CORVINUS
UNIVERSITY of
BUDAPEST

Doctoral School of Economics, Business and Informatics

THESIS BOOK

László Szepesváry

Informatics challenges in modern actuarial modelling:

Application of quantitative methods and computer algorithms in cash flow modelling and pricing

Ph.D. dissertation

Supervisor:

Erzsébet Kovács CSc

professor

Budapest, 2022

Department of Operations Research and Actuarial Sciences

THESIS BOOK

László Szepesváry

Informatics challenges in modern actuarial modelling:

Application of quantitative methods and computer algorithms in cash flow modelling and pricing

Ph.D. dissertation

Supervisor:

Erzsébet Kovács CSc

professor

Budapest, 2022

Contents

1. Research background and motivation for the topic

The concepts of risk and insurance are well-known from our daily lives. As described by Banyár (based on various encyclopaedias and textbooks), „risk is the danger from an action or undertaking, the possibility of material loss or damage", or „risk is the tendency that the actual outcome of a process under investigation is differing from the expected outcomes" (Banyár, 2016, pp. 60). There are several ways of managing risk, such as avoiding risk, preventing damage, building up a financial reserve or covering the risk with an insurance contract (Banyár, 2016). Due to the law of large numbers, the total risk underwritten by the insurer (the sum of many relatively small risks) is much more predictable than the individual risk, the insurer's risk is likely to be close to its expected value (Ohlsson & Johansson, 2010). Insurance can therefore be defined as "a cooperative strategy of risk management, done through the organisation of a risk pool" (Banyár, 2016 pp. 66). The economic risk is transferred from the policyholder to the insurance company by the insurance contract (Ohlsson & Johansson, 2010).

Insurances can be categorised according to several aspects (for instance, see (Banyár, 2016)). In the context of this thesis, we distinguish between life and non-life insurance. We refer to life insurance if it includes a coverage that depends on the insured's life (i.e. the occurrence or non-occurrence of death). Life insurance policies are often long-term contracts, for many years. Non-life insurance policies do not contain life coverage. Non-life insurances include, among others property and liability insurances, but also accident and health insurances. Non-life insurance policies are typically for a short period of time, typically one year (which does not mean that the insurance coverage cannot automatically be extended once it has expired).

Insurance is a very complex financial product, which requires special techniques for pricing and for further financial management (e.g. calculation of reserves). The discipline dealing with this is actuarial science, or insurance mathematics.

Over the last decades, the financial products linked to the underwritten risk, the insurance constructions for both life and non-life insurance, have undergone an essential transformation. The assessment of risks has been based on more complex models, leading to the genesis of modern insurance products. The development of the methodological and IT background was also essential for the transformation of the risk approach, which led to the birth of modern actuarial science. The traditional actuarial methods have nowadays been extended by a number

of novel tasks, which on one side require various statistical and modelling methodology, and on the other side, due to the huge data sets and large computational demanding algorithms a similar important element for the implementation is the appropriate IT background. These methods include the pricing of complex risk structures, but also the modelling of uncertain future cash flows associated with the underwritten risk. Latter is the basis of modern calculation of reserves. Several regulatory regimes have been established over the past decade which are closely linked to the modern approach, such as Solvency II, the framework for solvency capital requirement based on real nature of risks, or IFRS 17, the complex, modern standard for insurance contracts.

The thesis aims to present the processes outlined in both life and non-life insurance which have led to the new approach and to produce new results in some areas of modern actuarial science. The thesis presents the description of the methodology and its computational implementation, the dissertation deals with the synergy of these two fields in modern insurance mathematics.

The thesis examines the following research topics in life and non-life insurance. In the case of life insurance, the primary topic examined is modern cash flow modelling, its applications and certain methodological aspects (e.g. modelling techniques for investment yields, expenses, mortality, policyholder behaviour regarding life insurances), with a focus on their IT implementation as well. Monte Carlo simulations, time series analysis and other advanced statistical techniques, detailed sensitivity analysis are also applied. For non-life insurance, the relevant challenges of modern premium calculation are examined, different statistical and machine learning algorithms (generalized linear models, decision trees and random forests, neural networks) are applied to model insurance claims.

2. Used methods

With the spread of computers and specialised softwares, and the associated expansion of the statistical methodology, the actuarial modelling repository also has been extended in recent decades. The use of stochastic simulations (see for example (Bølviken, 2014)), machine learning techniques (see for example (Frees, Meyers, & Derrig, 2016)) and many other advanced statistical methods have become part of actuarial models. Even more, separate research fields have been developed within statistics as part of actuarial science, see for example mortality forecasting (Vékás, 2019). The development of the actuarial sciences has not only been motivated by researchers' motivation to knowledge, but also by the new challenges of the sector, the changing environment and increasing competition.

The aim of this thesis is not to review all the modern actuarial methods known today, as this would cover a vast amount of knowledge. In general, the thesis presents research hypotheses regarding relevant actuarial problems and elaborate their methodological and informatics solution.

2.1. Applications and methods used in life insurance modelling

In the subject of life insurance, the area of cash flow modelling is the central topic of the thesis. This field includes all applications based on the forecasting of future cash flows directly or indirectly related to a life insurance contract or group of contracts.

The thesis describes the following applications of cash flow modelling in more detail:

- **Profit test.** The profit test is based on the net present value technique, a method of company and business valuation known from finance (Banyár, 2016). More specifically, the profit test aims to calibrate the appropriate level of premiums and profitability based on the estimated future cash flows associated with a new contract, and therefore profit test is an important tool in modern premium calculation.
- **Solvency II**. The EU-wide regulation, which came into force on 1st January 2016, provides a framework for capital requirements for insurers based on the real nature of risks. The related solvency capital requirement (SCR) must be set at a level that ensures that the insurer can meet its financial obligations with a probability of at least 99.5% over a one-year period (2009/138/EC). This definition leads to the concept of *Value at Risk* (VaR) (see for example (Csóka, 2003)). The calculation of the capital requirement can be done using the so-called Standard formula (a set of rules that gives formula for

the calculation of the capital requirement for the different risk modules and a rule how to aggregate them) or using an internal model (calibrating the capital requirement to the 99.5% confidence level based on the distribution of risks and their interdependencies). The solvency capital requirement must be covered by the eligible part of the own funds. The own funds are the sum of the basic own funds and the ancillary own funds (2009/138/EC, Article 87). Basic own funds are the surplus of assets over liabilities plus subordinated loan capital (2009/138/EC, Article 88). Assets are valuated at market value and technical provisions included in liabilities are valued as follows. The technical provisions are the sum of the so-called best estimate and the risk margin (Directive 2009/138/EC, Article 87) The best estimate is the probability-weighted average of future cash flows related to insurance contracts, taking into account the time value of money. The best estimate should be calculated on the basis of current and reliable information and realistic assumptions, using appropriate, suitable and relevant actuarial and statistical methods. The best estimate requires the specification of an actuarial cash flow model.

- **IFRS 17**. IFRS 17, the new accounting standard for insurance contracts came into force from 1st January 2022. The new framework builds on a strong actuarial basis, with modelling of future cash flows as a central pillar. The standard defines the so called *fulfilment cash flow* which is the projected future cash flow adjusted with the time value of money (discount rate) and with the risk adjustment for non-financial risk (Risk adjustment or RA). The so-called CSM (contractual service margin) represents the expected unrealised profit arising from the future insurance services (IFRS 17, Article 38). The IFRS 17 calculations include an implicit profit test, known as the onerous test, which is necessary to group contracts. An important difference stands out between onerous and non-onerous contract groups. According to (IFRS 17, Article 47), if the amount calculated on the basis of the contractual cash flows at initial recognition is an expense to the insurer in total, the expected loss is recognised immediately in the P&L and the CSM margin is 0. The former loss is referred to as a loss component (LC) (IFRS 17, Article 49). An asymmetric situation is caused by the fact that the expected loss is recognised immediately in the profit and loss account for onerous contract groups, while for profitable contract groups the profit is only realised in the longer term, in proportion to the insurance coverage run-off.

The thesis identifies the most important variables of the cash flow model:

- Premiums paid by policyholders,

- Insurance service payments to customers,
- Costs associated with insurance obligations,
- Cash flows between insurers and agents (commissions),
- Changes in benefit reserves and deferred acquisition costs,
- Investment income and expenses,
- Reinsurance-related cash flows,
- Tax obligations,
- Other items.

The thesis also discusses the economic and non-economic factors that may influence the timing, quantity and probability of cash flows, which are the most important for the forthcoming hypotheses:

- Mortality,
- Lapses and other policyholder options,
- Future investment returns (returns of the insurer and participation of policyholders which increases their service),
- Future costs and inflation,
- Interrelationship between the variables.

Research hypotheses related to the life insurance cash flow modelling areas are presented in the section of life insurance in the thesis, and in many cases are proven by analysing empirical data. The main methodological techniques used in the case studies and in the proof of hypotheses are the following.

- **Using Monte Carlo simulations.** Determining the distribution of the present value or even the expected present value of cash flows from a multivariate interdependent system using explicit mathematical formulas can be a very difficult task. In such cases, it is usual to apply Monte Carlo simulations to approximate the analysed distribution. According to Glivenko's theorem known from statistics, if the sample size goes to infinity, then the empirical distribution function converges uniformly to the true distribution function almost surely (Tucker, 1959). The idea of Monte Carlo simulations is exactly the same, using random number generation to produce a sufficiently large number of realizations of the distribution under examination, and then using the simulated sample to estimate either the distribution or some related variable, such as the expected value (see for example (Bølviken, 2014)). In this thesis Monte Carlo

simulation technique is applied to the modelling of investment yields and the valuation of technical interest rate guarantees.

