which appears to exhibit an increasing trend at the beginning and a negative slope for the other entry age groups in a univariate view, as shown in Fig. 6. However, we decided to fuse different entry age groups together since this appeared more adequate after accounting for the impact of other covariates (light blue line).<sup>4</sup>

We can summarise the individual penalty types for each covariate in the model formula:

```
lapse ~ gender R + country R + insurance type R +
payment method R + native R.
+ payment frequency F + entry age F
+ contract duration T + dynamic premium increase per-
centage T + sum insured T
+ original term of the contract T + premium payment
duration T + yearly premium T
```

<sup>4</sup> Further reasons can motivate one or the other regularisation type beyond goodness of fit. For instance, the limitations of a pricing system for an insurance tariff may only offer a limited amount of different possible tariff cells. Or the sales team requires a stable premium over several contract years. In all cases, fused Lasso can be applied for grouping covariates.



**Fig. 5** Result of the model without interactions for different values of  $\lambda$  for the covariate contract duration

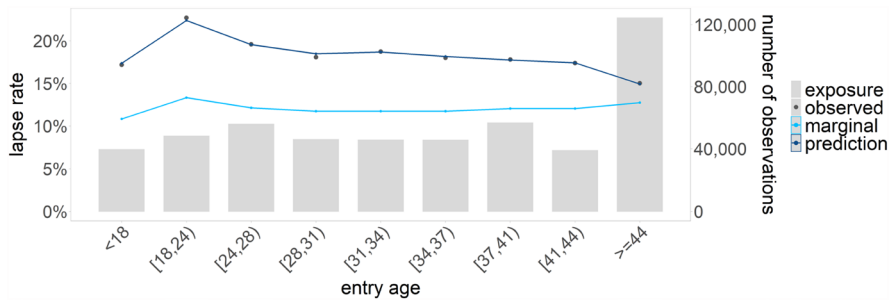
Here, the letters T, F and R denote trend filtering, fused Lasso and regular Lasso, respectively. For the trend filtering, we start with the category level in the middle. This decision can be justified by the exposure (starting in an area with higher exposure) or by the expected trend from a univariate analysis of the covariate. A sensitivity analysis shows that this decision has very little impact to our data set.

Based on Zou [38], Devriendt et al. [6] propose different penalty weights for the different regularisation types, since standard Lasso can lead to inconsistent variable selection. However, these weights have little influence for this data set and we chose to not use any penalty weights in this analysis.

### 4.3 The shrinkage factor $\lambda$

As explained in Sect. 3.2, the shrinkage factor  $\lambda$  significantly changes the result of the model, where an optimal value can be estimated by cross-validation. We want to visualise this effect with Fig. 5, where results of four different values of  $\lambda$  (very high—high—optimal—low) are shown for the covariate contract duration. The panels have the same structure as in Fig. 1, so the bars show the exposure and the points show the observed lapse rates. However, this time the results of the model are two-fold, as the GLM is a multivariate model. The dark blue line shows the average prediction for each contract duration and the light blue line shows the marginal effect of the covariate, i.e. the  $\beta$  for all relevant contract durations (all other  $\beta$ 's are manually set to zero for this visualisation). In a good model, the dark blue line should reflect the points reasonably well. From the shape of the light blue line, it is possible to explain and interpret the impact of the covariate, here contract duration, on lapse rates.

The top left panel shows the results of the model for a very high value of  $\lambda$ , i.e. a large penalisation of coefficients. Since the penalty is high, we get a very simple model (low variance). However, we are not able to detect the main trends of the covariate (high bias) and presumably end up underfitting. The fact that the dark blue



**Fig. 6** Result of the model without interactions for  $\lambda$  based on the 1-SE rule—visualised for the covariate entry age

