MODELLING OF LIFE LAPSE RATES

Radek Hendrych

Aktuárský seminář | 6/12/2019

ČSOB banka i pojišťovna

AGENDA

1. Motivation
2. Key factors determining lapse
3. Modelling techniques: Intro
4. Modelling techniques: Logistic regression
5. Modelling techniques: Other approaches

MOTIVATION

TERMINOLOGY

- The term ***lapse*** refers to the termination of an insurance contract before its maturity.
- Occasionally, it is distinguished between:
  - lapse = the termination of policy without payout to the policyholder
  - surrender = the termination of policy where a (cash) surrender value is paid out to the policyholder
- Using the single term “lapse” for both of these concepts is consistent with standard measures of lapse rates as they typically include lapsed policies as well as surrendered ones.

PROJECTION OF LIFE LAPSE RATES = KEY INPUT

- The projection of (life) lapse rates of insurance contracts is one of the key inputs of actuarial (liability) cash flow models since it is involved in the portfolio evolution and (eventually) in the surrender outflow:

$$CF_t^{LIAB} = PREMIUM_t - CLAIMS_t - EXPENSES_t - COMMISSIONS_t$$

- Lapse is one of the key embedded option in life insurance contracts (←dynamic policyholder behavior). It is important in the field of valuation and management of embedded options in life insurance contracts from the following points of view:
  - industry
  - regulatory
  - company

INDUSTRY VIEW I

CFO Forum: MCEV Principles (principle 7)

"Allowance must be made in the MCEV for the potential impact on future shareholder ***cash flows of all financial options and guarantees*** *within the in-force covered business. The allowance for the time value of financial options and guarantees must be based on* ***stochastic techniques using methods and assumptions consistent with the underlying embedded value****. All projected cash flows should be valued using economic assumptions such that they are valued in line with the price of similar cash flows that are traded in the capital markets."*

INDUSTRY VIEW II

Zurich Insurance Group (Annual results 2018 - Embedded value report)

*"… policyholders can exercise an option against the life company in certain circumstances, such as **to surrender a policy**. For example an increase in the lapse rates could be assumed when interest rates rise (or a corresponding reduction when interest rates fall). This dynamic effect in relation to **lapse rates** has been allowed for in the stochastic models."*

Baloise Group (Market consistent embedded value report 2017)

*"Management's selection of bonus rates and **policyholder lapse rates** are key variables for which dynamic assumptions – varying depending on the economic scenario – are applied in stochastic projections."*

REGULATORY VIEW I

Solvency 2 (Directive 2009/138/EC, Article 79)

"When calculating technical provisions, insurance and reinsurance undertakings shall take account of the value of financial guarantees and any contractual options included in insurance and reinsurance policies.

Any assumptions made by insurance and reinsurance undertakings with respect to the likelihood that policy holders will exercise contractual options, including ***lapses and surrenders****, shall be realistic and based on current and credible information. The assumptions shall take account, either explicitly or implicitly, of the impact that future changes in financial and non-financial conditions may have on the exercise of those options."*

REGULATORY VIEW II

- **S2 lapse risk sub-module (life underwriting risk module):**

  "Lapse risk is the risk of loss or adverse change in liabilities due to a change in the expected exercise rates of policyholder options…"

- The Solvency 2 capital requirement for the lapse risk is calculated as maximum of three stress scenarios:
  - *LAPSE_DOWN*: long-term decrease of lapse rates by 50% (it shall not exceed 20 pp)
  - *LAPSE_UP*: long-term increase of lapse rates by 50% (it shall not exceed 100%)
  - *LAPSE_MASS*: mass lapse event of 40% of all policyholders

  Note: Scenarios are applied only when the exercise of this option would result in an increase of S2 technical provisions without the S2 risk margin.

COMPANY VIEW

- ALM
  (duration matching ← liability cash flow profile)
- Design & pricing of life products
  (profitability measures, commissions set-up)
- Monitoring & reporting
  (profitability / portfolio monitoring)

POSSIBLE CONSEQUENCES OF LAPSE

Lapse could affect an insurer's liquidity and profitability:

1. The insurer might suffer losses from lapsed policies due to upfront investments for acquiring new business.
2. The insurer might face adverse selection with respect to mortality and morbidity as customers with adverse health are less likely to lapse their contract.
3. The insurer might be exposed to a **liquidity risk** when forced to pay the cash surrender value for lapsed policies.
4. High lapses can have a negative effect on the insurer's reputation that might result in even more lapses, as well as harm new business.

KEY FACTORS DETERMINING LAPSE

DETERMINANTS OF LIFE INSURANCE LAPSE

In the academic and practitioner's literature, the following two groups of drivers of life insurance lapse are commonly considered:

1. Market (environment) dependent
   [macro-economic indicators]
2. Market (environment) independent
   [product & contract features, policyholder characteristics]

MARKET (ENVIRONMENT) DEPENDENT FACTORS

- Macro-economic variables & market conditions:
  - economic growth rates (GDP)
  - unemployment rate
  - financial crisis indication
  - interest rates (and their volatility)
  - inflation
  - returns on stock market (and their volatility)
- Regarding market (environment) dependent factors, the ***interest rate*** and ***emergency fund*** hypotheses have been repeatedly analyzed.

INTEREST RATE HYPOTHESIS

The interest rate hypothesis assumes that savings through life insurance is sensitive to rates of return:

- Kuo et al. (2003) argue that policyholders lapse their policies to exploit higher interest rates and/or lower premiums in the market when market interest rates rise. Increasing interest rates act as opportunity cost for owning life insurance.
- Dar and Dodds (1989) conjecture that lapse rates are negatively related to internal rates of return, e.g., surplus participation, and positively related to external rates of return, e.g., market interest rates or rates of return on other financial assets.

EMERGENCY FUND HYPOTHESIS

The emergency fund hypothesis conjectures that personal financial distress forces policyholders to lapse their contracts in order to access the cash surrender value (unemployment rate commonly used as an indicator of adverse economic condition).