- **Use of time series analysis methods, Granger causality test.** For multivariate time series the fitting of vector autoregressive (VAR) models is used in the thesis and the examination of causality is done with the concept of Granger causality (see for example (Kirchgässner, Wolters, & Hassler, 2013)).

  For two endogenous variables X_t and Y_t time series, the VAR model is described in equation (1) for a maximum delay k (α, β coefficients are constants, ε_1, ε_2 are residual variables, assumed to be white noise, t is the time parameter).

$$\begin{aligned} X_t &= \alpha_0 + \alpha_{1,1}X_{t-1} + \cdots + \alpha_{1,k}X_{t-k} + \alpha_{2,1}Y_{t-1} + \cdots + \alpha_{2,k}Y_{t-k} + \varepsilon_{1,t} \\ Y_t &= \beta_0 + \beta_{1,1}X_{t-1} + \cdots + \beta_{1,k}X_{t-k} + \beta_{2,1}Y_{t-1} + \cdots + \beta_{2,k}Y_{t-k} + \varepsilon_{2,t} \end{aligned} \quad (1)$$

  Granger causality is defined as the following: Y_t is Granger cause of X_t if the delayed variables of the time series Y_t in the equation written for X_t have a significant effect on the value of X_t (the coefficients are non-zero), so that the past of Y_t has explanatory power for the present of X_t. The definition also does not exclude the possibility of looking at the Granger causality of a variable on itself.

  The Granger causality technique requires that the time series X_t and Y_t are stationary (their expected value and variance should be constant, and their autocovariance function should depend only on the distance between observations, be constant over time). Stationarity can be tested using the augmented Dickey-Fuller test, the null hypothesis is that the time series is non-stationary. To fit the VAR model, the ordinary least squares method can be used to estimate the coefficients. Finally, the Granger causality test can be performed using an F-test for the null hypothesis that for example, all the lagged values of Y_t in equation (1) have zero coefficients in the estimate of X_t. If the null hypothesis of the test is accepted, it means that Y_t is not the Granger cause of X_t.

  The concept of Granger causality is used in the thesis for hypotheses on insurance lapse rates.

- **Using multivariate statistical methods** (k-means clustering, survival models).

  The idea of k -means clustering is to form k cluster centres in the space formed by the selected variables, and then to classify each observation into k classes based on a defined distance to the nearest cluster centre. Based on the coordinates of the cluster centres, conclusions can be drawn about the properties of the cluster. For more details, see for example (Kovács, 2011). In the thesis the clustering is applied in the analysis of insurance lapses to classify contract groups.

Also in the context of modelling lapses, the so-called survival models are presented in the thesis. Let T denote the random variable that measures the number of months that elapse between the inception of a contract and its cancellation. The function defined by the formula $G(t) = P(T \geq t)$ is called the survival function, which gives the probability for each time t that the contract has been alive for at least t months. The two best-known survival models are the Kaplan-Meier estimator and Cox regression. The former estimates the survival function without explanatory variables, the latter with explanatory variables (for more details, see for example (Vékás, 2011)).

- **Sensitivity analysis techniques and dynamic actuarial modelling.** The use of sensitivity analysis techniques is a popular part of modern actuarial modelling and profit testing. In addition, dynamic modelling techniques are used in many analyses of the thesis (see next section on hypotheses for more details).

In addition to the techniques mentioned above, the thesis also systematises possible further elements of the methodological repertoire, with relevant literature review from both the theoretical and the applied perspective.

2.2. Applications and methods used in non-life insurance modelling

In the context of non-life insurance, the thesis deals with the issue of data analysis based modern premium calculation (pricing). The main objective is to investigate whether is it possible to use certain artificial intelligence-based machine learning algorithms to produce a model which is more fitting to the real nature of risks than the already widely used multivariate statistical methods, such as the generalized linear model. The thesis describes the process in non-life insurance that has led to the development of complexity in constructs and the emergence of differentiated premium rates according to risk structure.

The essence of modern pricing is to model the adjusted dependent variable Y (the so-called key ratio, which could be for example the claim frequency or the claim severity) by the rating factors $X = (X_1, X_2, \dots, X_p)$ (explanatory variables presenting different properties of the observation units). In other words, the aim of the statistical analysis is to find an f function that explains the $Y = f(X)$ risk model well. The pricing factors can be different features of the policyholder (e.g. age, residence, bonus-malus rating), different properties of the insured object (e.g. performance, make of car etc.).

In modern non-life insurance premium calculation, the following formula is very often used to determine the net premium (see for example (Ohlsson & Johansson, 2010), (Frees, Meyers & Derrig, 2016)):

$$\text{Pure premium} = \text{Claim frequency} \cdot \text{Claim severity}\,. \tag{2}$$

The formula has to be understood as two statistical models are constructed, one for the claim frequency and the other for the claim severity, and then the product of the expected frequency and severity per observation unit from the models gives the value of the net premium.

The following techniques are used in this thesis to model the function $Y = f(X)$. For given n observations for statistical model building, the values of the variables X and Y are denoted by y_i $(i = 1, \dots, n)$ and $x_{i,j}$ $(i = 1, \dots, n; j = 1, \dots, p)$.

- **Generalized linear model (GLM).** Its basic equation is:

$$y_i = \mu_i + \varepsilon_i, \qquad E(y_i) = \mu_i, \qquad g(\mu_i) = \sum_{j=1}^{p} b_j x_{i,j}, \qquad (i = 1, \dots, n). \tag{3}$$

  The function g is called a link function, g is assumed to be differentiable and monotone, and the distribution of y_i is assumed to belong to the exponential family of distributions. The observations y_i $(i = 1, \dots, n)$ are assumed to be independent. For more details see (Gray & Kovács, 2001).

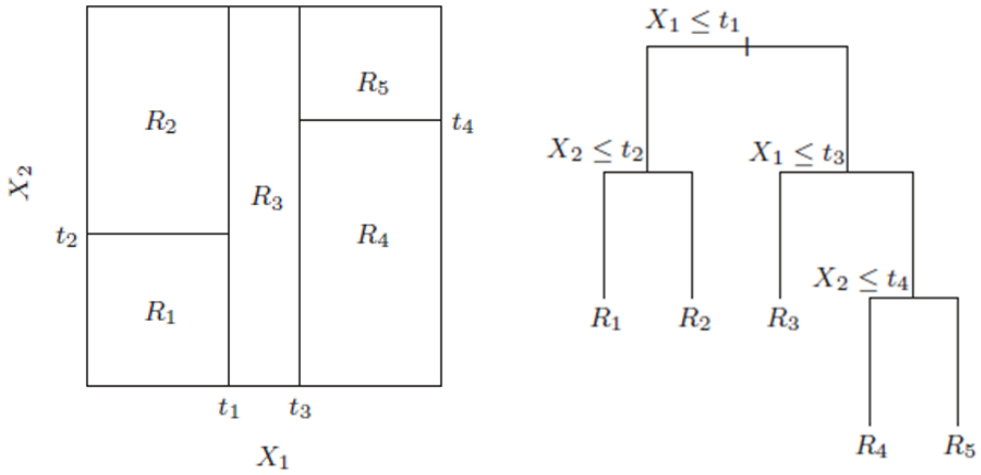
  The assumption on the link function is very important for GLM models. Ohlsson and Johansson point out that the logarithmic link function (log-link, $g(\mu_i) = \log(\mu_i)$) is often used for multiplicative model assumptions, the identity link function $g(\mu_i) = \mu_i$ is suitable for linear models, and the logit link function $g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right)$ should be applied for proportions or probabilities (Ohlsson & Johansson, 2010).

  In the GLM models, the parameters b_j $(j = 1, \dots, p)$ are estimated using the maximum likelihood principle, based on the density function of the assumed exponential distribution family distribution. The maximization of the log-likelihood is typically performed by numerical methods, such as the Newton-Raphson method (see, for example (Ohlsson & Johansson, 2010)).

  After the maximum likelihood estimation, the significance of the coefficients b_j of each explanatory variable can be tested for: the Wald test statistic can be used to test the null hypothesis $b_j = 0$. Nowadays, model fitting and related calculations are usually performed using computer software packages (e.g. R, SPSS) when applied to real data. A commonly used indicator for comparing different models is the deviance.

- **Decision trees and random forests.** Decision trees are popular and simply implementable machine learning algorithms that can be used to map non-linear structures. The mathematical model of classification and regression trees (CART for short) is reviewed in the thesis, based on the methodological book (Hastie, Tibshirani & Friedman, 2009).

  For regression trees, the dependent variable Y is a continuous variable, and for classification trees, it is a categorical variable. In both cases, the goal is to partition the space of explanatory variables $X = (X_1, X_2, \dots, X_p)$ with a series of cuts like $X_i \leq s_{i,j}$ and $X_i > s_{i,j}$ ($s_{i,j}$ denotes the j-th cut for the i-th explanatory variable). Then within the resulting partitions the values of the dependent variable are estimated as constants such that the difference between the real $(y_i\ (i = 1, \dots, n))$ and the estimated $(\widehat{y_i}\ (i = 1, \dots, n))$ values function should be as small as possible, based on a target function. Figure 1 illustrates the schematic structure of the decision tree and the partitioned space for a two-dimensional example.