**Table 1** Coefficients of the covariate entry age

Level	<18	[18, 24)	[24, 28)	[28, 31)	[31, 34)	[34, 37)	[37, 41)	[41, 44)	>=44
Coefficient	- 2.109	0.237	- 0.107	- 0.038	0.000	0.000	0.029	0.000	0.065
Coefficient cumulated	- 2.109	- 1.873	- 1.980	- 2.018	- 2.018	- 2.018	- 1.989	- 1.989	- 1.923
Marginal lapse rate	0.108	0.133	0.121	0.117	0.117	0.117	0.120	0.120	0.127

line and the light blue line are not identical shows that there are other covariates with coefficients not equal to zero. Decreasing  $\lambda$  will reduce the penalisation, and hence more and more coefficients remain in the model. This can be seen in the panel on the top right. Here, the effect of the trend filtering can be seen, since certain areas of the light blue line follow a trend and in between are several trend changes, for example in contract duration eleven. Remember, that all these trends and trend changes are identified purely data driven and are not based on any manual adjustments. The bottom left panel shows the result for  $\lambda$  based on the 1-SE rule, which we consider the optimal  $\lambda$  for this illustration. Based on visual inspection, the model seems to be able to find the main trends without overfitting the data. From the shape of the marginal effect, we can learn that higher contract durations imply a lower lapse rate. However, the degree of reduction is not constant but is tremendously reduced after three years. If we further decrease lambda and thus further reduce penalisation, the model appears to overfit the data set (as shown in the bottom right panel). The overall prediction is good (low bias), but the model is probably no longer robust (high variance). Note that for the case  $\lambda = 0$ , Eq. (2) reduces to Eq. (1), and we have a logistic regression model without any penalisation and presumably no coefficient would be set to zero.

Note that the shrinkage parameter  $\lambda$  is fitted using cross-validation in one step for all covariates jointly. This is one of the main advantages of the Lasso specification in this paper, since it replaces the iterative process of fitting a GLM by a one-step automated process.

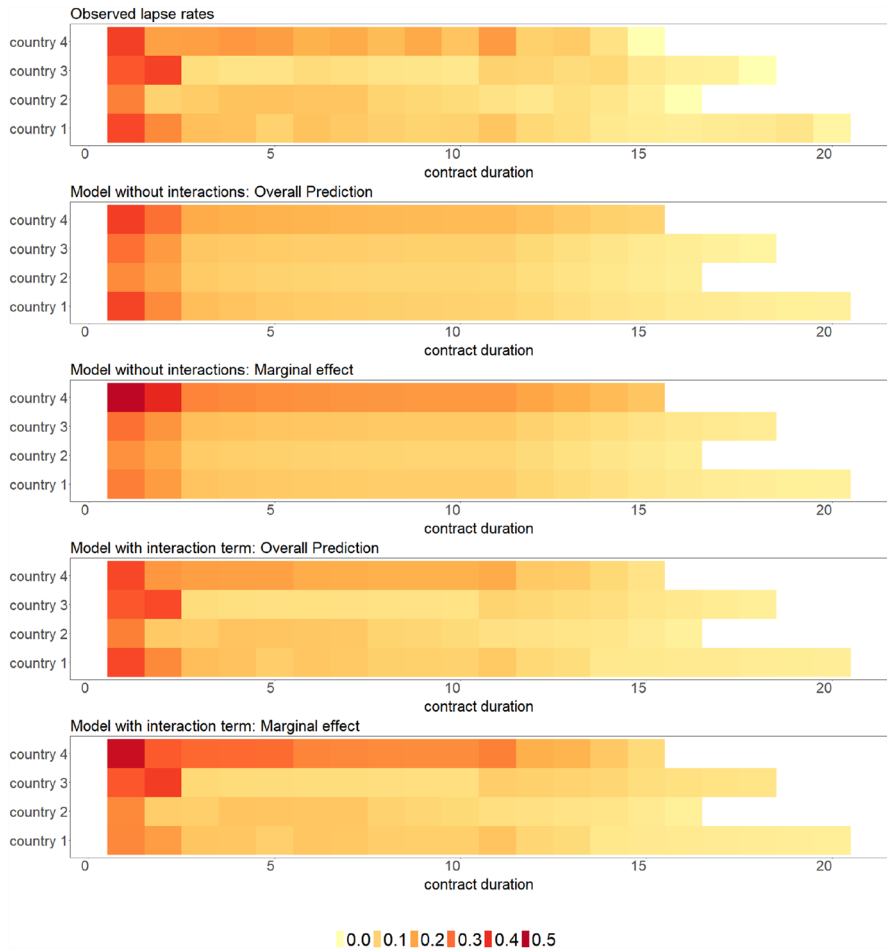
The covariate entry age is modelled with a fused Lasso in order to find entry ages that can be grouped together beyond the univariate pre-processing in Sect. 2. Figure 6 shows the result for the optimal value of  $\lambda$ . Here, the shape of the light blue line and the dark blue line differ. This illustrates that a univariate approach can be misleading for such a covariate as the slopes of the two lines have different signs in some areas. Table 1 shows the coefficients for entry age. To be more specific: the first row (coefficient) corresponds to the actual coefficients from the model. The second row (coefficient cumulated) translates those coefficients according to the contrast matrix we used for that covariate—here fused Lasso, see Eq. (3). The third row (marginal lapse rate) then applies the inverse of the logit function and therefore gives the marginal lapse rate, which corresponds to the light blue line from Fig. 6. This highlights the advantage of our approach since each coefficient can be explained and interpreted.