INTEREST RATE x EMERGENCY FUND HYPOTHESES

- Outreville (1990) studies the emergency fund hypothesis with lapse rate data of whole-life insurance in the USA and Canada. The results provide consistent evidence for the emergency fund hypothesis.
- Dar and Dodds (1989) test both hypotheses using endowment policies of UK life insurers. They find evidence in favor of the emergency fund hypothesis, but no significant relationship between surrenders and rates of return.
- Kuo et al. (2003) investigate the competing lapse rate hypotheses for US data using a cointegration analysis to address long-term lapse dynamics. They find that the interest rate effect is economically more significant than the unemployment rate in explaining the lapse rate dynamics. The interest rate hypothesis is favored over the emergency fund hypothesis.
- Kiesenbauer (2012) claims that the interest rate and emergency fund hypotheses do not hold for traditional life insurance products in the German market. However, both hypotheses are supported when other business representing almost exclusively unit-linked products is considered.

To conclude, the results of these studies examining the interest rate and emergency fund hypotheses are inconsistent.

MARKET (ENVIRONMENT) INDEPENDENT FACTORS

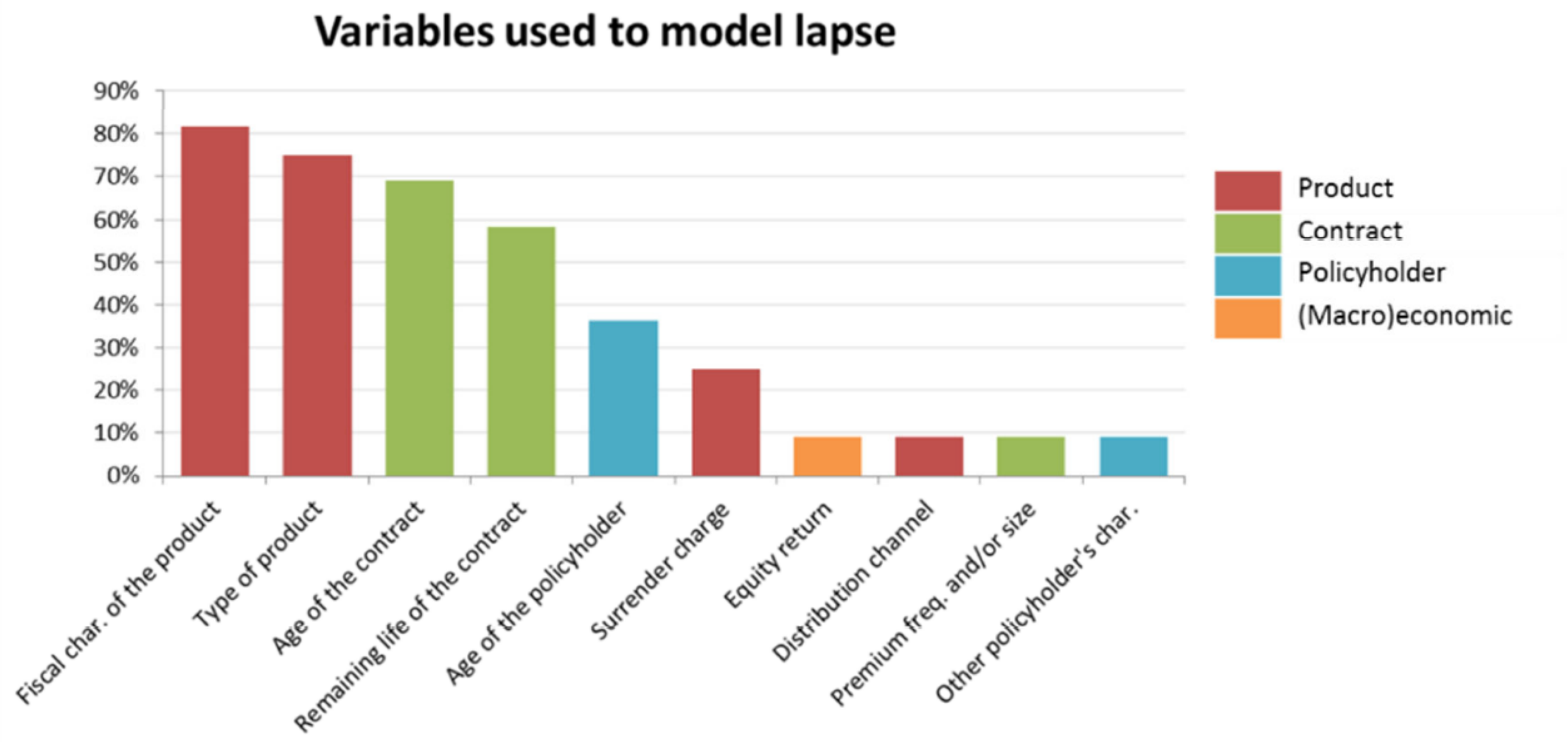
Product & contract features

- type of product
- distribution channel
- age of the contract
- remaining lifetime of the contract
- premium frequency and/or size
- value of the insurance
- surrender charge
- optimal moment of lapsation
- risk / saving premium
- sum insured
- policy status (active, paid-up, …)
- tax / regulation effects
- calendar year / seasonal effects

Policyholder characteristics

- age of policy holder
- gender
- mortality rate

FACTORS OF LIFE INSURANCE LAPSE: CASE STUDY



Results of Reacfin survey (2016) of the Belgian insurance market

MODELLING TECHNIQUES: INTRO

LAPSE RATE MODELLING TECHNIQUES

- There is no wide / clear consensus how to model (life) lapse rates.
- From the practical perspective, one should consider:
  - long-term cash flow projections (← methods)
  - modelling outputs would be efficiently implemented within the actuarial cash flow models
    [R, Python, SAS, etc. x Prophet]
- Applicable methodologies:
  - simple approaches
  - one-factor models
  - generalized linear models (GLM) and its extensions (GEE, GLMM, GAM)
  - regression trees: CART, random forests, gradient boosting
  - neural networks

TRADITIONAL (SIMPLE) APPROACH

- The lapse rate in the policy segment defined by factors $f_1, \dots, f_p$ for the time instant *t* can be derived using:

$$\hat{\ell}_t^{f_1,\dots,f_p} = \frac{\#lapses_t^{f_1,\dots,f_p}}{\#contracts\_in\_force_t^{f_1,\dots,f_p}}, t = 1, \dots, T$$

- Further, weighting or expert judgements are employed.
- It can be easily implemented within actuarial cash flow models.