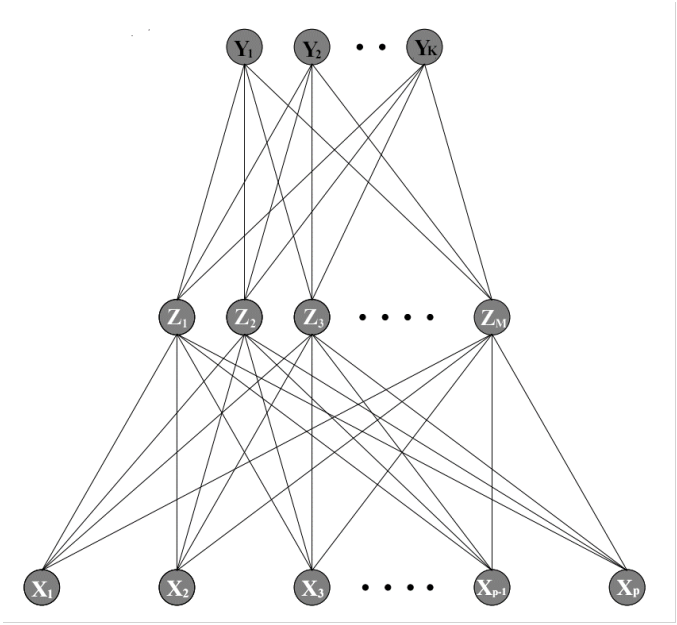


1. Figure: The schematic structure of the decision tree and the partitioned space.
Source: (Hastie, Tibshirani & Friedman, 2009)

Random forests are a so-called bootstrap aggregation (or bagging for short) procedure based on the method of classification and regression trees. It consists of taking a large number of random subsamples from the original training sample, training the algorithm on these subsamples, and then averaging the trained models to produce the final model. With the method the variance of the estimate can be reduced (Hastie, Tibshirani & Friedman, 2009).

- **Neural networks.** The idea behind the development of so-called artificial neural networks, or neural networks for short, stems from the consideration that the human brain learns and behaves in a completely different way to a conventional computer. Just as the neurons in the human brain form a complex network, the model of artificial neural networks builds a similar structure. Neural networks also develop through a learning process, and a further similarity is that the connections between neurons store the knowledge of the system (Hajek, 2005).

  The most commonly used neural network model nowadays is the multi-layer perceptron network (MLP). Their popular learning method is the back-propagation algorithm, which is based on gradient descent.



2. Figure: Schematic diagram of a one hidden layer neural network.
Source: (Hastie, Tibshirani & Friedman, 2009)

Figure 2 shows the schematic graph of a neural network with one hidden layer. The bottom layer is the input layer, $X = (X_1, X_2, \dots, X_p)$ input (explanatory) variables. The specific values for the n observations are denoted by $x_i = (x_{i,1}, x_{i,2}, \dots, x_{i,p}), i = 1, \dots, n$.

The middle layer is the hidden layer, which consists of a number M of latent derived variables. The relationship between the input and hidden layer variables can be described by the following formula:

$$Z_m = \sigma(\alpha_{0m} + \alpha'_m X), \qquad m = 1, \dots, M, \tag{4}$$

where α values are weight parameters and $\sigma(v)$ is the so-called activation function.

The top layer is the output layer $Y = (Y_1, Y_2, \dots, Y_K)$. In case of regression modelling (continuous output variable) typically $K = 1$ is used, and for classification modelling typically the number of output variables is equal to the number of groups, and Y_k $(k = 1, \dots, K)$ is the indicator for the k-th group with values $0 - 1$ (Hastie, Tibshirani és Friedman, 2009).

The relationship between the output layer and the hidden layer can be described by the following formulas:

$$\begin{aligned} T_k &= \beta_{0k} + \beta_k' Z, \quad k = 1, \dots, K, \\ \mathrm{Y}_k &= f_k(X) = g_k(T), \quad k = 1, \dots, K, \end{aligned} \tag{5}$$

where $Z = (Z_1, Z_2, \dots, X_M)$, $T = (T_1, T_2, \dots, T_K)$, β are weight parameters and $g_k(T)$ is the transformation function for the T output.

3. Results of the thesis

I briefly present the research hypotheses and results. Without exception, the examined hypotheses are based on one of my own or co-authored publications, and the research is presented in the thesis by citing the studies.

3.1. Yield modelling approaches in cash flow projection and valuation of the yield guarantee

This sub-chapter examines some modelling approaches for future investment yields related to life insurance cash flow projections and analyses the value of the yield guarantee in traditional life insurance products. In addition to a general description of the topic (related accounting rules, assumptions on future interest rates, interest rate models, valuation methods for the yield guarantee), the following hypothesis is stated

1. **Hypothesis:** Stochastic methods make it possible to quantify the value of the guarantee provided by the technical interest rate, which can have a significant impact on the value of liabilities, especially in a low yield environment.

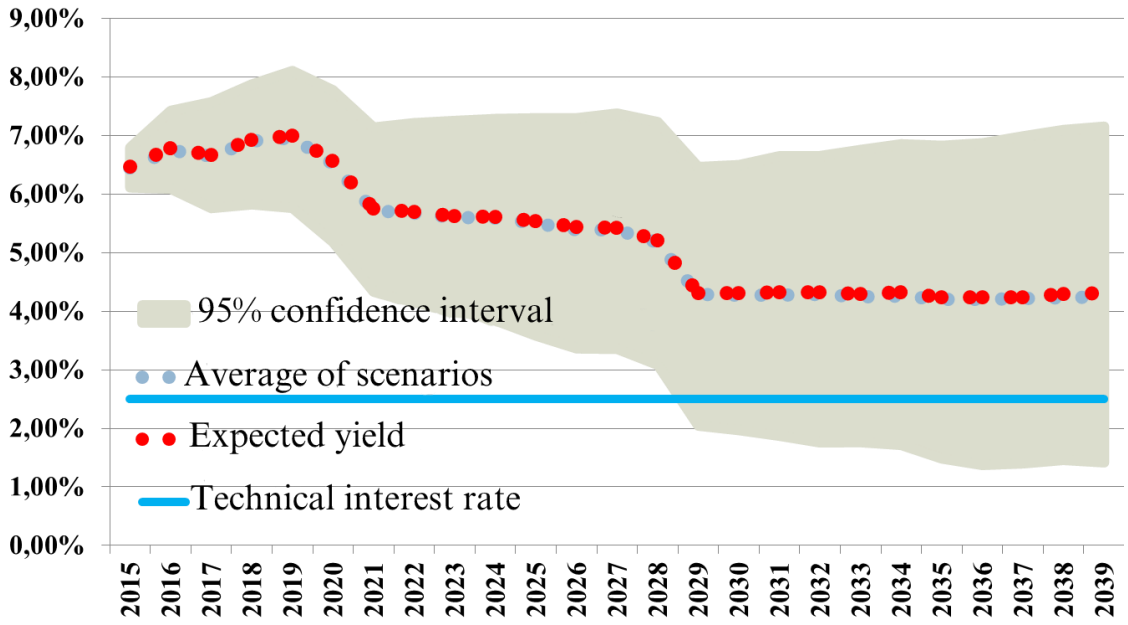
To prove the hypothesis, a case study is presented based on own research publication (Szepesváry, 2015). The hypothesis examines the value of the interest rate guarantee in case of traditional life insurance policies. The numerical tests are carried out in the Solvency 2 framework. A characteristic feature of traditional life insurance is the fixed technical interest rate (i), which represents a guaranteed return on the investment of the premium reserve. If the investment yield exceeds the technical interest rate, the policyholder will participate in the surplus over the technical interest rate. The study points out that the guarantee provided by the technical interest rate creates an asymmetric situation between the insurer and the client: if the return on the premium reserve is below the technical interest rate, the insurer suffers the loss, but if it exceeds it, the two parties share the surplus (Szepesváry, 2015). According to Hungarian accounting rules, assets behind the premium reserve of traditional life insurance policies must be valued according to the book value method (192/2000. (XI. 24.) Government Decree, 2000). It is possible to valuate the book values and yields on a security basis, and for the future it should be projected taking into account future market expectations. Furthermore, as it is necessary to adequately cover liabilities (reserves) with assets also in the future, this also requires a forecast of the matching of assets and liabilities (so-called *Asset-liability matching*, abbreviated by ALM).

In this study, three portfolios of policies with assets for their premium reserves were examined and modelled based on different assumptions (Szepesváry, 2015). The main differences between the portfolio samples are the maturity structures of assets and liabilities, the technical interest rate and the currency. A 4-step process was used:

1. **Dynamic ALM model (deterministic yield model).** The objective of the method is to determine the expected future accounting yields. The method first compares the expected run-off of the assets and premium reserves existing at the time of valuation, taking into account the expected cash flows on the asset and liability sides. On this basis, it makes a dynamic adjustment for the future: it sells or buys bonds according to the assumed future yield curve, based on changes in reserves and cash flows of the portfolio. The aim is to fully cover the expected reserves with assets at future dates and to provide liquidity in line with the expected future cash flows on the asset and liability sides. The output of the model is the expected future accounting yields of the dynamic asset portfolio calculated as above.
2. **Deviation variable based on empirical data, Monte Carlo simulation for future returns (stochastic yield model).** It is the point where the deterministic and stochastic models separate and the random variability of future returns emerges. Based on empirical insurance asset and return data, the distribution of the actual yield reported at the end of the month was analysed in relation to the yield that could be predicted from the asset portfolio data at the beginning of the month. This monthly variability is caused by random changes. From the estimated distribution as a kind of monthly random deviation for month i, independent monthly error terms (e_i) were simulated by Monte Carlo method. From this, we could define the cumulative deviation from the time of the evaluation to the end of month i:

$$E_i = \sum_{j=1}^{i} e_j. \tag{6}$$

For E_i a sample can be simulated, and by adding these values to the deterministic expected returns from the dynamic ALM model, stochastic yield scenarios can be generated (see for example Figure 3).