The fused Lasso identifies many coefficients to be 0 and therefore effectively removes them from the model. The corresponding category levels are fused together, and the light blue line is constant. However, a closer inspection of the coefficients shows that some of the grouped coefficients are very close to zero, yet not equal to zero, see for example the coefficient for the level [37, 41). This phenomenon of the Lasso is already known from other applications, see Zou [38]. The reason for this is that the  $L_1$  norm is the best approximation of a sparse solution with a convex optimisation, where the actual sparse solution would be with respect to the  $L_0$  pseudo-norm. In a comment on the presentation of Tibshirani [30], Peter Bühlmann, therefore, suggested to “interpret the second ‘s’ in Lasso as ‘screening’ rather than ‘selection’”. For the modelling of lapse rates, this phenomenon may not per se be problematic. If a parsimonious model is required or desirable, the screened parameter can be adjusted manually, as we will illustrate in Sect. 4.5.

#### 4.4 Interactions

So far we have not included any interaction terms, leading to a model with one estimated coefficient for each category level of each covariate. But what if the lapse rate for one category level depends on the category level of another covariate? This appears to be the case for the covariates country and contract duration. The top panel of Fig. 7 shows the observed lapse rates for different values of contract duration on the x-axis and for the four countries on the y-axis. We observe a decreasing trend for all countries with increasing contract duration. However, as in Fig. 1, both level and slopes differ by country.

For example, the observed lapse rate in country four is higher for all values of contract duration. A different level could be modelled without interactions, as the coefficient for country four would be higher in the multivariate model. However, structurally different slopes can only be modelled with interactions, as illustrated in the second and third panel of Fig. 7. The second panel shows the overall prediction of the model without interactions which corresponds to the dark blue line in Fig. 5. The third graph shows the marginal effect of the covariates contract duration and country for the model without interactions, which corresponds to the light blue line



**Fig. 7** Observed and estimated lapse rates for model with and without interaction for the covariates contract duration and country

in Fig. 5. The model without interactions captures the overall decreasing trend and the different level of the lapse rates per country, e.g. country four has a higher coefficient. But the different slopes for increasing contract durations for different countries cannot be detected by this model. For example, the estimates for country three appear too high for contract durations three to ten.

To capture these differences, it is necessary to include the interaction term for contract duration  $\times$  country in the model. In our framework, we can include interactions straightforwardly and just need to assign a sub-penalty for this new interaction, see Sect. 3.4. We choose a penalty such that the different countries are penalised with the regular Lasso, since there is no ordinal scale for country, and the different values for contract duration are penalised with the fused Lasso for the interaction term. The model formula can be adjusted accordingly:

```

lapse ~ gender R + country R + insurance type R + pay-
ment method R + native R
+ payment frequency F + entry age F
+ contract duration T + dynamic premium increase per-
centage T + sum insured T
+ original term of the contract T + premium payment
duration T + yearly premium T
+ contract duration F * country R

```

The resulting model with the interaction term  $\text{contract duration} \times \text{country}$  is then able to identify the overall main trend, the different overall level of lapse rates per country and the different slopes for the different countries. This can be seen in the fourth and fifth panel of Fig. 7: Now the different lapse rates in country three are captured.

