LISBON SCHOOL OF ECONOMICS & MANAGEMENT
UNIVERSIDADE DE LISBOA

Master
Actuarial Science

Master's Final Work
Internship Report

Inferring competitiveness without price information: an application to the motor insurance portfolio of Fidelidade

Johannes Meindersma

July 2020

LISBON SCHOOL OF ECONOMICS & MANAGEMENT
UNIVERSIDADE DE LISBOA

Master
Actuarial Science

Master's Final Work
Internship Report

Inferring competitiveness without price information:
an application to the motor insurance portfolio of Fidelidade

Johannes Meindersma

Supervision:
João Manuel de Sousa Andrade e Silva
João Pedro Senhorães Senra Pinto

July 2020

Acknowledgements

The process of writing this thesis was a challenging but rewarding experience. I am indebted to many for their encouragement and support.

I would like to express my gratitude to my supervisor at Fidelidade, João Pedro Senhorães Senra Pinto, for his tireless support and availability during this project.

I would like to pay my special regards to Professor João Andrade e Silva for his supervision during this and other projects.

I would like to express my gratitude to Professor Maria de Lourdes Centeno for recommending me at Fidelidade.

I would like to thank the head of the department, Rui Esteves, for inviting me to the team, as well as for his feedback and advice during my internship.

I thank Fébio Nuno Alves Levezinho for taking the time for our daily meetings, and for his effort to accommodate all my problems.

In addition, I would like to acknowledge my colleagues at Fidelidade for making me feel welcome right from the start.

I would like to thank my fellow graduate students for making my time in Lisbon so worthwhile.

I would like to thank my friends - in particular my dear friend Ana - for their willingness to listen and for helping me overcome frustration.

Finally, I am grateful for the lifetime support of my family, who gave me all opportunities I could have wished for.

Abstract

This work proposes several novel methods for inferring competitiveness of motor insurance policies in a setting of limited availability of price information. State-space functionalities are employed to filter noise from observations by introducing underlying time-dependent structures for transition and conversion data. Transition data about the insurance companies of vehicles in the Portuguese insurance market was collected to analyze the evolution of the incoming transition probabilities of insurers. The binomial hidden Markov model is somewhat restricted due to its assumption of discrete state-space. The Kalman smoother is more successful in removing noise from the observations. The smoother provides intuitive results that are interpretable for a non-technical audience. Furthermore, conversion data was used to infer weekly segment-specific estimates of competitiveness changes. We have proposed a penalized regression framework where time is included as a random walk structure. The model uses credibility weighting on each segment's changes using the full portfolio's changes as the complement. The powerful hierarchical fashion of the model produces estimates of competitiveness changes that are more interpretable than those of generalized linear models, where time is included as a categorical variable. Moreover, the proposed method outperforms the generalized linear models in terms of predictive performance. Both methods can serve as a tool to support the price decision-making process by insurers when the availability of reliable price information is limited.

Keywords: Motor insurance, competitiveness, hidden Markov model, Kalman filter, penalized regression.

Contents

List of Figures

List of Tables

1 Introduction

This report is the result of a curricular internship at the Portuguese insurance company Fidelidade that took place between February and June 2020.

1.1 Background

Traditionally, insurers have performed their ratemaking process using a risk-based approach, where the price of a motor insurance policy is determined solely by the expected costs of the claims that a policyholder will make. Recently, the increasing level of competition in the insurance market has caused a shift towards a more sophisticated system of price determination. Insurers are looking to optimize the policy premiums to find the right balance in the trade-off between profitability and growth of the portfolio. To successfully apply price optimization, it is essential to have a thorough understanding of the market and the behavior of potential customers.

The first building block in optimization is the aforementioned analysis of a customer's risk profile. Furthermore, the expected present value of future profits that a customer generates during his entire future relationship with the insurer should be determined for different premium levels. Finally, a model for the customer's reaction to price changes should be created. In an ideal world, the insurer has perfect information about the individual demand curve of every customer. The demand curve is directly related to the concept of price elasticity of demand, which is a measure used in economics to define the responsiveness of the quantity demanded to a change in the price of that good or service.

Estimating individual demand curves would require a controlled environment which makes it an infeasible venture in practice. Insurers generally strive to determine demand curves for several segments of homogeneous risks, the premise being that customers with similar characteristics will behave similarly when faced with price changes. The most accurate way to gather the required data is by price tests. A price test works by offering a randomly selected sample of customers from a segment a different price and analyzing the conversion rate at the different price levels (Parodi, 2014). Price tests are not meant to test extreme scenarios, but rather focus on premium levels around the current premium. We measure how the demand varies around a region where the demand is roughly linear. The slope of this linear approximation of the demand function is the local elasticity of demand.