3. Figure Simulated annual yields for sample portfolio 1.
Source: (Szepesváry, 2015)

3. **Evaluation of the cash flow model (deterministic and stochastic return model).** The *best estimate* from the deterministic return model is calculated as the expected present value of the modelled cash flows according to the deterministic yield scenario. For the stochastic return model, the best estimate is the average of the expected present values calculated separately for the simulated scenarios.
4. **Analysis of the yield guarantee.** The following two indicators are introduced to quantify the value of the yield guarantee:
   a. „Simulation effect (S). The difference between the stochastic and deterministic best estimate. It quantifies the change in the expected present value of the liabilities when modelling random fluctuations in yields" (Szepesváry, 2015). With formula:

$$S = BE_S - BE_D = \frac{\sum_{s=1}^{N}\sum_{t=1}^{n} CF_t^{\,s} \cdot d_t}{N} - \sum_{t=1}^{n} CF_t \cdot d_t, \qquad (7)$$

   b. „Loss due to yield guarantee (H). The expected present value of the yield deficiency below the technical interest rate to the promised level based on stochastic return scenarios. This is the expected amount of returns (that the investment has not earned), which have to be credited to the policy due to the activation of the guarantee" (Szepesváry, 2015). With formula:

$$H_s = \sum_{t=1}^{n} (V_{t-1} + N_t) \cdot \max(i - r_t{}^s, 0) \cdot d_t$$

$$H = \frac{\sum_{s=1}^{N} H_s}{N} \tag{8}$$

In the formulas the CF variables are the cash flows of different scenarios, d_t is the discount curve, n is the number of periods, N is the number of stochastic scenarios, r_t is the vector of future accounting yields, $(V_{t-1} + N_t)$ is the amount invested in a given month (premium reserve).

Policy	Term (year)	Years left	Initial Sum Assured (in HUF)	Premium reserve (V)	Best estimate (BE)	H	H / |BE|	S / |BE|
1	10	3	1 000 000	840 512	889 844	0	0,00%	0,00%
2	10	5	1 000 000	503 327	533 030	0	0,00%	0,01%
3	15	10	1 500 000	457 020	438 935	0	0,00%	0,13%
4	25	20	2 500 000	404 088	239 699	1 821	0,76%	2,43%
5	30	25	3 000 000	391 658	193 557	4 223	2,18%	5,73%

1. Table: Results for different contracts in sample portfolio 1.
Source: (Szepesváry, 2015)

The value of the yield guarantee has been evaluated on different samples (see for example Table 1). The thesis also discusses the criteria for assessing the value of the yield guarantee as significant (material). The aspects of the changes in the yield environment are presented also. The investigations presented in the thesis in detail prove hypothesis 1.

3.2. Onerous testing in IFRS 17 and expense modelling approaches in cash flow projections

This subsection examines some modelling possibilities for future expenses related to life insurance cash flow projections and presents how this is linked to the onerous test under IFRS 17. In addition to a general description of the topic (types of costs in life insurance, classical expense assumptions, expense modelling in modern cash flow projections, cost allocation methods, role of inflation), the following hypothesis is stated:

2. **Hypothesis:** *onerous contracts* under IFRS 17 have a significant impact on financial income indicators at *initial recognition*, which can be optimised by calibrating the

model's assumptions on expenses or by making the cost structure of the premium calculation and the real cost structure consistent over time.

To prove the hypothesis, a case study based on own research publication is presented in the thesis. The study is based on an analysis of empirical insurance data (Szepesváry, 2019).

The following cost allocation methods are defined (Szepesváry, 2019).

- ***Per policy cost allocation***: the total projected cost of product is allocated to the future projected number of contracts using a unit cost method (i.e. each contract is allocated an equal cost).
- ***Per premium cost allocation***: the former costs are allocated to each contract in proportion to portfolio premium.
- ***Mixed cost allocation***: mixture of *Per policy cost allocation* and *Per premium cost allocation*. The so-called *per policy cost%* indicator shows what proportion of the total costs is considered to allocated per policy basis, and its complement is considered to allocated per premium.

Other important concepts for the application in the IFRS 17 framework are the so called *directly attributable* and *non-directly attributable* costs. According to (IFRS 17, B65 and B66 paragraphs) the latter category of costs are not part of fulfilment cash flows and should not be taken into account in the onerous testing. Directly attributable costs are defined using a percentage indicator as a proportion of total costs in the cited study (Szepesváry, 2019).

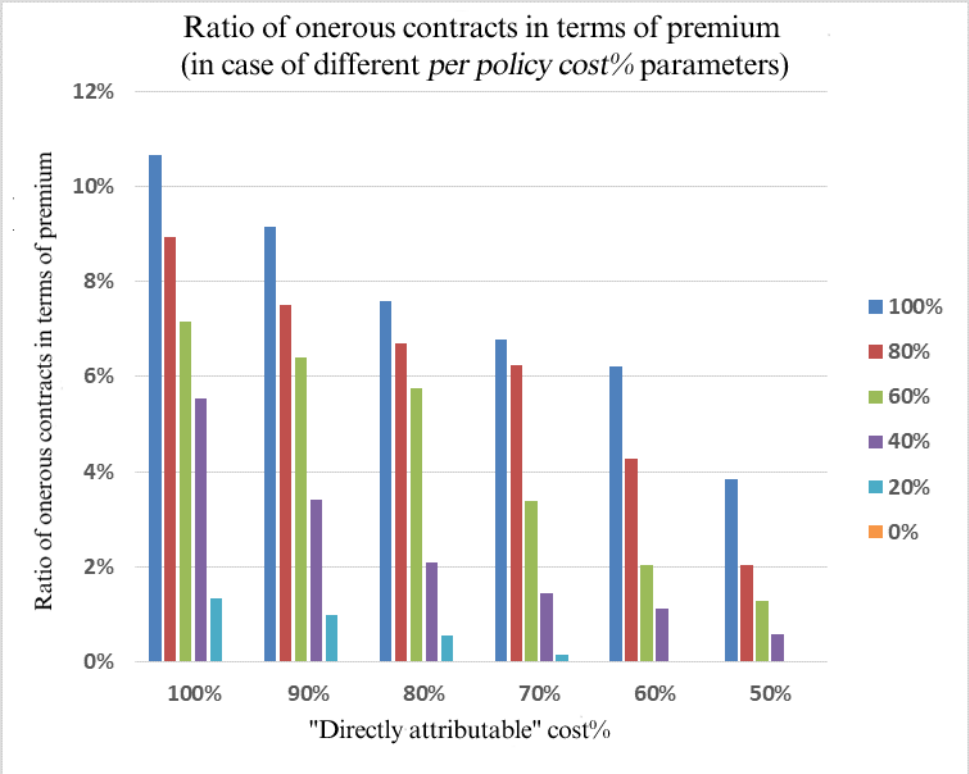
For a given single premium insurance product and a given new business sample, the paper and the thesis present the results of the contract-by-contract onerous test and some of its implications for the indicators that can be calculated at initial recognition under IFRS 17. The initial recognition profit test and the initial *contractual service margin* (CSM) calculation is analysed in details.

The thesis focuses on the separation of initially onerous and initially non-onerous sets, and does not examine the case of getting onerous with significant probability later on. In the corresponding cash flow model, expected future cash flows per contract are determined, which after discounting and adding the risk adjustment, become the measure of initial profitability. Based on the sign of the former quantity, the onerous grouping can be performed.