If necessary, further interactions can be added to the model formula in a similar fashion. However, interactions always come with a significant number of additional parameters. In our example, the interaction  $\text{contract duration} \times \text{country}$  added  $4 \times 20 = 80$  potential additional parameters to the model.<sup>5</sup> Among them, the Lasso can again identify the most important parameters, such that only 35 parameters are effectively added. All parameters (marginal effects and interactions) are again estimated in one step using cross-validation. Sometimes the addition of an interaction parameter may make a marginal parameter obsolete. This may also be justified from a performance point of view as it is controlled by the cross-validation. However, for practical applications it may be desirable that most of the effects are explained by the univariate effects and only structures going beyond that are explained by interaction terms. To achieve this, it is possible to include interactions in two steps. In a first step, a model without interactions is built. This model produces forecasts for each line item. In a second step, interactions can be analysed with a similar Lasso as in step one, but now without the marginal effects (and therefore only the additional interaction terms) and the forecast of the first step as an offset, i.e. a bias for each line item. By such a two-step approach, the model is able to capture the marginal effects within the offset and can then add relevant interaction terms if needed.

To illustrate this, we start with the Lasso model without interactions and add the interaction term  $\text{contract duration} \times \text{country}$ . Again,  $\lambda$  is chosen based on a 5-fold cross-validation using the 1-SE rule. In this setup, only 20 parameters are effectively added.

## 4.5 Comparison and sensitivity analysis

In order to quantitatively assess and compare the different models, we first implement an intercept only model as a baseline. As the name suggests, the intercept only model consists of the intercept but no other covariates. Essentially, the predicted

<sup>5</sup> Actually there are 69 potentially new parameters, since there are some combinations of contract duration and country with no observations, as also seen in Fig. 7.

lapse rate for each observation is just the average lapse rate of the portfolio. Several modelling approaches have been discussed so far: First of all the Whittaker–Henderson approach. We implement one model using only the contract duration, resulting in one curve—the blue line from Fig. 1 and one model using contract duration and country, which would lead to four curves. Secondly, we implement a regular GLM without interactions. Thirdly, we implement the proposed Lasso without interaction terms with  $\lambda$  based on the 1-SE rule. This model serves as the base model afterwards.

We consider several sensitivity analyses:

- the effect of  $\lambda$ , see Sect. 4.3: we exemplary look at the model with the best  $\lambda$  (without the 1-SE rule).
- the effect of different penalty types: we analyse a model using regular Lasso penalty for all covariates and another model using fused Lasso penalty for all covariates.<sup>6</sup>
- the regularisation type: we analyse two elastic net models ( $\alpha = 0.0$ , i.e. Ridge regularisation and  $\alpha = 0.5$ ).
- calendar year and macroeconomic covariates: such variables are not reflected in the base model. For the first model, we add the macroeconomic covariates swap rate, inflation rate and Eurostoxx performance. We map these values using the calendar year, and then remove calendar year from the model. The second model on the other hand adds only the calendar year covariate. We use trend filtering for all new covariates.
- interactions: interaction terms can be analysed with the Lasso efficiently. Here, we include the interaction contract duration  $\times$  country. This is reflected via the design matrix, and alternatively by an offset (based on the predicted lapse rates of the base model). Adding further interactions may further improve the model.
- selection vs. screening: to deal with the phenomenon that the Lasso performs screening rather than selection, we also illustrate manually adjusted Lasso models. The smallest coefficients (in absolute terms) in the base model are manually set to zero. The reduced design matrix is then used for building a GLM (without additional penalty). Hereby, we can analyse the effect of screened coefficients.
- binning: we analyse the binning effect by building a model which uses expert judgement to derive the bins, which means
  - entry age [in years] from 0 to 70 in steps of 5 and everything above 70 as one bin
  - original term of the contract [in years] from 5 to 50 in steps of 5 and everything above 50 as one bin
  - premium payment duration [in years] from 1 to 36 in steps of 5, zero as individual bin and everything above 36 as one bin
  - sum insured [in EUR] from 0 to 150,000 in steps of 10,000 and everything above 150,000 as one bin

<sup>6</sup> Covariates which are nominal scaled are still penalised with the regular Lasso.