1.2 Problem statement

It is not always feasible or desirable to apply price testing for new customers, because of legal constraints or the cost of reputational damage. Often, we have to resort to observational data for the determination of demand functions. When there is a lack of knowledge about the price policy of competitors, determining an accurate demand curve becomes an increasingly difficult task. While working on demand functions of new customers at Fidelidade, we were facing such a severe lack of price information. Therefore, it was challenging to objectively determine the competitiveness of the premium offered to new customers. For an insurer willing to carry out a successful price decision-making process, it is essential to have reliable information about the competitiveness of their offer.

We recognize that the term competitiveness deserves further clarification. In this work, we consider competitiveness to be the attractiveness of the product and premium offered to new customers. We stress that a distinction is made between renewals and new business. Competitiveness is broader than simply the relative price level, although it is certainly a component with a significant impact. Factors such as the attractiveness of a brand and product are also part of the term competitiveness, including marketing strategy and brand loyalty. Furthermore, third-party intermediaries can significantly influence customers' conversion behavior, by steering them towards a specific insurer or product.

We propose two methods for inferring the competitiveness of auto insurance policies in an environment of limited historical and present price information. Both methods attempt to capture changes in competitiveness over time by incorporating state-space model functionalities. Firstly, we track the competitiveness of the main insurers in the Portuguese market using data of transitions between insurers. We use several filtering approaches to remove noise from the data, by introducing an underlying time structure of competitiveness. Secondly, we infer segment-wise competitiveness levels within the full portfolio using conversion data. We present a method where time-series components are added within a penalized regression framework to apply credibility weighting over time and between segments.

1.3 Literature review

Actuaries frequently use generalized linear models to solve a broad range of response modeling problems. However, the model is not equipped to handle time-dependent structures in an appropriate way, which is important when we are working with data that spans multiple periods. Time-dependent behavior is best modeled through state-space models, which are powerful in dealing with time-series because of their flexibility. State-space modeling techniques are used in this work to filter noise from the observed competitiveness level.

There are numerous instances in actuarial literature where state-space models are employed. De Jong and Zehnwirth (1983) were the first to adopt a Kalman filtering approach to smooth development patterns in claims reserving. Similarly, Zehnwirth (1996) uses Kalman filtering to smooth reserving estimates, and Evans and Schmid (2007) filter measurement errors from observed severity and frequency data. De Jong (2005) provides a comprehensive overview of the use of state-space models in actuarial science. He gives an example of an application in mortality modeling, where age-specific log-mortality rates can be smoothed across time.

Korn (2018) proposes a method for incorporating a subset of state-space model functionality into a penalized regression framework. The method describes how to add time-dependent processes such as a random walk process. The author expresses his criticism at the inclusion of a time variable as a categorical variable in a generalized linear model, as is common in actuarial work. When fitting a penalized regression model to such a structure, credibility weighting is performed against the overall mean, ignoring the relationship between consecutive years. On the contrary, by including the time variable as a random walk process, the complement of credibility for each year is the fitted value of the previous year. The author claims that the method is superior in performance compared to the categorical variable approach. Moreover, it produces more intuitive results and it is easier to discover trends that may be affecting particular segments by including an interaction term between the segmentation and the random walk variable. The method is demonstrated by an example involving yearly loss ratios.

The penalized regression methods apply a penalty to the value of the coefficients to reduce the variance of the estimates, accepting a small bias to improve performance and increase interpretability. Ridge regression (Hoerl and Kennard, 1970) and lasso regression (Tibshirani, 1996) are most commonly used. Ridge and lasso penalize based on the squared value and absolute value of the coefficients, respectively. A benefit of lasso regression over ridge regression is that it can aid in variable selection, setting some coefficients equal to zero. On the other hand, lasso regression is not good at handling strongly correlated variables. Ridge regression does not shrink variables to zero and will, therefore, report small changes every period, even if no actual changes have occurred, which greatly reduces the interpretability of the results. Zou and Hastie (2005) introduced the compromising elastic net model, which is a linear combination of the ridge and lasso regression. The elastic net model can perform variable selection and handles correlated variables well, which makes it the preferred choice when dealing with time-series variables.

1.4 Research objective

This work attempts to contribute to the existing literature by providing several methods to infer competitiveness of auto insurance policies in an environment of limited or non-existent price information. We propose two methods using different data sources. Firstly, we observe transition behavior of customers in the Portuguese market and infer quarterly competitiveness levels for the main insurers using state-space model functionalities. Secondly, we use conversion data to infer weekly competitiveness levels of different segments in the portfolio of Fidelidade. We apply credibility weighting among segments and among periods on the conversion data by introducing a penalized regression framework. Both methods are capable of handling time-series effects and stress the relevance of an underlying time structure of competitiveness. The problem is relevant to insurers because it is often costly, time-consuming, or simply infeasible to gather price information of competitors. We propose an alternative solution that can be used to mitigate this loss of information. The methods can help insurers to improve their decision-making process in situations where price information is lacking. There is currently no other literature that has provided a similar methodology to infer competitiveness levels, to the author's knowledge.