ONE-FACTOR MODELS

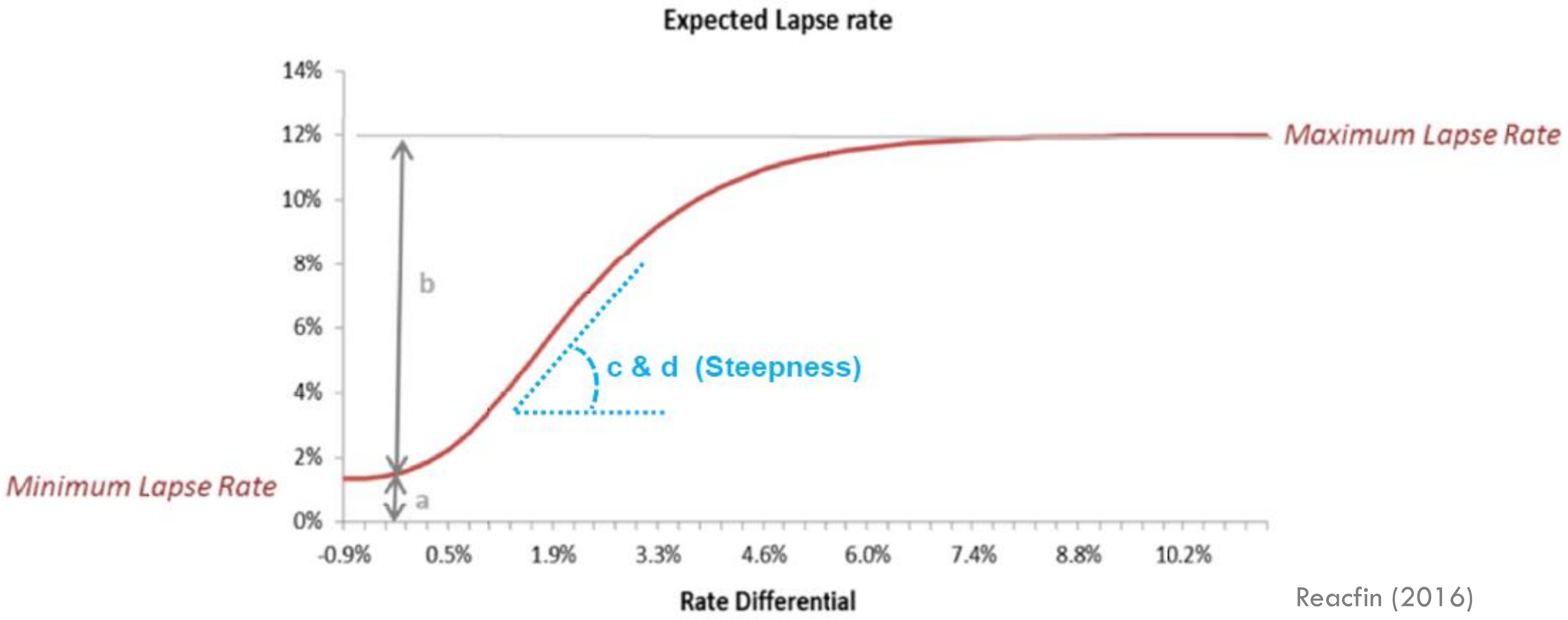
- One-factor models suggest a simple relationship between a single variable and lapse rates.
- A common predictor is the interest rate differential Δr , i.e. the difference between a market rate and the expected rate on the life insurance contract (Reacfin, 2016).

- Arctan model: $\ell_t = a + b \cdot \arctan(c \cdot \Delta r_t - d)\,, t = 1, \dots, T$
- Parabolic model: $\ell_t = a + b \cdot \operatorname{sign}(\Delta r_t) \cdot (\Delta r_t)^2, t = 1, \dots, T$
- Exponential model: $\ell_t = a + b \cdot e^{-c \cdot e^{-d \Delta r_t}}, t = 1, \dots, T$

[a, b, c, d are real parameters with particular constraints → to be calibrated]

ONE-FACTOR MODELS: EXPONENTIAL MODEL

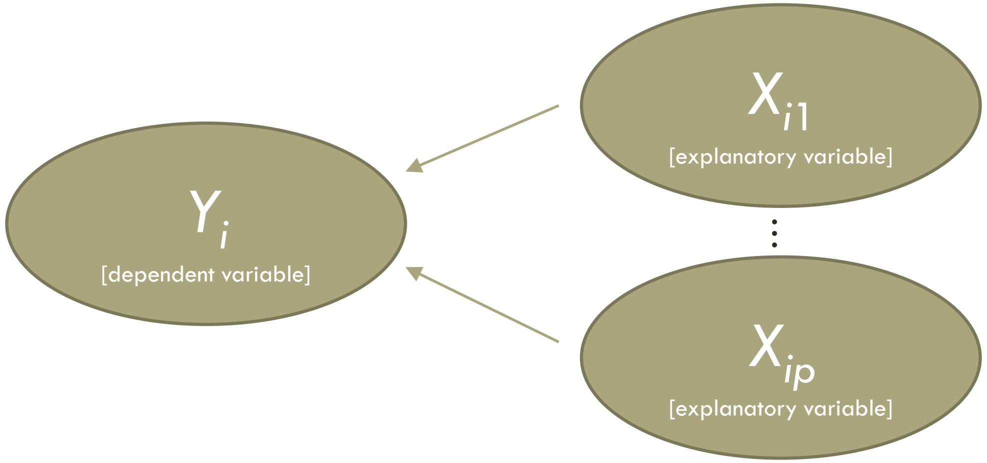
- $f(\Delta r_t) = a + b \cdot e^{-c \cdot e^{-d\Delta r_t}}, c, d > 0 \leftrightarrow f(-\infty) = a \geq 0 \leftrightarrow f(+\infty) = a + b \leq 1$



Reacfin (2016)

GENERALIZED LINEAR MODELS (GLM) I

- Academic and practitioner's literature frequently applies the Generalized Linear Models (GLM) to model and/or predict lapse rates of (life) insurance contracts since they are able to:
  - include many explanatory variables (drivers),
  - capture interactions between explanatory variables,
  - be simply translated into the standard actuarial software (Prophet, EARNIX).