**Table 2** Quantitative comparison of different lapse models

Sensitivity	Model	Number of parameters	Residual deviance	$1 - \frac{D_M}{D_0} (\%)$	AUC
Reference models	Intercept only	1	470,092	0.00	0.500
	Whittaker–Henderson (one curve)	20	438,604	6.70	0.678
	Whittaker–Henderson (four curves)	69	436,432	7.16	0.686
	GLM without interactions	77	412,005	12.36	0.732
Proposed base model	Lasso base	44	413,039	12.14	0.731
Lambda	Lasso with best $\lambda$	54	412,578	12.23	0.731
Penalty type	Lasso all regular	70	412,633	12.22	0.730
	Lasso all fused	51	412,685	12.21	0.731
Regularisation type	Elastic net ( $\alpha = 0.0$ )	77	412,437	12.26	0.732
	Elastic net ( $\alpha = 0.5$ )	55	412,987	12.15	0.732
Additional covariates	Lasso with macroeconomic covariates	72	407,499	13.32	0.742
	Lasso with calendar year covariate	74	407,205	13.38	0.743
Interaction term	Lasso with interaction term	79	409,658	12.86	0.740
	Offset Lasso	64	410,300	12.72	0.738
Manual selection	Manual Lasso with 30 parameters	30	413,162	12.11	0.729
	Manual Lasso with 20 parameters	20	418,876	10.89	0.718
	Manual Lasso with 10 parameters	10	434,326	7.61	0.682
Binning	Lasso with expert judgement binning	53	419,662	10.73	0.725

- yearly premium [in EUR] from 0 to 30,000 in steps of 1,000 and everything above 30,000 as one bin.

For the specification of an optimal model with cross-validation, we already used the deviance. Therefore, a measure based on the deviance allows for a consistent comparison of the models. In addition to the deviance, we also consider the quantity  $1 - \frac{D_M}{D_0}$  which is similar to the  $R^2$  measure for normal distributed data. Here,  $D_M$  is the deviance of the model and  $D_0$  is the null deviance, i.e. the deviance of the intercept only model. This quantity can be interpreted as the improvement over the reference model. We also include the area under the curve (AUC) as another well-known measure. Table 2 summarises the results for the full data set.

Among the analysed models, the Whittaker–Henderson approach has by far the worst performance. This is mainly due to the fact that the approach is univariate. Including another dimension, e.g. country, would slightly improve the results, but also the number of estimated parameters is multiplied by the number of covariate

levels (four in the case of the country). Note that some combinations do not contain any observations, see Fig. 7, leading to 69 parameters. This already exceeds the number of parameters of the multivariate Lasso base model with 44 parameters. For the GLM, the number of estimated parameters is quite high (77). Note that we have used the same specification as for the Lasso which implies that for each category level of each covariate a separate parameter is estimated. In order to get a robust GLM, the manual and iterative process of grouping levels would start here.

The alternative proposed in this paper is to use the Lasso to identify these structures automatically. The Lasso model shows a similar good fit (12.14% vs 12.36% improvement over the intercept only model) as the (presumably overfitted) GLM but only has 44 parameters. Using the optimal  $\lambda$  (without the 1-SE rule), the improvement over the intercept only model increases slightly (12.23% vs. 12.14%). However, the number of parameters increases accordingly (54 vs. 44). The models using only one penalty type (all regular and all fused, respectively) show a similar performance with respect to deviance and AUC. But the number of parameters is significantly higher, especially for the all regular Lasso. This is mainly due to the fact that adjacent category levels are not fused together. The sensitivity of the different regularisation types shows a similar effect as the sensitivity of the penalty type: Again, the performance remains very similar and the number of parameters increases, especially for the elastic net with  $\alpha = 0.0$ . Just like the GLM (with no regularisation), this model leads to 77 coefficients, since Ridge does not perform variable selection. Increasing  $\alpha$  to 0.5 reveals the selecting property of the Lasso, leading to 55 parameters.

Adding additional information to the model will almost certainly improve the performance. We analyse the effect of both new covariates and an interaction term of existing covariates: Firstly, the additional macroeconomic covariates improve the model performance (13.32% vs. 12.14%), but also increase the complexity (72 vs. 44 parameters). Interestingly, the model with the additional calendar year covariate performs equally well to the model with additional macroeconomic covariates. It seems like the additional information can mostly be described through the calendar year effect and not through the structure of the macroeconomic covariates. Secondly, the interaction term increases the predictive power (12.86% vs. 12.14%). Again, this additional covariate increases the number of parameters (79 vs. 44). The offset model shows a similar effect (12.72% vs. 12.14% and 64 vs. 44 parameters).

The reduced design matrices used in the manually selected models lead to very parsimonious models with 30, 20 and 10 parameters, respectively and decreasing performances. However, the manual model with 30 parameters shows that the number of parameters can be reduced significantly (30 vs. 44) without losing too much predictive power (12.11% vs. 12.14%). The manual Lasso with only 10 parameters is still better than the Whitaker–Henderson models—increased performance and less complexity. Lastly, the binning effect is analysed with the Lasso with expert judgement binning. We can see that the choice of the bins in the pre-processing has an impact on the results.