1.5 Outline

Section 2 gives an overview of the methodological approach of this report. We discuss the collection and manipulation of different data sources. Subsequently, we explain the main ideas of the methods and models used and their application in this work. Section 3 and 4 describe, explain, and interpret the results of the conducted methods using the transition and conversion data, respectively. In Section 5, we discuss the significance of our findings, as well as the limitations of both methods. In Section 6, we conclude our work and provide several directions for future research.

2 Methodology

This section gives an outline of the research methods that were followed. First, we describe the collection and manipulation of different data sources. We then explain the methodological approach taken for the main models presented in this work. We explain the methodology for the competitiveness inference models that are build using transition data. We formulate the binomial transition framework where we apply a hidden Markov model and Kalman smoothing to remove noise from the observed data. Furthermore, we propose a penalized regression framework that applies credibility weighting to conversion data. We describe an approach for customer segmentation and explain elastic net regression. Again, we provide a formal description of the proposed model. Both problems are answered in the setting of the Portuguese auto insurance market. Finally, we provide a brief overview of the software used to build the models.

2.1 Data collection and manipulation

We collect transition data from a publicly available tool provided by the Portuguese supervisor *Autoridade de Supervisão de Seguros e Fundos de Pensões* (ASF), where drivers can check where a particular license plate is insured in case of an accident. Upon stating a license plate number and date, the tool returns the insurance provider, the policy number, and the start and end date of the policy. We collect a random sample of 12,110 license plates issued in 2005. The query starts on 01/01/2006 and the next query is performed at the date 90 days after the end date of the previous policy. This procedure is continued until all end dates exceed 31/12/2019. The data is manipulated into quarterly categorical data, with one insurer sequence for every unique license plate. Finally, the data is summarized to find the number of transitions from each insurer to another by quarter. The structure of the transition data after manipulation is shown in Figure 1.

	Period 1	Period 2	...	Period T
License plate 1	Insurer X_{11}	Insurer X_{12}	...	Insurer X_{1T}
License plate 2	Insurer X_{21}	Insurer X_{22}	...	Insurer X_{2T}
...	...	...	...	...
License plate n	Insurer X_{n1}	Insurer X_{n2}	...	Insurer X_{nT}

Figure 1: Structure of the transition data after manipulation.

Furthermore, we use conversion data collected from simulations performed by potential customers between 01/01/2018 and 01/10/2019, consisting of ratemaking variables related to the car and the driver. The data is divided into weeks, where Week 1 corresponds to the first week of 2018 and Week 81, the last week, corresponds to Week 29 of 2019. Moreover, information is available on whether or not the customer has converted one of the simulations. In case a customer has converted at least one simulation, additional information is given on the plan that was accepted.

2.2 Inferring competitiveness from transition data

We use state-space models (SSM) to remove noise from the observed time-varying transition probabilities. We outline the hidden Markov model (HMM) following the notation from the well-known tutorial by Rabiner (1989). We consult the textbook by Zucchini et al. (2017) for additional information. Furthermore, we move from discrete to continuous state-space using Kalman's seminal paper on Kalman filtering (Kalman, 1960). Finally, we develop a framework for inferring competitiveness of insurers from transition data. We discuss the application of HMMs and Kalman smoothing with regard to this framework.

2.2.1 Hidden Markov model

Consider a system that can be, at any time, in one of N distinct states, i.e. one state from the set of states $S = \{S_1, ..., S_N\}$. We denote the actual state at time $t = 1, 2, ...$ as q_t. The Markov property states that for a discrete, first-order Markov chain, the probabilistic description of the next state depends only on the current state, i.e.

$$P(\, q_{t+1} = S_j \mid q_t = S_i, q_{t-1} = S_k, ... \,) = P(\, q_{t+1} = S_j \mid q_t = S_i \,)$$

For details on the properties of Markov processes, we refer to Ross (2014) or other textbooks on stochastic modelling.

The Markov model assumes that each state corresponds to an observable event and is therefore too restrictive to be applicable to many real-world problems. The hidden Markov model is an extension where the observation is a probabilistic function of the state. In other words, the state is a stochastic process satisfying the Markov property, that can only be observed through another set of stochastic processes that produce a sequence of observations.

Denote the observation and state at time t as O^t and S^t, and the observation history and state history up to and including time t as $O^{(t)}$ and $S^{(t)}$. Then, the conditional independence property states that

$$P(O^t \mid O^{(t-1)},\ S^{(t)}) = P(\, O^t \mid S^t \,).$$

In other words, the current observation depends only on the current state, and not on the history of observations and sequences. This process is illustrated in Figure 2.