GENERALIZED LINEAR MODELS (GLM) II

GLM is defined by the following components / assumptions:

1. Each component of $\boldsymbol{Y}$ is independent and follows a distribution of the family of exponential distributions.
2. The linear combination of the covariates in $\boldsymbol{X}$ give the predictor: $\eta = \boldsymbol{X\beta}$, where $\boldsymbol{\beta}$ is the vector of real parameters.
3. There exist a link function g (smooth and invertible) such that $\mathrm{E}[\boldsymbol{Y}|\boldsymbol{X}] = \boldsymbol{\mu} = g^{-1}(\boldsymbol{X\beta})$.

ČSOB banka i pojišťovna

GENERALIZED LINEAR MODELS (GLM) III

Distribution	Support of distribution	Link name	Link function $X\beta = g(\mu)$	Mean function
Normal	$(-\infty, +\infty)$	Identity	$\boldsymbol{X\beta} = \boldsymbol{\mu}$	$\boldsymbol{\mu} = \boldsymbol{X\beta}$
Exponential	$(0, +\infty)$	Negative inverse	$\boldsymbol{X\beta} = -\boldsymbol{\mu}^{-1}$	$\boldsymbol{\mu} = -(\boldsymbol{X\beta})^{-1}$
Gamma	$(0, +\infty)$	Negative inverse	$\boldsymbol{X\beta} = -\boldsymbol{\mu}^{-1}$	$\boldsymbol{\mu} = -(\boldsymbol{X\beta})^{-1}$
Poisson	$\{0,1,2, \dots\}$	Log	$\boldsymbol{X\beta} = \log(\boldsymbol{\mu})$	$\boldsymbol{\mu} = \exp(\boldsymbol{X\beta})$
Bernoulli	$\{0,1\}$	Logit	$\boldsymbol{X\beta} = \log\left(\frac{\boldsymbol{\mu}}{1-\boldsymbol{\mu}}\right)$	$\boldsymbol{\mu} = \frac{\mathbf{1}}{\mathbf{1} + \exp(-\boldsymbol{X\beta})}$
Binomial	$\{0,1, \dots, K\}$	Logit	$\boldsymbol{X\beta} = \log\left(\frac{\boldsymbol{\mu}}{1-\boldsymbol{\mu}}\right)$	$\boldsymbol{\mu} = \frac{\mathbf{1}}{\mathbf{1} + \exp(-\boldsymbol{X\beta})}$
…	…	…	…	…

MODELLING TECHNIQUES: LOGISTIC REGRESSION

LOGISTIC REGRESSION: FRAMEWORK

- The most frequent GLM framework used for modelling / predicting lapse rates of (life) insurance contracts.
- Y_i is the binary dependent variable (the Bernoulli distribution).
  [0 = *i*-th contract not lapsed, 1 = *i*-th contract lapsed]
- The lapse probability is expressed as:

$$p_{\mathrm{i}} = p(\boldsymbol{X}_i, \boldsymbol{\beta}) = \mathrm{P}[Y_i = 1 | \boldsymbol{X}_i, \boldsymbol{\beta}] = \frac{\exp\left(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{\beta}\right)}{1 + \exp\left(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{\beta}\right)}$$

i.e.

$$\operatorname{logit}(p_i) = \log\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}$$

LOGISTIC REGRESSION: KEY MODELLING STEPS

1. Data [cleaning, NAs handling, explanatory analysis]
2. Model identification [selecting regressors, interactions, offsets]
3. Model estimation [maximum log-likelihood]
4. Model verification [model diagnostics, model evaluation]

- R: `glm(formula, family=binomial(link="logit"), data)`
- SAS: `PROC GENMOD / PROC LOGISTIC`

LOGISTIC REGRESSION: MODEL IDENTIFICATION

- Selecting explanatory variables:
  - stepwise selection [forward selection, backward or bidirectional elimination]
  - graphical diagnostics
  - expert judgement [ev. offset terms]
- Interactions
- Multicollinearity diagnostics [VIF, generalized VIF, etc.]

LOGISTIC REGRESSION: MODEL ESTIMATION

- The unknown model parameters are estimated using MLE:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \mathrm{P}(Y = y_i | \boldsymbol{X} = \boldsymbol{x}_i, \boldsymbol{\beta}) = \prod_{i=1}^{n} p(\boldsymbol{x}_i, \boldsymbol{\beta})^{y_i} \big(1 - p(\boldsymbol{x}_i, \boldsymbol{\beta})\big)^{1-y_i}$$

$$\Downarrow$$

$$\underset{\boldsymbol{\beta}}{\operatorname{argmin}} - \frac{\ell(\boldsymbol{\beta})}{n} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} - \frac{1}{n} \sum_{i=1}^{n} \big[y_i \log\big(p(\boldsymbol{x}_i, \boldsymbol{\beta})\big) + (1 - y_i) \log\big(1 - p(\boldsymbol{x}_i, \boldsymbol{\beta})\big) \big]$$

- Regularization can be applied (by penalizing the loss function).