## 5 Conclusion

In order to assess lapse risk of a portfolio of life insurance contracts, the determinants of lapse rates need to be identified and reflected in a lapse model. Best estimate lapse rates are derived from such a model and fed into cash flow projection models used for market consistent valuations, e.g. under Solvency II. For the derivation of best estimate lapse rates, the insurance portfolio is typically divided into sub-portfolios based on contract characteristics like type of contract, country, or distribution channel. Lapse rates are then derived for each sub-portfolio independently, considering the dependency on factors like time since inception. A typical example is the Whittaker–Henderson approach. However, ignoring dependencies between these sub-portfolios can lead to inaccurate best estimate assumptions and thus unreliable cash flow projections. To address this, multivariate lapse models have been developed; they model lapse rates on the individual contract level using all available covariates simultaneously. If set up properly, these models can take dependencies between sub-portfolios into account and provide more reliable estimates. However, the specification of a sophisticated lapse model is associated with a considerable effort since it includes the selection of covariates and the identification of structures within a covariate. In fact, adjustments in one covariate potentially require a complete readjustment of all other covariates. Therefore, the increasing number of potential covariates requires a thorough time-consuming analysis and models are still prone to over- or underfitting.

The application of data science methods can replace this largely manual process by a more automated process. In this paper, we use the Lasso variable selection method to derive a lapse model for a European life insurance company. Our proposed approach is based on a novel methodology that combines different versions of the Lasso approach to jointly identify structures in the covariates of a model, including trends, groups and interactions. By that, the model is able to identify structures within covariates, which a regular Lasso or GLM is not able to. In particular, we adapt recently developed extended versions of the Lasso algorithm that apply different sub-penalties for individual covariates. This provides a high degree of flexibility as well as interpretability similar to GLMs and GAMs but can be fitted automatically in only one step. Furthermore, a model selection process based on cross-validation is applied. In contrast to the other models (Whittaker–Henderson/GLM), the model is therefore objectively calibrated based on the data. Thus, the risk of manually triggered overfitting is almost eliminated. Even though the model is fully data-driven, it allows for some flexibility by choice of parameters and modelling decisions for individual covariates. This fine-tuning requires a profound understanding of the underlying algorithms and the specifics of lapse behaviour in life insurance.

The advantages of the method are illustrated based on data from a European life insurer operating in four countries. Our lapse model is compared to alternative lapse models which have been proposed in the literature and are frequently used in practice. We discuss advantages and disadvantages of the Lasso model. In particular, we assess the screening property of the Lasso and show how the model can be further improved by removing selected coefficients that do not improve goodness of

fit. The proposed lapse model outperforms competing lapse models with respect to goodness of fit and parsimony. In contrast to random forests or neural networks, the predictions of our lapse model remain fully interpretable and explainable. Therefore, this research should be of interest to anybody who is concerned with lapse risk, e.g. life insurers, regulators and auditors.

Our analysis points at several fields for further research. One extension could be a more thorough analysis of the calendar year covariate. We already added the calendar year as an additional covariate to the model and saw an improvement of the predictive power. The impact of this finding on the prediction of future lapse rates needs to be assessed in a second step. Another way of using the calendar year information is by weighting the observations according to their calendar year (with higher weight to more recent observations). Another extension may address the issue of extrapolation, in the sense of prediction of lapse rates for realisations of covariates not included in the data set (e.g. contract durations beyond 20). This has high practical relevance for the long-term prediction of future lapse rates. In order to assess the quality of the proposed model, it would also be interesting to apply black box machine learning models and consider their predictive power as a benchmark. It would be promising to extract information from these models to further improve the Lasso model.