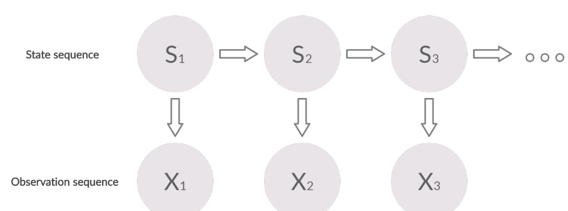

Figure 2: Illustration of a hidden Markov model.

The elements of an HMM are formally defined in the following way:

1. The number of unobservable 'hidden' states N. As in the Markov model, the state space is $S = \{S_1, ..., S_N\}$ and q_t is the state at time t.

2. The set of M observation symbols corresponding to the output of the system, denoted by $V = \{v_1, ..., v_M\}$.

3. The state transition probability distribution $A = \{a_{ij}\}$ where

$$a_{ij} = P(q_{t+1} = S_j \mid q_t = S_i), \quad 1 \leq i, j \leq N$$

4. The observation symbol probability distribution in state j, $B = \{b_j(k)\}$, where

$$b_j(k) = P(v_k \mid q_t = S_j), \quad 1 \leq j \leq N, \; 1 \leq k \leq M.$$

5. The initial state distribution $\pi = \{\pi_i\}$ where

$$\pi_i = P(q_1 = S_i), \quad 1 \leq i \leq N.$$

When all elements are appropriately defined, an HMM can be used as a generator of an observation sequence $O = O_1, O_2, ..., O_T$. Conversely, we can estimate what HMM was most likely to generate an observed sequence. We use the following compact notation for the complete specification of an HMM:

$$\lambda = \{A, B, \pi\},$$

where A is the state transition probability matrix, B the observation symbol probability distribution, and π the initial state distribution.

Three problems of HMMs

HMMs raise three fundamental problems. We briefly explain the problems and give an overview of how a solution can be obtained.

1. Likelihood: Given the observation sequence $O = O_1, O_2, ..., O_T$ and model $\lambda = \{A, B, \pi\}$, how do we efficiently compute the likelihood function $P(O \mid \lambda)$?

The probability of the observation sequence O for the state sequence $Q = q_1, ..., q_T$ is

$$P(O \mid Q, \lambda) = \prod_{t=1}^{T} P(O_t \mid Q, \lambda)$$

where we assume conditional independence of observations. Hence, we get

$$P(O \mid Q, \lambda) = b_{q_1}(O_1) \cdot b_{q_2}(O_2) \cdots b_{q_T}(O_T).$$

The probability of state sequence Q is simply given by

$$P(Q \mid \lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} \cdots a_{q_{T-1} q_T}$$

The joint probability of O and Q is given by

$$P(O, Q \mid \lambda) = P(O \mid Q, \lambda) P(Q, \lambda).$$

By summing over all possible state sequences Q, we obtain the probability of O given by

$$\begin{aligned} P(O \mid \lambda) &= \prod_{\text{all Q}} P(O \mid Q, \lambda) P(Q \mid \lambda) \\ &= \prod_{q_1, \ldots, q_T} \pi_{q_1} b_{q_1}(O_1) a_{q_1 q_2} b_{q_2}(O_2) \cdots a_{q_{T-1} q_T} b_{q_T}(O_T) \end{aligned}$$

The forward-backward procedure is used to calculate the likelihood (Rabiner, 1989). This procedure is very efficient due to its use of dynamic programming.

2. Decoding: Given the observation sequence $O = O_1, O_2, ..., O_T$ and model $\lambda = \{A, B, \pi\}$, how do we find the state sequence $Q = q_1, q_2, ..., q_T$ that describes the observed sequence in the 'optimal' way?

There are several ways to define optimality in this problem, for example by choosing the states that are individually most likely. The Viterbi algorithm (Viterbi, 1967) finds the state sequence Q that is most likely to have generated the observation sequence O. It is a dynamic programming algorithm that maximizes the likelihood function in an efficient way, i.e.:

$$\max_{q_1,\dots,q_T} P(q_1, ..., q_T, O_1, ..., O_T \mid \lambda).$$

3. Learning: How do we choose the parameters $\lambda = \{A, B, \pi\}$ to maximize the likelihood function $P(O \mid \lambda)$?

The most challenging problem of HMMs is to determine a method to adjust the model parameters λ to maximize the probability of the observations sequence O. The Baum-Welch algorithm (Baggenstoss, 2001) is a special case of the expectation–maximization algorithm (Dempster et al., 1977) and performs maximum likelihood estimation of parameters in an HMM using an iterative procedure.