LOGISTIC REGRESSION: MODEL DIAGNOSTICS I

- Pearson residuals:

$$r_i = \frac{y_i - \hat{p}_i}{\sqrt{\hat{p}_i(1 - \hat{p}_i)}}$$

- Deviance residuals [`residuals()` in R]:

$$d_i = +\sqrt{-2\{y_i \log(\hat{p}_i) + (1 - y_i)\log(1 - \hat{p}_i)\}} \text{ for } y_i = 1$$

$$d_i = -\sqrt{-2\{y_i \log(\hat{p}_i) + (1 - y_i)\log(1 - \hat{p}_i)\}} \text{ for } y_i = 0$$

- RSS-like statistics as goodness of fit tests:
  - deviance: $D = \sum_{i=1}^{n} d_i^2$
  - Pearson: $X^2 = \sum_{i=1}^{n} r_i^2$
- Hosmer & Lemeshow test (goodness of fit), Wald test (significance)

LOGISTIC REGRESSION: MODEL DIAGNOSTICS II

- Standardized Pearson and deviation residuals:

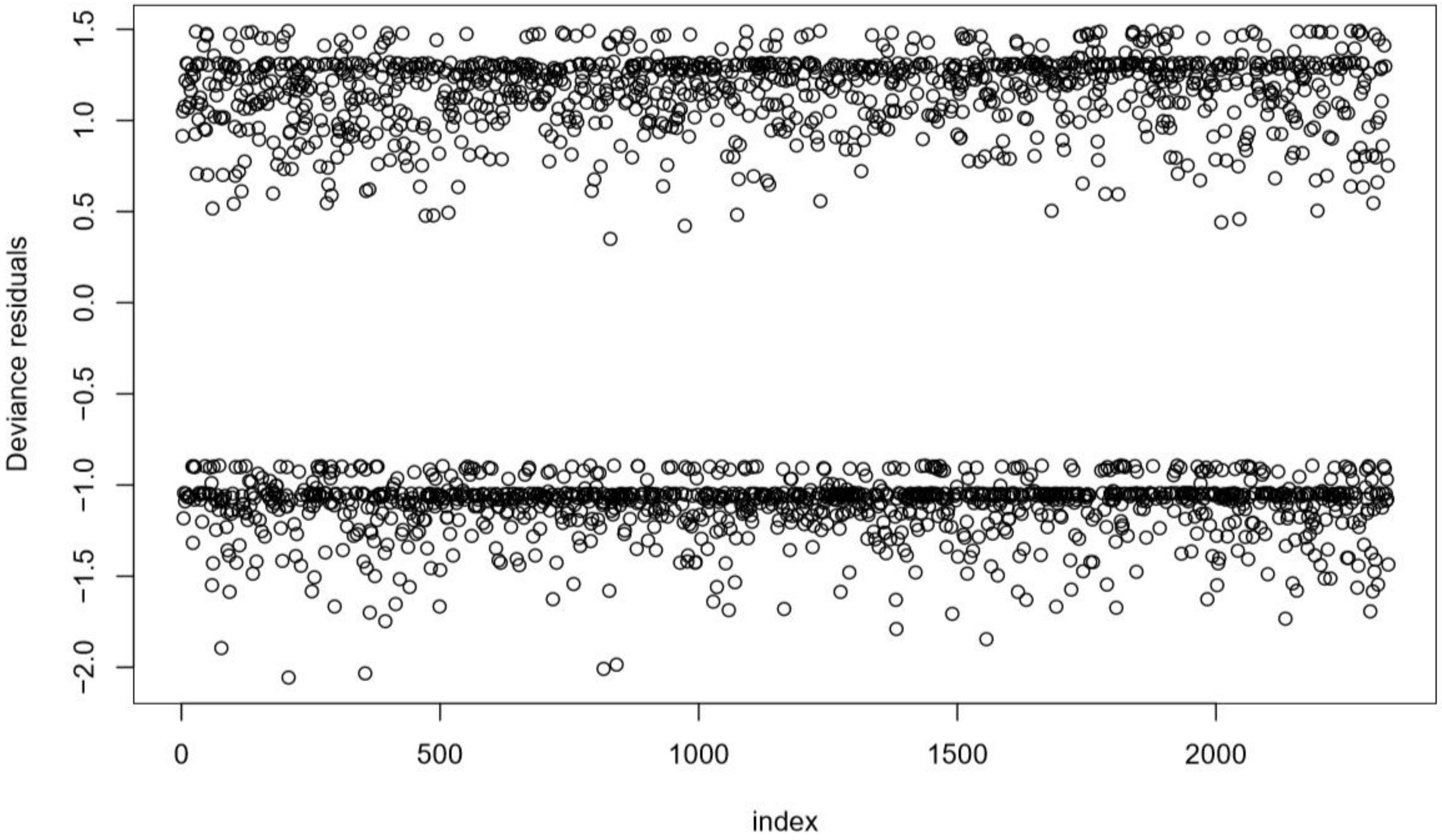
$$r_i^S = \frac{r_i}{\sqrt{1-h_{ii}}} \quad \text{and} \quad d_i^S = \frac{d_i}{\sqrt{1-h_{ii}}}$$

[h_{ii} denotes the i-th diagonal element of the hat matrix]