For the onerous profile, the following indicators are used in the paper and the thesis:

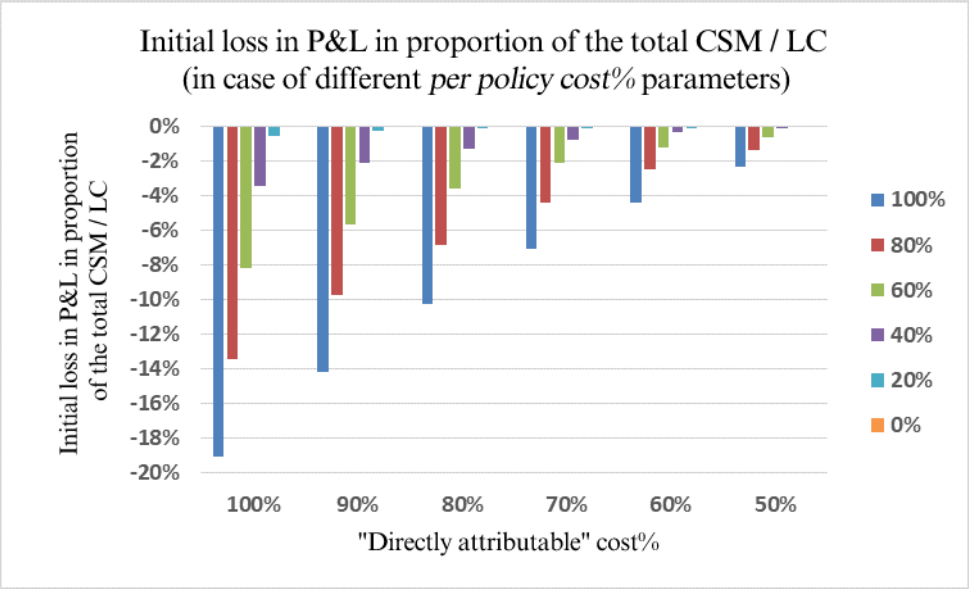
- **„CSM / LC**[1] **in proportion to premium**. The expected future profit/loss per premium, i.e. a measure of profitability for a given segment.
- **Ratio of onerous contracts.** What proportion of the segment is in the onerous contract group?
- **Ratio of onerous contracts in terms of premium.** What proportion of the segment is in the onerous contract group in terms of premium?
- **Initial loss in P&L in proportion to the premium.** How much loss should be recognised in the profit and loss account initially due to onerous contracts, in proportion to the premium?
- **Initial loss in P&L in proportion to the total CSM / LC.** How much loss should be recognised in the profit and loss account initially due to onerous contracts, in proportion to the total expected future profit/loss of contracts?" (Szepesváry, 2019)

These indicators are suitable measures of IFRS 17 onerous profile and are important for illustrating the results and proving the research hypothesis. Figures 4 and 5 show the values of some of the indicators in the sample, according to the sensitivity analysis of the cost allocation procedure parameters.



4. Figure: Ratio of onerous contracts in terms of premium in case of different cost allocation parameters
Source: (Szepesváry, 2019)

[1] Loss component.



5. Figure: Initial loss in P&L in case of different cost allocation parameters

Source: (Szepesváry, 2019)

Figures 4 and 5, as well as the results presented in details in this thesis, provide strong confirmation of the first part of the hypothesis that onerous contracts can have a significant impact on the financial income indicators at initial recognition under IFRS 17.

The sensitivity analysis on costs is the basis of the second part of the examined hypothesis, that the IFRS 17 onerous profile and the financial income indicators at initial recognition can be optimised by calibrating the model's assumptions on costs. The thesis concludes that an optimal onerous profile can be achieved if the cost structure of the pricing is fitting with the real cost structure. This requires a very complex interconnection between the different areas and processes of the insurer in the long run: the real cost pattern and its theoretical model must be consistent with the pricing principles to reach the optimal state.

Using the right cost allocation methodology means not only choosing the parameters, but also justifying them. "Determining this is obviously a company-specific task and may need to be supported by data and analysis for use in the cash flow model. In many cases, the proportion of directly attributable costs is more given for a company and cost situation, whereas in case of per policy or per premium or mixed cost allocation there is usually considerable discretion for the company for the allocation principles" (Szepesváry, 2019).

It is emphasised that "in current market practice, business considerations may cause some onerous contracts for certain products to be sold, which could have a strong impact on the financial indicators of the new accounting standard. Recognising this and taking the necessary management actions may be central when preparing for IFRS 17." (Szepesváry, 2019).

In the last part of the section, additional sensitivity tests are presented, as well as possible solutions to the IT difficulties of IFRS 17 profitability assessment.

3.3. Modelling lapses and policyholder behaviour options, analysing their relationship with economic and non-economic variables

The topic of this subchapter is to discuss some modelling options for future customer behaviour related to life insurance cash flow forecasting and to analyse its relationship with different economic and non-economic variables. In addition to a general description of the topic (policyholder behaviour options, mathematical modelling of lapses and survival models, key external and internal factors affecting lapses, modelling possibilities of dynamic policyholder behaviour), the following hypotheses are stated:

3. **Hypothesis:** In the case of traditional regular premium life insurances, relevant insurance data do not show that if the benchmark yields exceed the value of the technical interest rate, the lapse rates increase.
4. **Hypothesis:** For investment-focused single premium life insurance constructions a dependence on the external or internal interest rate environment can be detected from relevant insurance data. If other forms of investment offer higher returns, or if the level of interest available within a given contract decreases, then lapse rates may increase depending on the characteristics of the contract group.

The hypothesis is verified by a case study based on own research publication (Szepesváry, 2022). At the time of finalizing the thesis, the article was submitted, awaiting review status, and if successfully accepted, will be published in the journal *Financial and economic review* in June 2022. The study was based on an analysis of empirical insurance data.

The study examines traditional single and regular premium life insurance policies with a savings element. The time series of lapse rates of the different products are analysed in relation to external, economic and non-economic time series and events: changes in the yield and

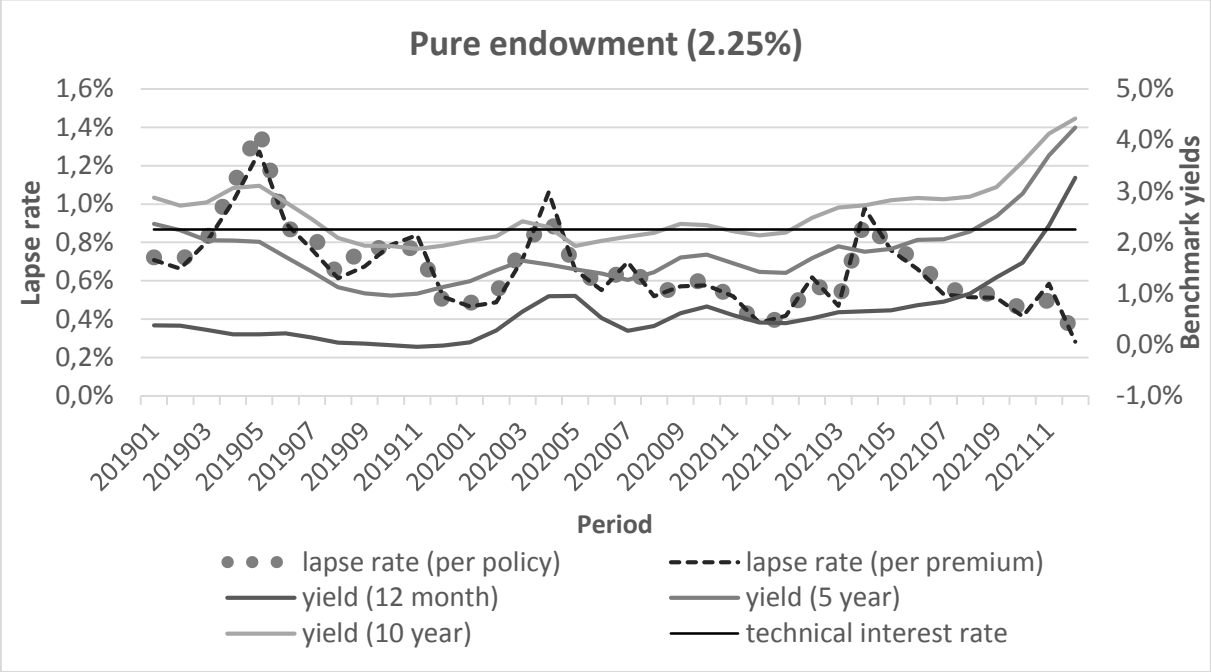
inflation environment and the impact of COVID-19 lockdowns on lapses are the main points of analysis.

The external interest rate environment is measured by the time series of the 1, 5 and 10-year benchmark yields[2] of the Government Debt Management Agency, inflation is measured by the seasonally adjusted core inflation indicator published by the HNB[3] and the impact of the COVID-19 lockdowns by the so-called stringency index indicator[4] (data for Hungary). The lapse rate was defined as the ratio of the volume of contracts cancelled (surrendered or payed up) in a given month to the average number of contracts were in force in that month (both per policy and per premium versions were calculated and displayed). An important question asked in the study is whether the rising yield and inflation environment in the second half of 2021 has had a detectable impact on lapse rates.

The results for one of the analysed current premium products are shown graphically in Figure 6. The graph shows the lapse rates for the period 2019 - 2021 monthly (scale on the left vertical axis), as well as the benchmark yields and the technical interest rate level for the product (scale on the right vertical axis). The percentage value displayed next to the name of the construction in the header of the graph shows the technical interest rate for the contract group. As shown in (Figure 6), the increase in benchmark yields in the second half of 2021 did not increase the lapse rates of the products, even when the level of the available interest rate exceeded the technical interest rate.

This was also analysed using time series analysis methods. The Granger causality test (whether the reference yield is the Granger cause of cancellation) was performed on the per policy and per premium lapse rates of the regular premium insurance types examined with the time series of the reference yields, pairwise. Based on the values of the F-test, it was statistically proven that the reference yield had no significant effect on lapse rates for the analysed products during the period under study, which proves Hypothesis 3.