2.2.2 Kalman filtering and smoothing

In this section, we follow the notation from the book of Durbin and Koopman (2012) and use the vignette for the KFAS package in R (Helske, 2017). Unlike the discrete state-space HMM, each hidden state α_t in a linear Gaussian state-space model is modeled as a continuous random variable with a multivariate normal distribution. We denote y_t as the observations and α_t as a vector of latent state processes at time t. The linear Gaussian state-space model with continuous states and discrete time intervals $t = 1, ..., n$ is given by

$$\begin{aligned} y_t &= Z_t \alpha_t + \epsilon_t, \quad \text{(observation equation)} \\ \alpha_{t+1} &= T_t \alpha_t + R_t \eta_t, \quad \text{(state equation)} \end{aligned}$$

where $\epsilon_t \sim N(0, H_t)$, $\eta_t \sim N(0, Q_t)$ and $\alpha_t \sim N(a_1, P_1)$. We assume that y_t, α_{t+1}, and η_t are vectors of length p, m, and k respectively. We denote p, m, and k as the number of observation sequences, the number of states, and the number of disturbances, respectively. The system matrices Z_t, T_t, and R_t, together with the covariance matrices H_t, Q_t, and P_1 depend on the particular model definition. In most cases, these matrices will be time-invariant.

It is possible to use a state-space model for a distribution from the exponential family (Kitagawa, 1987), using a linear Gaussian approximation based on the iterative weighted least squares method (Durbin and Koopman, 2012). The model has the same state equation as in the Gaussian case, but the observation equation has the form

$$p(y_t \mid \theta_t) = p(y_t \mid Z_t \alpha_t),$$

where $\theta_t = Z_t \alpha_t$ is the signal and $p(y_t \mid \theta_t)$ the observational density. The signal θ_t is the linear predictor which is connected to the expected value $E(y_t) = \mu_t$ with a link function $l(\mu_t) = \theta_t$.

The transition data that we use in this work is binomially distributed. For the binomial distribution with size parameter u_t and probability parameter π_t, a logit-link function can be used such that $\theta_t = \text{logit}(\pi_t)$. This results in the following properties:

$$E(y_t \mid \theta_t) = u_t \pi_t$$
$$\text{Var}(y_t \mid \theta_t) = u_t \pi_t (1 - \pi_t).$$

In 1960, R.E. Kalman published his famous paper on the Kalman filter. The Kalman filter is a set of equations that provides an efficient computational solution of the least-squares method. In most cases, the primary goal of state-space modelling is to recover the latent states α given the observations y. Even when the precise model is not completely specified, the Kalman filter can support estimations of past, present, and future states. It is a recursive Bayesian algorithm, updating predictions as new information arrives.

We briefly explain the difference between prediciton, filtering, and smoothing in the context of state-space models. Prediction is an a priori form of estimation of what the state will be at some point in the future. Filtering extracts information about the state at time t, by using data measured up to and including time t. Finally, the smoothed estimate is obtained a posteriori using data measured over the interval $[0, T]$, where $t < T$.

We obtain the one-step-ahead predictions and the prediction errors from the Kalman filtering:

$$a_{t+1} = E(\alpha_t \mid y_t, ..., y_1)$$
$$v_t = y_t - Z_t \alpha_t$$

and the related covariance matrices

$$P_{t+1} = \text{Var}(\alpha_{t+1} \mid y_t, ..., y_1)$$
$$F_t = \text{Var}(v_t) = Z_t P_t Z_t^T + H_t.$$

Then, the state smoothing equations running backwards in time yield

$$\hat{\alpha}_t = E(\alpha_t \mid y_T, ..., y_1)$$
$$V_t = \text{Var}(\alpha_t \mid y_T, ..., y_1).$$

We refer to Durbin and Koopman (2012), Kalman (1960), Kitagawa and Gersch (1996) for details on the algorithm and its technical properties.

2.2.3 Formulation of the proposed framework

We formulate a framework for inferring competitiveness from observed customer transition behavior. The observed transition probabilities are a sequence of noisy estimates of true transition probabilities. We attempt to remove noise from the observations to recover the true time-varying probability parameter. We asses competitiveness on a quarterly basis, because the current collected data is rather limited. We formulate the framework in a general setting.

Consider license plate $k = 1, ..., n$ insured by insurer $i = 1, ..., m$ at the start of period $t = 1, ..., T$. Let n_{ij}^t be the observed number of transitions from insurer i to insurer j for all license plates during period $t = 1, ..., T$. We do not consider any segmentation of license plates.