Lastly, the proposed methodology may be extended to other types of policyholder options, e.g. the paid-up option and the partial surrender option. Instead of modelling just the two states lapse or no lapse, this would imply that transitions are modelled jointly, as e.g. the drivers for lapse and paid-up rates may be similar.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Aleandri M (2017) Modeling dynamic policyholder behaviour through machine learning techniques. *Scuola de Scienze Statistiche* (**submitted**)
2. Aleandri M, Elettì A (2020) Modelling dynamic lapse with survival analysis and machine learning in CPI. *Decis Econ Finance* 44:37–56. <https://doi.org/10.1007/s10203-020-00285-9>
3. AMECO swap rate (2022) Interest rates, Long term nominal (ILN). [https://dashboard.tech.ec.europa.eu/qs\\_digit\\_dashboard\\_mt/public/sense/app/667e9fba-eea7-4d17-abf0-ef20f6994336/sheet/2f9f3ab7-09e9-4665-92d1-de9ead91fac7/state/analysis](https://dashboard.tech.ec.europa.eu/qs_digit_dashboard_mt/public/sense/app/667e9fba-eea7-4d17-abf0-ef20f6994336/sheet/2f9f3ab7-09e9-4665-92d1-de9ead91fac7/state/analysis). Accessed 16 May 2022

4. Azzone M, Barucci E, Moncayo GG, Marazzina D (2022) A machine learning model for lapse prediction in life insurance contracts. *Expert Syst Appl* 191:116261. <https://doi.org/10.1016/j.eswa.2021.116261>
5. Barucci E, Colozza T, Marazzina D, Rroji E (2020) The determinants of lapse rates in the Italian life insurance market. *Eur Actuar J* 10(1):149–178. <https://doi.org/10.1007/s13385-020-00227-0>
6. Devriendt S, Antonio K, Reynkens T, Verbelen R (2021) Sparse regression with multi-type regularized feature modeling. *Insur Math Econ* 96:248–261. <https://doi.org/10.1016/j.insmatheco.2020.11.010>
7. ECB Eurostoxx Performance (2022) Dow Jones Euro Stoxx 50 Price Index—historical close, average of observations through period. [https://sdw.ecb.europa.eu/quickview.do?SERIES\\_KEY=143.FM.M.U2.EUR.DS.EI.DJES50I.HSTA](https://sdw.ecb.europa.eu/quickview.do?SERIES_KEY=143.FM.M.U2.EUR.DS.EI.DJES50I.HSTA). Accessed 16 May 2022
8. ECB Inflation Rate (2022) EU (changing composition)—HICP—overall index, Annual average rate of change, Eurostat, Neither seasonally nor working day adjusted. [https://sdw.ecb.europa.eu/quickview.do?SERIES\\_KEY=122.ICP.A.D0.N.000000.4.AVR](https://sdw.ecb.europa.eu/quickview.do?SERIES_KEY=122.ICP.A.D0.N.000000.4.AVR). Accessed 16 May 2022
9. EIOPA (2011) EIOPA Report on the fifth Quantitative Impact Study (QIS5) for Solvency II. EIOPA-TFQIS5, 11/001, pp 77–79
10. Eling M, Kiesenbauer D (2014) What policy features determine life insurance lapse? An analysis of the German market. *J Risk Insur* 81(2):241–269. <https://doi.org/10.1111/j.1539-6975.2012.01504.x>
11. Eling M, Kochanski M (2013) Research on lapse in life insurance: what has been done and what needs to be done? *J Risk Finance* 14(4):392–413. <https://doi.org/10.1108/JRF-12-2012-0088>
12. Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 33(1):1–22. <https://www.jstatsoft.org/v33/i01/>
13. Haberman S, Renshaw AE (1996) Generalized linear models and actuarial science. *J R Stat Soc Ser D (The Statistician)* 45(4):407–436. <https://doi.org/10.2307/2988543>
14. Hoerl AE, Kannard RW, Baldwin KF (1975) Ridge regression: some simulations. *Commun Stat Theory Methods* 4(2):105–123. <https://doi.org/10.1080/03610927508827232>
15. James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning. Springer, Berlin
16. Joseph AW (1952) The Whittaker–Henderson method of graduation. *J Inst Actuar* 78(1):99–114. <https://doi.org/10.1017/S0020268100052495>
17. Kiermayer M (2021) Modeling surrender risk in life insurance: theoretical and experimental insight. arXiv preprint [arXiv:2101.11590](https://arxiv.org/abs/2101.11590)
18. Kiesenbauer D (2012) Main determinants of lapse in the German life insurance industry. *North Am Actuar J* 16(1):52–73. <https://doi.org/10.1080/10920277.2012.10590632>
19. Kim SJ, Koh K, Boyd S, Gorinevsky D (2009)  $\ell_1$  trend filtering. *SIAM Rev* 51(2):339–360. <https://doi.org/10.1137/070690274>
20. LeDell E, Gill N, Aiello S, Fu A, Candel A, Click C, Kraljevic T, Nykodym T, Abouyoun P, Kurka M, Malohlava M (2022) h2o: R interface for the ‘H2O’ scalable machine learning platform. R package version 3.36.0.4. <https://CRAN.R-project.org/package=h2o>
21. Loisel S, Piette P, Tsai CHJ (2021) Applying economic measures to lapse risk management with machine learning approaches. *ASTIN Bull J IAA* 51(3):839–871. <https://doi.org/10.1017/asb.2021.10>
22. McCullagh P, Nelder JA (1989) Generalized linear models, 2nd edn. Chapman & Hall/CRC, Boca Raton
23. Milhaud X, Dutang C (2018) Lapse tables for lapse risk management in insurance: a competing risk approach. *Eur Actuar J* 8(1):97–126. <https://doi.org/10.1007/s13385-018-0165-7>
24. Milhaud X, Loisel S, Maume-Deschamps V (2011) 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
25. Nelder JA, Wedderburn RW (1972) Generalized linear models. *J R Stat Soc Ser A (Gen)* 135(3):370–384. <https://doi.org/10.2307/2344614>
26. R Core Team (2022) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>
27. Reynkens T, Devriendt S, Antonio K (2021) smurf: sparse multi-type regularized feature modeling. R package version 1.1.2. <https://CRAN.R-project.org/package=smurf>
28. Schweizerische Aktuarvereinigung (2018) Richtlinie der Schweizerischen Aktuarvereinigung zur Bestimmung ausreichender technischer Rückstellungen Leben gemäss FINMA