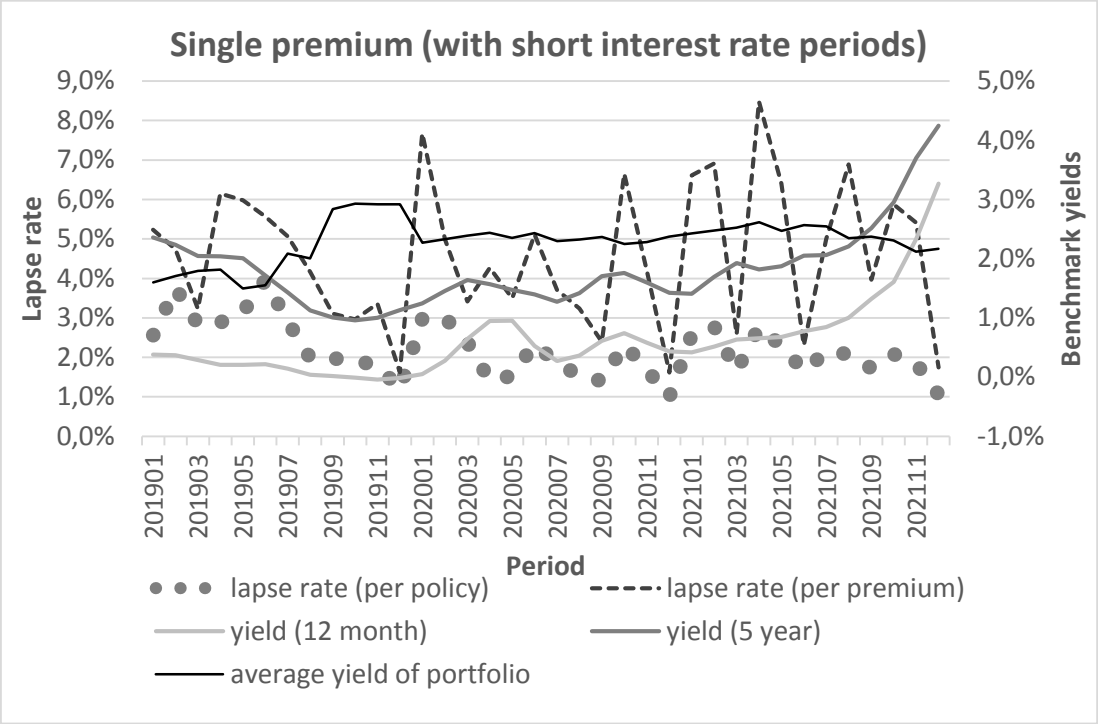
[2] https://www.akk.hu/statisztika/hozamok-indexek-forgalmi-adatok/referenciahozamok
[3] https://www.mnb.hu/statisztika/statisztikai-adatok-informaciok/adatok-idosorok/vi-arak
[4] https://ourworldindata.org/grapher/covid-stringency-index?tab=chart&region=Europe&country=~HUN



6. Figure: Lapse rates of regular premium insurance with saving element depending on yields
Source: (Szepesváry, 2022)

For single-premium insurance, a special construction was examined, where customers share in the returns according to short-term interest periods fixed in advance. This pre-fixed rate allows each client to know until the end of the month the interest the insurer is offering for the following month and to decide whether to keep or surrender their savings. A *decrease in yield* indicator is defined, which measures how much the level of yield credited to a given client would change in the following month. This indicator turns out to be an important proxy for the probability of lapse.

A graph similar to Figure 6 was used to analyse the cancellation rates of the single premium product (Figure 7). Here, the average interest rate over the short-term interest periods is plotted instead of the technical rate. The figure does not show a relationship between benchmark yields and lapse rates (this was later confirmed by the Granger causality definition also).



7. Figure: Lapse rates of single premium insurance in relation with yields

Source: (Szepesváry, 2022)

It is notable in the graph that the per premium cancellation rate is significantly higher and more volatile than the per policy lapse rate. This indicates that the client portfolio is not homogeneous, the size of the invested premium affects the cancellation rate, clients with a higher savings amount being more likely to call the cancellation option. To better understand the reasons for the volatile behaviour, a more detailed analysis was carried out with data broken down to the level of each client (Szepesváry, 2022).

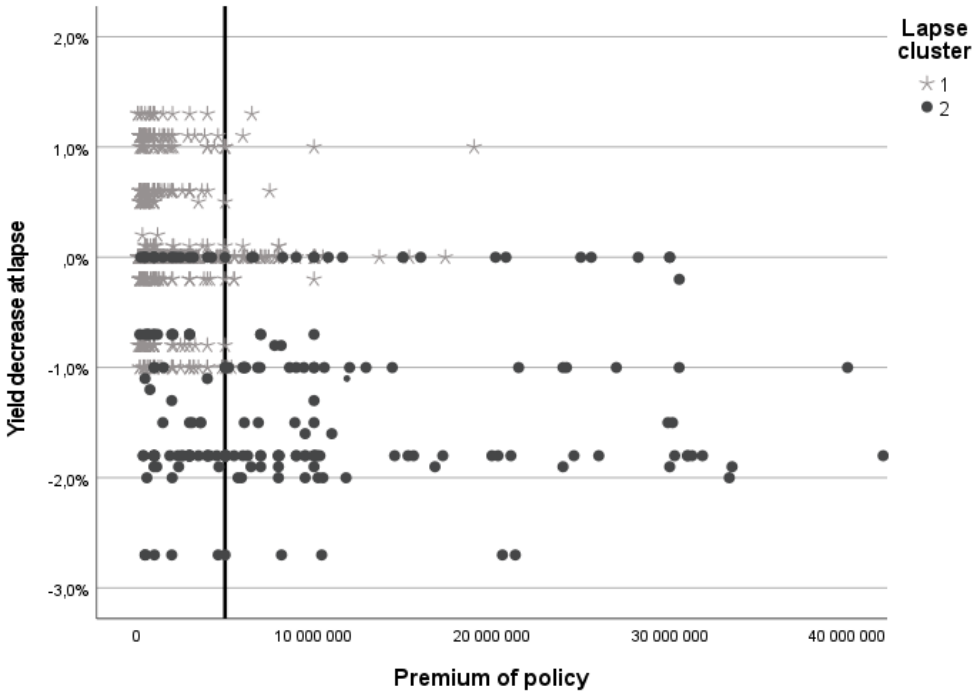
k-means clustering was used to classify the cancellation events and draw conclusions about the portfolio, in order to finally group the clients by contract parameters. Two groups are identified (lapse clusters):

- Customers with lower premiums, less sensitive to yield changes regarding lapse rate
- Customers with higher premiums, more sensitive to yield changes regarding lapse rate

Finally, a simpler clustering is defined (premium clustering, using the premium coordinates of the lapse cluster centres), which is no longer applicable only to cancelled contracts, but can achieve a good hit ratio for lapse cluster groups.

(Figure 8) shows, in a cross-plot of the size of the premium and the yield decrease indicator, how the lapse clusters are positioned relative to the premium clusters (the latter is the left and

right side of the vertical line, where the higher premium clients, typically with higher cancellation probabilities, are positioned on the right).



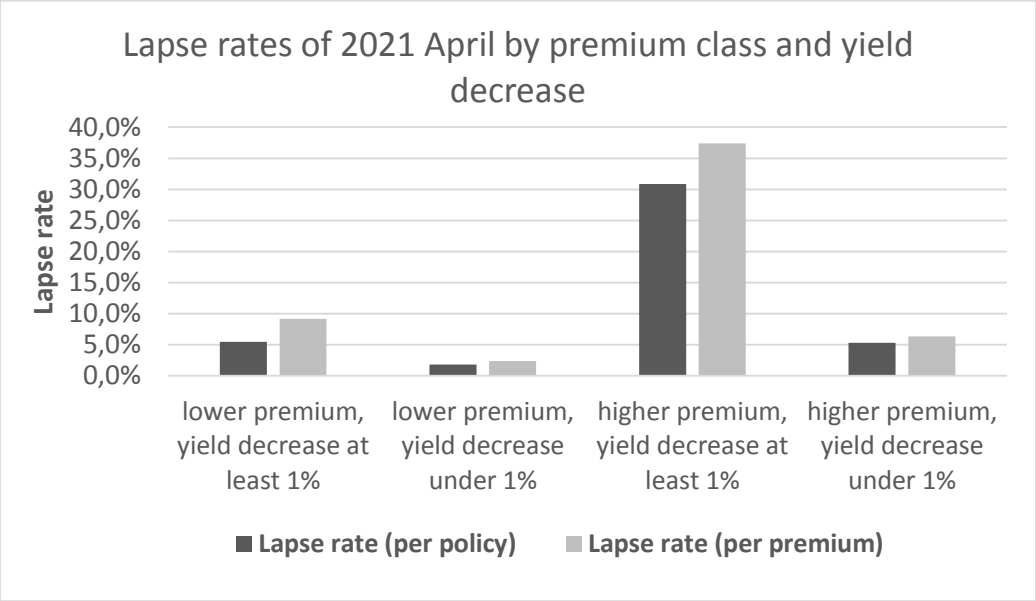
8. Figure: Clusters by lapses and by premium (in HUF) and dimension of the yield decrease indicator
Source: (Szepesváry, 2022)

We split the portfolio into two parts according to the cut-off resulting from the clustering by premium to further investigate the probability of cancellation. For several selected periods, we have analysed how the yield decrease indicator has developed for the contracts in the two groups, both for the full sample and for the contracts cancelled in a given month (for example, Figure 9 shows the data for month April of 2021).

In April 2021, the lapse rate was definitely influenced by the yield decrease indicator (Figure 9). In the lower premium group, the lapse rate was between 3 and 3.87 times (per piece and per premium lapse rates) for contracts that would have experienced a yield decrease of at least 1% (in absolute terms) from the following month, compared to contracts that did not experience such a yield decrease (or had a yield increase).