We define the exposure e_i^t for insurer i during period t as the total number of outgoing transitions from all insurers except from the insurer itself, i.e.

$$e_i^t = \sum_{j \neq i} \sum_{l \neq j} n_{jl}^t.$$

Let x_k^t be the provider of insurance for license plate k at the start of period t. Suppose there exists true probability $p_{i,k}^t$ that license plate k transitioning from an insurer $j \neq i$ during period t, transitions to insurer i, i.e.

$$p_{i,k}^t = P(x_k^t = i \mid x_k^{t-1} = j \wedge x_k^t \neq j), \quad j \neq i.$$

Again, we assume this probability to be the same for all license plates k, i.e.

$$p_{i,k}^t = p_i^t.$$

Let y_i^t be the total observed number of incoming transitions for insurer i during period t, i.e.

$$y_i^t = \sum_{j \neq i} n_{ji}^t$$

which is assumed to be a random draw from a binomial distribution, such that the random variable Y_i^t is as follows:

$$Y_i^t \sim B(e_i^t,\ p_i^t).$$

Observing a single realization of y_i^t, we find the maximum likelihood estimate of p_i^t to be

$$\hat{p}_i^t = \frac{y_i^t}{e_i^t}.$$

In reality, we observe a sequence of transition probabilities over time, instead of observing individual realizations. We expect serial time-dependence between observations, with an underlying time structure in the true transition probability p_i^t. We attempt to capture this using state-space models. Two different methods will be proposed that use the methodologies previously discussed.

Method 1: Binomial-HMM

We propose a binomial-HMM for every insurer i consisting of the following elements:

1. An unobserved state parameter process $\{P_i^t, t = 1, 2, ...\}$ satisfying the Markov property. The state sequence is assumed to be a homogeneous Markov chain, i.e.

$$\pi_{j,k} = P(P_i^{t+1} = k \mid P_i^t = j)$$

   for all $t = 1, ..., T$ and $j, k = 1, ..., N$. Moreover, we assume stationarity such that $\delta_j = P({P_i}^t = j)$ is constant for all t.

2. A state-dependent process $\{Y_i^t, t = 1, 2, ...\}$, where the distribution of Y_i^t depends only on the current state P_i^t.

Suppose that P_i^t has N possible states for every insurer i, i.e. $P_i^t = \{P_1, ..., P_N\}$ for all t. Then, p_i^t is the state representing the true incoming transition probability of insurer i during period t. Again, the state-dependent observation distribution for insurer i at time t is then given by

$$Y_i^t \sim B(e_i^t,\ p_i^t),$$

where e_i^t is the exposure of insurer i during period t.

The state transition probability matrix A and the initial state probabilities π are defined as in the general case. We have assumed stationarity, which means that we effectively impose the constraint $\delta = A'\delta$ on the initial state distribution $\delta = (\delta_1, ..., \delta_N)$. The solutions to the three fundamental problems are explained for the general case in the previous section.

Method 2: Kalman smoothing

We move from a discrete state-space model to a continuous state-space model by introducing a Kalman smoother to estimate the true incoming transition probabilities for all insurers over time. We define the state-space model as follows:

$$\begin{aligned} Y_i^t &\sim B(e_i^t,\ \pi_i^t) \\ \pi_i^t &= \text{logit}(p_i^t) \\ p_i^t &= p_i^{t-1} + \eta_{i,t} \end{aligned}$$

where $\eta_{i,t} \sim N(0, \sigma_{i,t})$, with $\sigma_{i,t}$ to be estimated. In other words, the probability parameter follows a random walk process before the logit transformation. The logit transformation is necessary to ensure that the probability parameter remains between 0 and 1.

We apply Kalman smoothing to the observed transition probability process Y_i^t, that now has continuous state-space, unlike in the Binomial-HMM. We aim to recover the true underlying transition probability ${P_i}^t$ for every insurer i.

2.3 Inferring competitiveness from conversion data

We explain how decision trees can be used to perform customer segmentation. Subsequently, we discuss logistic regression and different penalized regression models. We then formulate a model that adds a time-dependent structure to a penalized regression model to apply credibility weighting. Finally, we discuss how the binary classification models that we build can be evaluated.

2.3.1 Customer segmentation

Customer segmentation is the process of dividing customers into homogeneous groups of similar characteristics, based on their purchasing behavior. In our work, we use a classification tree that finds the most relevant variables to perform segmentation. A classification tree is built by first finding a single variable that bests splits the data in two groups. After the first split, the process is recursively applied separately to each subgroup, until the subgroups either reach a minimum size or until no further improvement can be made. We will later explain how the 'best' split is defined. Decision trees can be applied to many different kinds of data. In this work, all the variables used in the decision tree are categorical.

First, we will briefly introduce some notation regarding decision trees following the vignette from the R package rpart (Therneau and Atkinson, 2019). The sample population consists of n observations from C classes. The conversion variable is binary (a customer either converts or not) and therefore $C = 2$ in our case. The model breaks the observations into k sub-groups and each of the groups has a predicted class assigned.