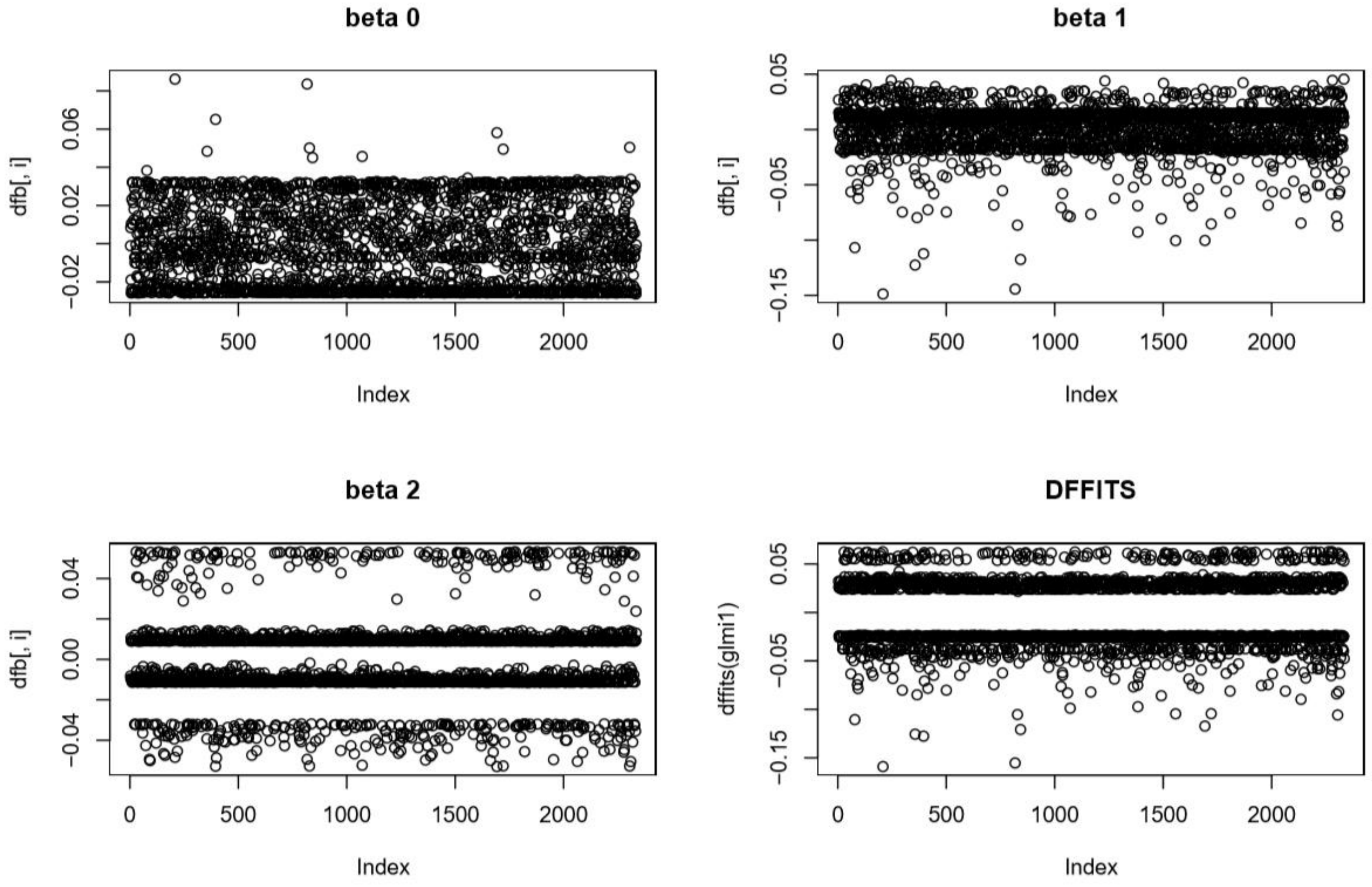
- Influential observations:
  - DFBETAS (measures the effect of an individual observation on each estimated parameter of the fitted model)
  - DFFITS (measures the effect of an individual observation on the fitted value)
  - Cook's distance (measures the distance between the full-data estimate/prediction and the estimate/prediction obtained by deleting an individual observation)
  - C and CBAR (measure the influence of an individual observation on $\widehat{\boldsymbol{\beta}}$):

$$C_i = \frac{h_{ii} r_i^2}{(1-h_{ii})^2} \quad \text{and} \quad \bar{C}_i = \frac{h_{ii} r_i^2}{(1-h_{ii})}$$

LOGISTIC REGRESSION: MODEL DIAGNOSTICS III



LOGISTIC REGRESSION: MODEL DIAGNOSTICS IV



LOGISTIC REGRESSION: MODEL SELECTION I

- Consistency with expert judgements (significance, signs of parameter estimates)
- Information criteria (AIC, BIC)
- $R^2_{McFadden} = 1 - {}^{\ell_{model}}/_{\ell_{null}} \in [0,1]$
- Backtesting (predicted lapses *vs.* real lapses)

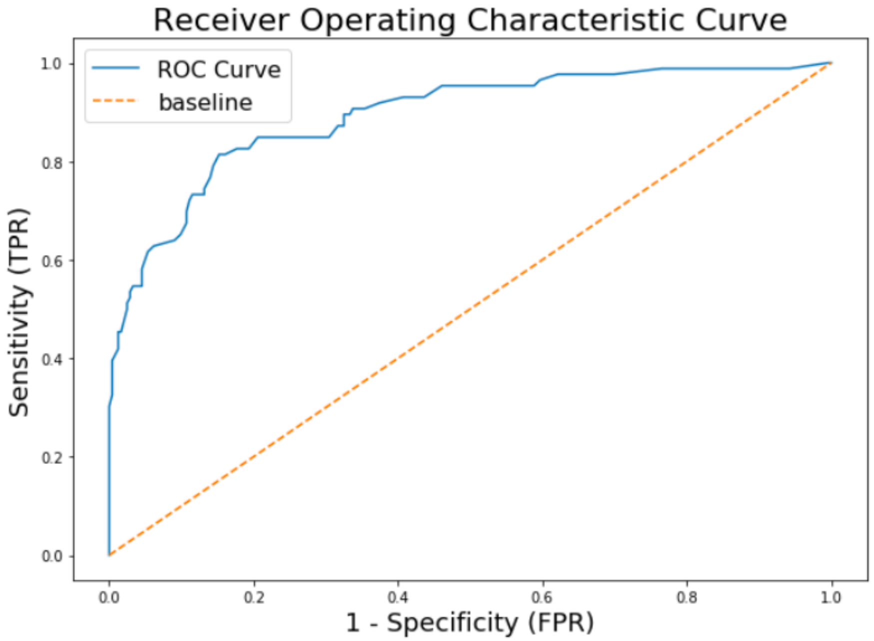
ČSOB banka i pojišťovna

LOGISTIC REGRESSION: MODEL SELECTION II

- Confusion matrix
- ROC curve / AUC [based on confusion matrices for various thresholds]