The same indicator represented a 5.8 to 5.9-fold increase in the lapse rate in the higher-premium group. This implies that the internal rate of return for the analysed product is a very strong indicator of cancellation. A 1% decrease in yields already encourages a significant proportion of clients to surrender. This is even more true for the higher premium classes (where there are proportionally more yield-sensitive customers), but even for the lower premium classes, significant cancellation surplus has occurred before the yield decline. In June 2019, the launch

of the so called MÁP+ government bond, which has very attractive features, is found to explain the increase in cancellations in the cited study (Szepesváry, 2022).



9. Figure: Lapse rates by premium class and yield decrease (April 2021)
Source: (Szepesváry, 2022)

The results presented in more detail in the thesis prove research hypothesis 4. In addition to examining the dependence on yields, the relationship of cancellation rates with the inflation index and the stringency index associated with COVID-19 closures is also analysed.

3.4. Modelling the probability of non-life insurance claims using different machine algorithms

In this section, I present the results of a study based on my own co-authored research (Burka, Kovács & Szepesváry 2021) in the thesis, which supports the following:

5. **Hypothesis:** It can be verified on empirical data that machine learning methods and their combinations can be used to create a better predictive tool than the generalized linear model in non-life insurance claims modelling, and that they can be formalised with certain approximations with losing some of the variance.

The results of this research paper are a joint intellectual creation with the co-authors, but the co-authors have contributed to use the results in my thesis. In the related research, I was mainly involved in data preparation, statistical evaluation of the results and actuarial applications, while the co-authors were responsible for programming the machine algorithms, model fitting and

variable selection. Accordingly, I focused on actuarial application and statistical evaluation in the thesis. The study is attached to the Thesis book with further details.

The study investigates the occurrence of a claim event on the motor third party liability (MTPL) insurance data of a Hungarian non-life insurer using different machine learning algorithms. The dependent variable in the analysis is a binary variable that indicates whether a claim event occurred in a given policy year. A statistical model is constructed using more than ten explanatory variables to estimate the probability of loss. For model building, the sample is divided into training, validation and test sets. In the study it is highlighted that for model fits used in non-life insurance pricing, most authors evaluate the performance of models using deviance, see for example (Yeo, 2011), (Kafková & Krivánková, 2014). The disadvantage of this approach is that the deviance is only suitable for comparison when the estimates are made in the same model framework.

Since the study uses several different statistical models, a different type of evaluation methodology is required for the comparison. One tool for evaluation is based on the cut-off value and confusion matrix known from the methodology of classification procedures. Plotting the ROC curves and calculating the area under curve (AUC) can be used to model comparison (see for example (Kovács, 2011).

Actual \ Model	0	1
0	*a*	*b*
1	*c*	*d*

2. Table: A schematic example of the confusion matrix.

Source: own editing

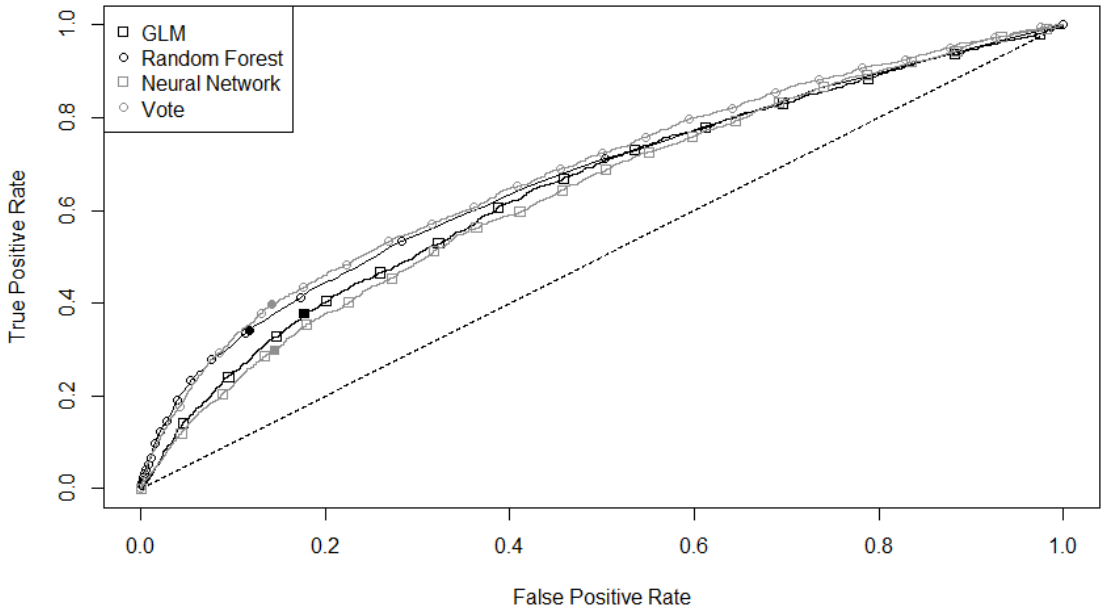
The other tool used for the evaluation in the presented study is also given by the logic of the confusion matrix, based on the following considerations (Burka, Kovács & Szepesváry, 2021):

- The misclassification among contracts with and without claim is not symmetric from a financial point of view, since the profit earned on clients without claim is typically much lower than the loss incurred on clients with claim in the examined MTPL sector,
- The company is basically interested in maximising profit, but the ROC curve is not an effective tool for analysing this.

Therefore, in this study we have introduced a utility function with the following considerations. It is assumed that the company uses the predictive model as an underwriting tool and accepts those customers for whom the model predicts no claim at the applied cut-off value. Thus, the company accepts $a + c$ number of contracts (see Table 2). Let us suppose that the company has

one unit of 'profit' on each contract without claim and L unit of 'loss' ($L < 0$) on each contract with claim. The company's goal is to maximise $U(\alpha) = a + L \cdot c$ utility function (which under the simplifying assumptions used is the profit of the company) with the optimal α cut-off value. In addition, the ratio of the retained portfolio $(a + c) / (a + b + c + d)$ is calculated also (Burka, Kovács & Szepesváry, 2021). The idea of using the general assessment function of a confusion matrix was also used by (Figini & Uberti, 2010).

Generalized Linear Model (GLM), Random Forest (RF) and Neural Network (NN) procedures were fitted to explain the claim occurrence variable. Furthermore, using the three fitted models, a fourth mixed model was defined that combines the advantages of each procedure. Latter is called Vote model in the paper, which means that the probability estimates are obtained by averaging the estimated probabilities of each basic model with different weights. The optimal weights are given by 1 GLM + 2 Random Forest + 1 Neural Network.



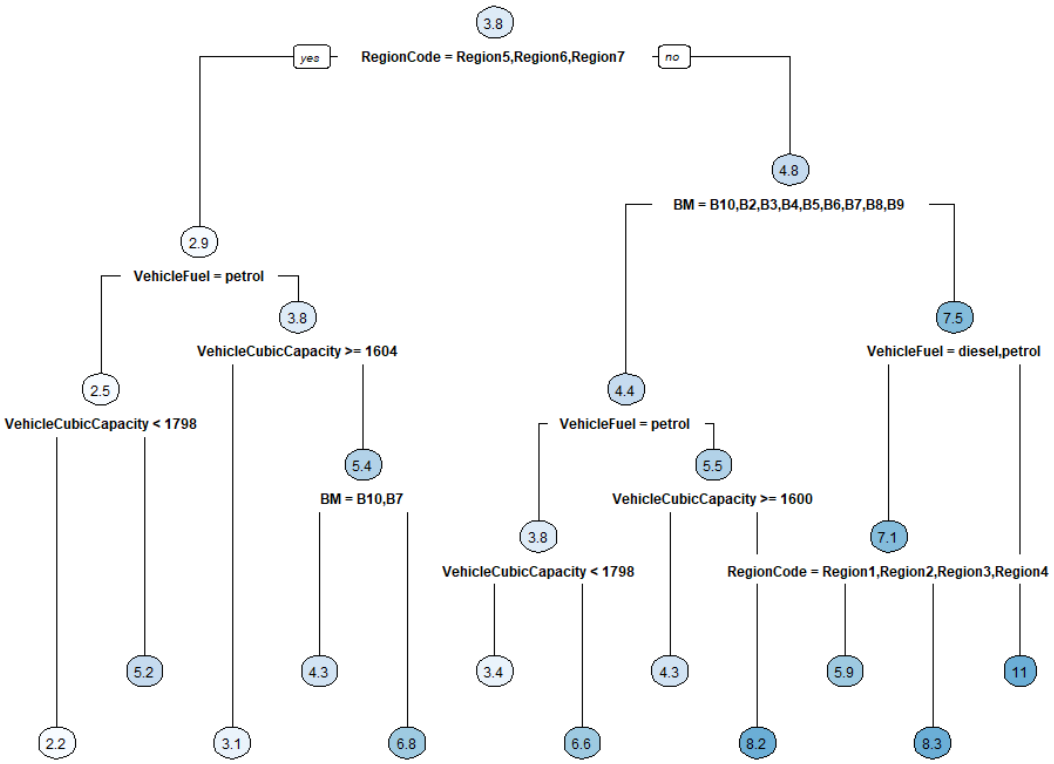
10. Figure: ROC curves for different processes.
Source: (Burka, Kovács & Szepesváry, 2021)

Based on the ROC curves (Figure 10) and AUC values, the Vote model obviously performed the best on the test set. The analysis related to the utility function U(α) was also performed for all four models. Based on the outputs, except for the $L = -10$ case, the performance of the Vote model was always the best. The paper highlights that based on the public Hungarian data (considering the average premiums and losses for Hungarian MTPL market), a size of $-20 \leq L \leq -10$ parameter can be realistic (Burka, Kovács & Szepesváry, 2021).