Let π_i be the prior probability of class $i = 1, ..., C$. We define $L(i, j)$ as the loss matrix for incorrectly identifying a class i observation as a class j observation. Obviously, $L(i, i) = 0$ as the observation is correctly identified. The nodes of the tree are defined as A, where $\tau(A)$ is the class assigned to A, if A were to be taken as a final node. Similarly, $\tau(x)$ is the true class of an observation x, where x is the vector of predictor variables. The number of observations in the sample that are in class i and node A, are n_i and n_A, respectively. Similarly, the number of observations from class i that are in node A is denoted by n_{iA}.

The probability for a future observation to cross node A is given by

$$P(A) = \sum_{i=1}^{C} \pi_i \times P\{x \in A \mid \tau(x) = i\} \approx \sum_{i=1}^{C} \pi_i \times n_{iA}/n_i.$$

It follows that the conditional probability of the true class being i, given that the vector of predictor variables x crosses node A, is given by

$$\begin{aligned} p(i \mid A) &= P\{\tau(x) = i \mid x \in A\} \\ &= \pi_i \times P\{x \in A \mid \tau(x) = i\}/P\{\tau(x)\} \\ &\approx \pi_i \times \frac{n_{iA}/n_i}{\sum_{i=1}^{C} n_{iA}/n_i}. \end{aligned}$$

We define the ‘risk’ of node A as

$$R(A) = \sum_{i=1}^{C} p(i \mid A) L(i, \tau(A))$$

and the risk of the entire tree T as

$$T(A) = \sum_{j=1}^{k} P(A_j) R(A_j),$$

where A_j is the subset of nodes that are terminal.

In case of binary classification, it is common to set $L(i, j) = 1$ for $i \neq j$. Setting the prior probabilities π_i equal to the observed class frequencies in the sample population, we simply have $p(i \mid A) = n_{iA}/n_A$ and $R(T)$ is the proportion incorrectly identified.

Several measures of impurity can be used to determine the best possible splits in a tree. Let f be some impurity function. We define the impurity of a node A as

$$I(A) = \sum_{i=1}^{C} f(p_{iA})$$

where p_{iA} is the proportion of those in A that belong to class i for future samples. A pure node, a node with all observations coming from a single class, should have $I(A) = 0$ and therefore we want $f(0) = f(1) = 0$. The Gini impurity measure was used in this work and is defined as

$$f(p) = p(1 - p).$$

The split with maximal impurity reduction

$$\Delta I = p(A)I(A) - p(A_L)I(A_L) - p(A_R)I(A_R)$$

is chosen, where A_L and A_R are the children nodes of node A.

The question arises how to define the stopping criteria that prevent the model from becoming too complex. There are several control parameters that can be set to avoid overfitting of the model. One of them is the minimum number of observations in a node for which the routine will try to compute a split. Similarly, one can set the minimum number of observations in a terminal node. Generally the most useful parameter is the complexity parameter, which is used to control the size of the decision tree and to select the optimal tree size. The complexity parameter is specified according to the following formula:

$$R_{cp}(T) = R(T) + \text{cp} \times |T| \times R(T_1)$$

where T_1 is the tree with no splits, $|T|$ is the number of splits for a tree, and R is the risk.

The greatest advantage of using a decision tree for segmentation is that the resulting segments are intuitive and easy to explain to a non-technical audience. The decision tree easily finds the most important variables in the data. Furthermore, a decision tree does not require scaling or normalization of the data. Compared to other algorithms, decision trees require less data pre-processing. A disadvantage is that a small change in the data can cause a large change in the structure of the decision tree, causing instability in the resulting segments.

2.3.2 Logistic regression

The generalized linear model (GLM) (Nelder and Wedderburn, 1972) is a generalisation of the classical linear model. GLMs are used to analyse the effect that different covariates or factors have on a response variable, equivalent to the classical linear model. The additional advantage is that non-normal data can now be considered. GLMs have many applications and are used extensively in actuarial work.

A GLM is characterized by the following three components:

1. A dependent variable Y_i whose distribution with parameter θ_i belongs to the exponential dispersion family.
2. A set of independent variables $x_{i1}, ..., x_{im}$ and a linear predictor $\eta_i = \sum_{j=1}^{m} \beta_i x_{ij}$.
3. A linking function $\theta_i = g(\eta_i)$ connecting the parameter θ_i of the distribution of Y_i with linear predictor η_i.

We refer to De Jong and Heller (2008) for details on the properties of GLMs.

Now, consider a binary response variable Y with possible outcomes 0 and 1, where $\mu = P(Y = 1)$ is defined as the probability of success. Hence, $Y \sim$ Bernouilli(μ) and

$$f_Y(y;\mu) = \mu^y(1-\mu)^{(1-y)}$$

is the probability function of Y. The Bernouilli distribution belongs to the exponential dispersion family and can therefore be modelled through a GLM. The choice of the link function should be such that the probability μ is bound between probability 0 and 1. This is satisfied by the logit link function, given by

$$g(\eta) = \ln\left(\frac{\eta}{1-\eta}\right) = \mu.$$

A useful guide for applying logistic regression is provided by Hosmer Jr et al. (2013).