Predicted Class

		Positive	**Negative**	
Actual Class	**Positive**	True Positive (TP)	False Negative (FN) **Type II Error**	**Sensitivity** $\frac{TP}{(TP+FN)}$
	Negative	False Positive (FP) **Type I Error**	True Negative (TN)	**Specificity** $\frac{TN}{(TN+FP)}$
		Precision $\frac{TP}{(TP+FP)}$	**Negative Predictive Value** $\frac{TN}{(TN+FN)}$	**Accuracy** $\frac{TP+TN}{(TP+TN+FP+FN)}$



MODELLING TECHNIQUES: OTHER APPROACHES

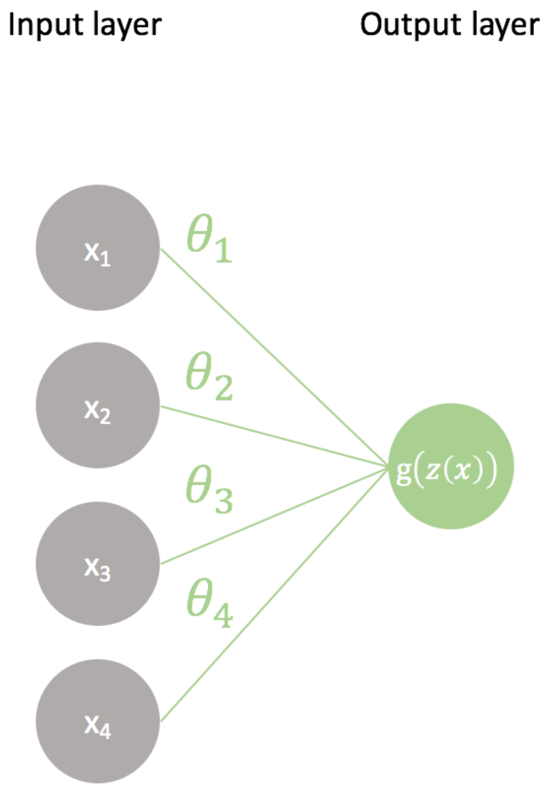
LOGISTIC REGRESSION: EXTENSIONS

- **Generalized estimating equations** (GEE) allow for the correlation between observations without the use of an explicit probability model for the origin of the correlations (“population-averaged” effects).
- **Generalized linear mixed models** (GLMM) include random effects in the linear predictor with an explicit probability model for the origin of the correlations (“subject-specific” effects).
- **Generalized additive models** (GAM) are extension in which the linear predictor $\boldsymbol{\eta}$ is not restricted to be linear in the covariates $\boldsymbol{X}$ but is the sum of smoothing functions:

$$\eta_i = \beta_0 + f_1(x_{i1}) + \cdots + f_p(x_{ip}),$$

where smoothing functions $f_j(\cdot)$ are estimated from the data.

LOGISTIC REGRESSION → NEURAL NETWORK



Logistic Regression

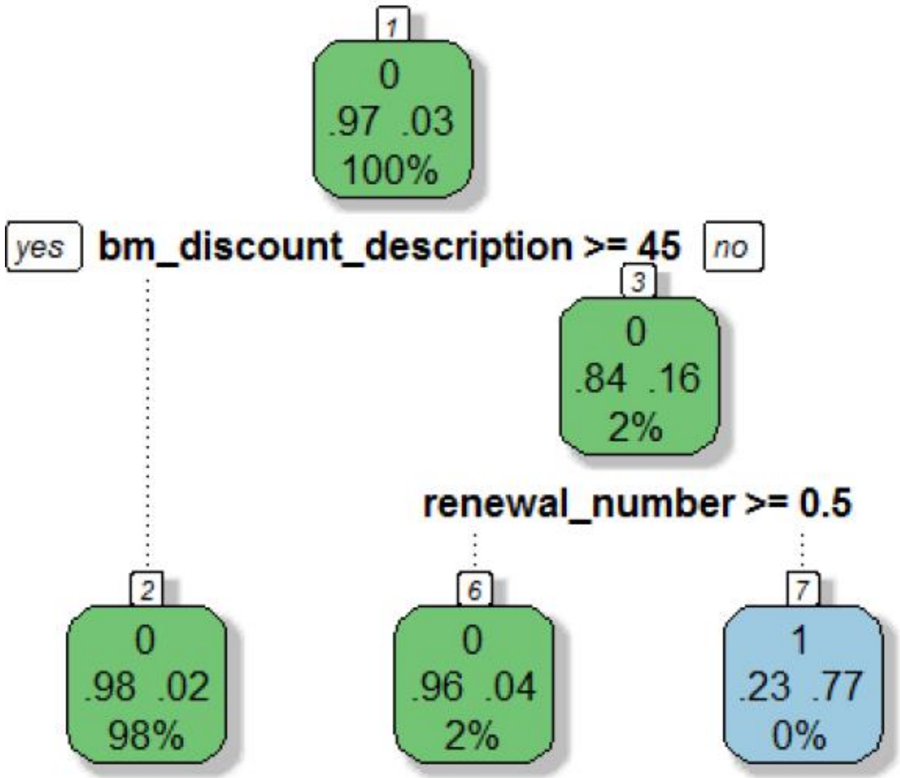
$$z(x) = w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4 + b$$

$$g(z) = \frac{1}{1+e^{-z}}$$

CLASSIFICATION AND REGRESSION TREES (CRT)