The study presents two options for applying the generated Vote Model to tariff calculation. The first option is to use the results of the Vote model in an unchanged form: for any combination of the explanatory variables' values, it is possible to determine the unique probability according to the model, which can be the first pillar of the tariff calculation. The advantage of the method is that it preserves the full variance of the model, but the disadvantage is its complex structure and the fact that it is almost impossible to describe or publish the results (Burka, Kovács & Szepesváry, 2021). In Hungary, it is compulsory to publish MTPL insurance tariffs in written form, so a non-written machine algorithm does not meet the legal requirements.

To solve these problems, the following procedure was used in the study. A visualisable explanatory model was constructed for the estimated probabilities of the Vote model. This was done using a regression tree: the Vote model probabilities as dependent variables were modelled using the original explanatory variables. The single leaves of the resulting tree form clusters, for which a decision rule can be easily defined, and which groups can be considered to have the same probability of claim occurrence under a simplifying assumption in the premium calculation. The advantage of the method is the descriptive structure, but the disadvantage is the loss of variance compared to the original Vote model (Burka, Kovács & Szepesváry, 2021). The optimal number of clusters was determined using the rank correlation indicator and a formula similar to the sum of squares within group (SSW) in case of ANOVA model.

The study suggests two possible cluster numbers for the specific data: 12 clusters or 21 clusters may be the optimal choice, but more detailed results are presented for the cluster number 12 for better visualisation of the results (Burka, Kovács & Szepesváry, 2021). Figure 11 shows the decision tree for $k = 12$ clusters. For each cut, the left direction represents the "yes" and the right direction the "no" split, and the circled numbers represent the average estimated probabilities of claim occurrence for the corresponding group. The 12 clusters results from the decision tree preserve 64% of the variance of the original Vote model. Even this simplified model shows several interactions between explanatory variables when estimating the probability of loss.



11. Figure: Cluster classification proposed by the decision tree for a tree depth of 4 (cluster number 12). Source: (Burka, Kovács & Szepesváry 2021)

With the clusters resulting from the decision tree, the criterion of publishability of the results is already achieved. The next important question is the predictive performance of the new clustering model: should the Vote model be downgraded to simplified clusters in the light of the loss of variance? Based on the previous evaluation techniques, the performance of the method was investigated by ranking the results of the decision tree simplified Vote model for different cluster numbers compared to the basic methods.

Model	AUC, (Rank)
GLM	0.6446, (4.)
Random forest	0.6657, (2.)
Neural network	0.6347, (6.)
Vote model	0.6791, (1.)
Vote mode (k=12 clustered)	0.6361, (5.)
Vote model (k=21 clustered)	0.6523, (3.)

3. Table: AUC values supplemented with results from the decision tree simplified Vote model. Source: (Burka, Kovács & Szepesváry, 2021)

$L = -20$	GLM	Random forest	Neural network	Vote model	Vote mode (k=12 clustered)	Vote model (k=21 clustered)
Maximum utility	16 927	18 495	16 257	19 043	17 280	17 877
Kept portfolio (%)	81.6%	88.0%	82.0%	85.0%	82.1%	80.0%
Optimal cut-off value	5.0%	8.6%	4.7%	5.9%	4.4%	4.2%
Rank	5	2	6	1	4	3

4. Table: Analysis of the utility function $U(\alpha)$ with $L = -20$ parameter, supplemented with results from the decision tree simplified Vote model
Source: (Burka, Kovács & Szepesváry, 2021)

Table 3 shows the AUC values of the different methods, and Table 4 shows the analysis of the utility function $\boldsymbol{U(\alpha)}$ with $\boldsymbol{L = -20}$ parameters, supplemented with the results of the decision tree simplified Vote model. Both tables show that the Vote model and the random forest methods prove to be the best performing. However, with increasing the number of clusters to $\boldsymbol{k = 21}$, the simplified Vote model becomes the 3rd best procedure, so that both GLM and neural network methods can be outperformed by the new technique.

By elaborating this thought in details, the cited study and the thesis provides evidence for Hypothesis 5.

4. References

1. 192/2000. (XI. 24.) *Goverment Decree*.
2. 2009/138/EC. *Directive 2009/138/EC of the European Parliament and of the Council*
3. Banyár, J. (2016). *Életbiztosítás (2. javított, bővített kiadás)*. Budapesti Corvinus Egyetem, Budapest.
4. Bølviken, E. (2014). *Computation and Modelling in Insurance and Finance*. Cambridge University Press, Cambridge.
5. Csóka, P. (2003). Koherens kockázatmérés és tőkeallokáció. *Közgazdasági Szemle*, *50*(10), 855–880. Retrieved from: http://epa.oszk.hu/00000/00017/00097/pdf/2csoka.pdf (last download: 03.01.2022)
6. Figini, S., & Uberti, P. (2010). Model assessment for predictive classification models. *Communications in Statistics - Theory and Methods*, *39*(18), 3238–3244. https://doi.org/10.1080/03610920903243751
7. Frees, E. W., Meyers, G., & Derrig, R. A. (2016). *Predictive modeling applications in actuarial science: Volume 2, case studies in insurance* (E. Frees, G. Meyers, & R. Derrig, Ed.). Cambridge University Press, Cambridge.
8. Gray, R., & Kovács, E. (2001). Az általánosított lineáris modell és biztosítási alkalmazásai. *Statisztikai Szemle, 79(8)*, 689–702.
9. Hajek, M. (2005). *NEURAL NETWORKS*. University of KwaZulu - Natal, KwaZulu - Natal.
10. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer. https://doi.org/10.1007/978-0-387-84858-7
11. IASB. (2017). *IFRS Standards – IFRS 17 Insurance contracts*.
12. Kafková, S., & Krivánková, L. (2014). Generalized linear models in vehicle insurance. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, *62*(2), 383–388. https://doi.org/10.11118/actaun201462020383
13. Kirchgässner, G., Wolters, J., & Hassler, U. (2013). Introduction to modern time series analysis. In *Springer Verlag*. Springer.
14. Kovács, E. (2011). *Pénzügyi adatok statisztikai elemzése*. Tanszék Kft., Budapest.
15. Ohlsson, E., & Johansson, B. (2010). *Non-Life Insurance Pricing with Generalized Linear Models*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-10791-7
16. Tucker, H. G. (1959). A Generalization of the Glivenko-Cantelli Theorem. *The Annals of Mathematical Statistics*, *30*(3). https://doi.org/10.1214/aoms/1177706212
17. Vékás, P. (2011). Túlélési modellek. In E. Kovács (Ed.), *Pénzügyi adatok statisztikai elemzése* (IV. bővítetett kiadás, 173–194). Tanszék Kft., Budapest.
18. Vékás, P. (2019). *Az élettartam-kockázat modellezése*. Budapesti Corvinus Egyetem, Budapest.
19. Yeo, A. C. (2011). Neural Networks for Automobile Insurance Pricing. In *Encyclopedia of Information Science and Technology, Second Edition*, 2794–2799. https://doi.org/10.4018/978-1-60566-026-4.ch446

5. List of publications

Journal articles (international)

1. Burka, D., Kovács, L., & Szepesváry, L. (2021). Modelling MTPL insurance claim events: Can machine learning methods overperform the traditional GLM approach? *Hungarian Statistical Review*, *4*(2), 34–69. https://doi.org/10.35618/hsr2021.02.en034
2. Burka, D., Puppe, C., Szepesváry, L., & Tasnádi, A. (2022). Voting: A machine learning approach. *European Journal of Operational Research*, *299*(3), 1003–1017. https://doi.org/10.1016/j.ejor.2021.10.005

Journal articles (Hungarian)

3. Szepesváry, L. (2019). Onerous test, avagy az IFRS 17 szerinti veszteségességi vizsgálat : Aktuáriusi és informatikai kihívások egy életbiztosítási portfólió példáján. *Biztosítás és Kockázat*, *6*(2), 18–37. https://doi.org/10.18530/bk.2019.2.18
4. Szepesváry, L. (2022). Életbiztosítások törlési rátáinak elemzése – Hogyan hatottak a hozam- és inflációs környezet változásai valamint a COVID 19-cel kapcsolatos lezárások az ügyfélviselkedésre? *(At the time of finalizing the thesis, the article was submitted, awaiting review status, and if successfully accepted, will be published in the journal Financial and economic review in June 2022.)*

Conference papers (Hungarian)

5. Szepesváry, L. (2015). Dinamikus modellek alkalmazása életbiztosítások cash flow előrejelzésére. In *Tavaszi szél 2015 Konferenciakötet II. kötet* (pp. 581-599.). Líceum Kiadó, Eger, Doktoranduszok Országos Szövetsége.

Book chapter (Hungarian)

6. Szepesváry, L. (2016). Túlélési modellek. In B. Szüle (Ed.), *Többváltozós adatelemzési számítások* (pp. 92–110). Retrieved from: http://unipub.lib.uni-corvinus.hu/2438/ (last download 2022.03.06)