- Rundschreiben 2008/43 “Rückstellungen Lebensversicherung”. [https://www.actuaries.ch/de/downloads/aid!b4ae4834-66cd-464b-bd27-1497194efc96/id!39/Richtlinie%20%C3%9Cberpr%C3%BCfung%20technische%20R%C3%BCckstellungen%20Leben\\_Version%202018.pdf](https://www.actuaries.ch/de/downloads/aid!b4ae4834-66cd-464b-bd27-1497194efc96/id!39/Richtlinie%20%C3%9Cberpr%C3%BCfung%20technische%20R%C3%BCckstellungen%20Leben_Version%202018.pdf)
29. Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B (Methodol)* 58(1):267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
  30. Tibshirani R (2011) Regression shrinkage and selection via the lasso: a retrospective. *J R Stat Soc Ser B (Methodol)* 73(3):273–282. <https://doi.org/10.1111/j.1467-9868.2011.00771.x>
  31. Tibshirani RJ (2014) Adaptive piecewise polynomial estimation via trend filtering. *Ann Stat* 42(1):285–323. <https://doi.org/10.1214/13-AOS1189>
  32. Tibshirani RJ, Taylor J (2011) The solution path of the generalized lasso. *Ann Stat* 39(3):1335–1371. <https://doi.org/10.1214/11-AOS878>
  33. Tibshirani R, Saunders M, Rosset S, Zhu J, Knight K (2005) Sparsity and smoothness via the fused lasso. *J R Stat Soc Ser B (Stat Methodol)* 67(1):91–108. <https://doi.org/10.1111/j.1467-9868.2005.00490.x>
  34. Wuthrich MV (2019) From generalized linear models to neural networks, and back. SSRN Manuscript ID 3822407. <https://doi.org/10.2139/ssrn.3491790>
  35. Xong LJ, Kang HM (2019) A comparison of classification models for life insurance lapse risk. *Int J Recent Technol Eng (IJRTE)* 7(5S):245–250
  36. Yu L, Cheng J, Lin T (2019) Life insurance lapse behaviour: evidence from China. *Geneva Pap Risk Insur Issues Pract* 44(4):653–678. <https://doi.org/10.1057/s41288-018-0104-5>
  37. Yuan M, Lin Y (2007) Model selection and estimation in the Gaussian graphical model. *Biometrika* 94(1):19–35. <https://doi.org/10.1093/biomet/asm018>
  38. Zou H (2006) The adaptive lasso and its oracle properties. *J Am Stat Assoc* 101(476):1418–1429. <https://doi.org/10.1198/016214506000000735>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.