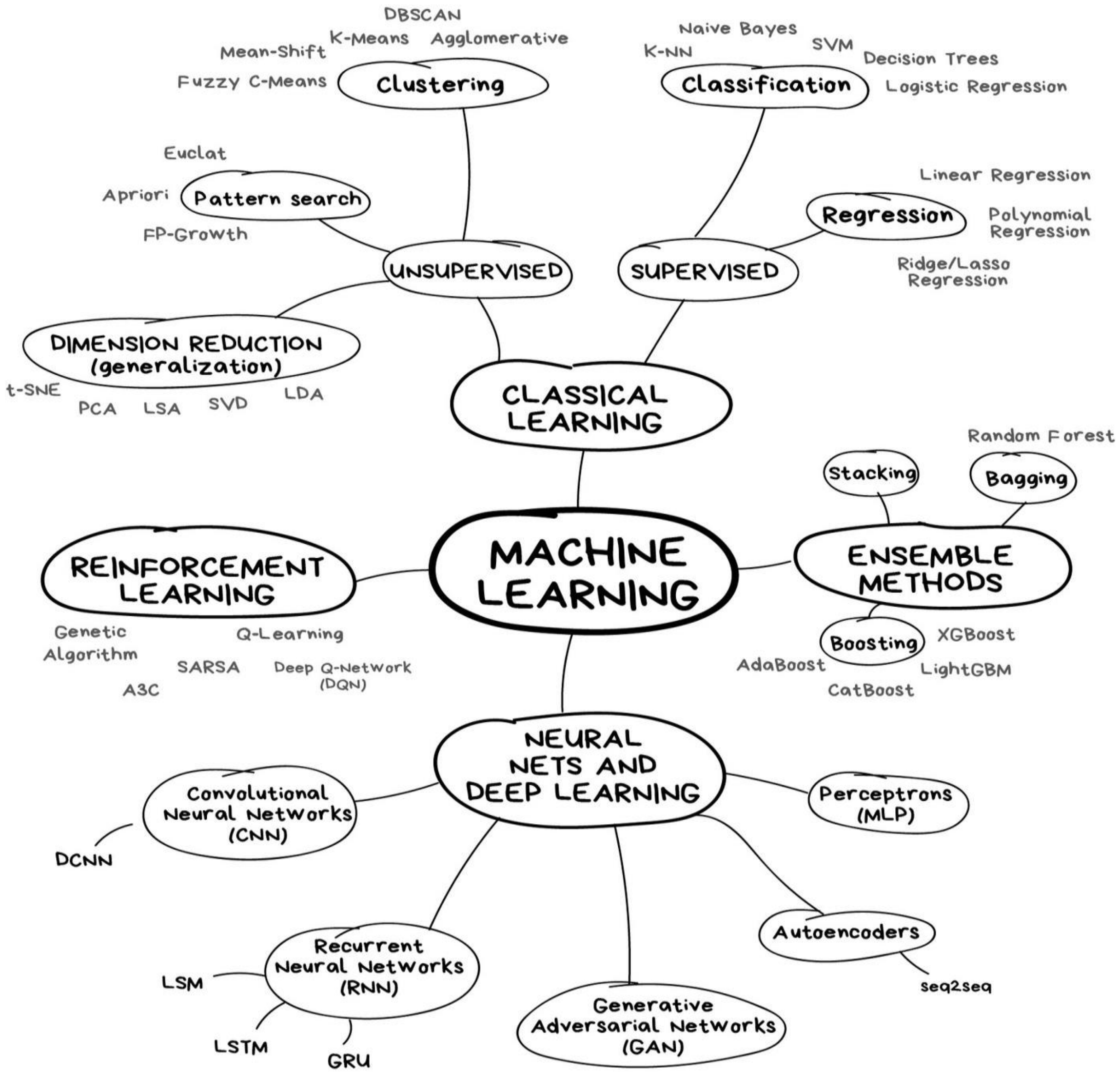
- Occasionally, CRTs are used for lapse rate modelling in literature.
- However, they might be unstable.
- Extension to ***random forests*** or ***gradient boosting***: implementation complexity (Prophet, EARNIX)?

➢ R: `randomForest`



Geschiere (2017)

...



DBSCAN
K-Means
Agglomerative
Mean-Shift
Fuzzy C-Means
Clustering
Naive Bayes
SVM
K-NN
Decision Trees
Classification
Logistic Regression
Euclat
Apriori
Pattern search
FP-Growth
Linear Regression
Regression
Polynomial Regression
Ridge/Lasso Regression
UNSUPERVISED
SUPERVISED
DIMENSION REDUCTION (generalization)
t-SNE
PCA
LSA
SVD
LDA
CLASSICAL LEARNING
Random Forest
Stacking
Bagging
REINFORCEMENT LEARNING
MACHINE LEARNING
ENSEMBLE METHODS
Boosting
XGBoost
AdaBoost
LightGBM
CatBoost
Genetic Algorithm
Q-Learning
SARSA
Deep Q-Network (DQN)
A3C
NEURAL NETS AND DEEP LEARNING
Convolutional Neural Networks (CNN)
DCNN
Perceptrons (MLP)
Autoencoders
seq2seq
Recurrent Neural Networks (RNN)
LSM
LSTM
GRU
Generative Adversarial Networks (GAN)

REFERENCES

REFERENCES

- Dar, A. and Dodds, C. 1989. Interest Rates, the Emergency Fund Hypothesis and Saving through Endowment Policies: Some Empirical Evidence for the U.K. *Journal of Risk and Insurance*, **56**(3): 415–433.
- Geschiere, M. 2017. *Predicting the Lapse Rates of AllSecur*. [link]
- Kiesenbauer, D. 2012. Main Determinants of Lapse in the German Life Insurance Industry. *North American Actuarial Journal*, **16**(1): 52-73
- Kuo, W., Tsai, C., and Chen, W.-K. 2003. An Empirical Study on the Lapse Rate: The Cointegration Approach. *Journal of Risk and Insurance*, **70**(3): 489–508.
- Outreville, J. F. 1990. Whole-life Insurance Lapse Rates and the Emergency Fund Hypothesis. *Insurance: Mathematics and Economics*, **9**(4): 249–255.
- Reacfin. 2016. *Lapse rate models in life insurance and a practical method to foresee Interest Rates dependencies*. White Paper on Life Insurance. Louvain-la-Neuve (BE).

THANK YOU FOR YOUR ATTENTION!

Radek Hendrych
radek.hendrych@csobpoj.cz

ČSOB banka i pojišťovna