2.3.3 Penalized logistic regression

Regression problems with many potential predictor variables require us to perform statistical model selection to find an optimal model. An optimal model is as simple as possible while still providing good predictive performance. Traditional stepwise selection methods suffer from several drawbacks, such as high variability and low prediction accuracy, especially when there are correlated predictor variables. Penalized regression methods have in recent years shown to be a good alternative, because of their capacity to predict more accurately while being computationally efficient. The loss function of a penalized regression can be viewed as a constrained version of the ordinary least squares (OLS) regression loss function.

We deviate from OLS estimates to increase prediction accuracy. Least squares estimates often have low bias but large variance. In some cases we can improve the prediction accuracy of our model by shrinking some coefficients, or by setting them to zero. We accept a small bias in the estimates in order to reduce the variance of the predicted values. Furthermore, we can build interpretable models even when we have a large number of predictors, by applying a penalty factor. We can determine a smaller subset of variables that exhibit the strongest effect.

We briefly explain two penalized regression methods: the ridge regression and the lasso regression. Then, we introduce the elastic net model, which is a linear combination of the ridge and lasso regression.

Ridge regression

Ridge regression (Hoerl and Kennard, 1970) shrinks the regression coefficients by imposing a penalty on their size. The ridge coefficients minimize a penalized residual sum of squares

$$\widehat{\beta}_{\text{ridge}} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{N} (y_i - g(\eta_i))^2 \quad \text{subject to} \quad \sum_{j=1}^{p} \beta_j^2 \leq t$$

where $g(\eta_i)$ is the prediction for observation i using the link function $g(\cdot)$ and linear predictor η_i. Here, t is a pre-specified free parameter that determines the amount of regularisation and p is the number of covariates.

Equivalently, the ridge estimation problem can be written in the Lagrangian form as

$$\widehat{\beta}_{\text{ridge}} = \underset{\beta}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{N} \left(y_i - g(\eta_i)\right)^2 + \lambda \|\beta\|_2^2$$

where $\|\beta\|_2^2 = \sum_{j=1}^{d} \beta_j^2$ is the square 2-norm of the vector β. The penalty $\lambda\|\beta\|_2^2$ is based on the L_2 norm of the parameter and is therefore called the L_2 penalty.

Lasso regression

Lasso (least absolute shrinkage and selection operator) regression (Tibshirani, 1996) finds the regression coefficient by solving the following minimization problem:

$$\widehat{\beta}_{\text{lasso}} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{N} \left(y_i - g(\eta_i)\right)^2 \quad \text{subject to} \quad \sum_{j=1}^{p} |\beta_j| \leq t,$$

where $g(\eta_i)$ is again the prediction from the GLM. Again, t is a pre-specified free parameter that determines the amount of regularisation and p is the number of covariates. Note that the difference between the ridge constraint and the lasso constraint is within the constraint. Because of the nature of the constraint in the lasso regression, it tends to produce some coefficients that are exactly 0. Therefore, the lasso regressions give more interpretable results.

Once again, we can rewrite the estimation problem in the Lagrangian form as

$$\widehat{\beta}_{\text{lasso}} = \underset{\beta}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{N} \left(y_i - g(\eta_i)\right)^2 + \lambda \|\beta\|_1$$

where $\|\beta\|_1 = \sum_{j=1}^{d} |\beta_j|$ is the 1-norm of the vector β. The penalty $\lambda\|\beta\|_1$ is based on the L_1 norm of the parameter and is therefore called the L_1 penalty.

Elastic net regression

In practice, ridge regression has a higher predictive accuracy than lasso regression. Also, ridge regression does well when there are predictor variables that are highly correlated, by shrinking the grouped variables proportionally. However, a property of ridge regression is that it does not reduce the number of variables, i.e. none of the parameter estimates are zero. Often, it is favorable to eliminate some variables in order to simplify the model. On the other hand, lasso regression uses penalization where some coefficients are actually set to zero. The elastic net regression is a compromise between the ridge and the lasso regression, attempting to combine the advantages of both.

The elastic net regression (Zou and Hastie, 2005) is a linear combination of the ridge and lasso regression. The problem can be written in the Lagrangian form as follows

$$\widehat{\beta}_{\text{elastic net}} = \underset{\beta}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{N} \left(y_i - g(\eta_i)\right)^2 + \lambda(\alpha\|\beta\|_1 + (1-\alpha)\|\beta\|_2^2)$$

where λ is the penalty factor and the parameter α determines whether we use ridge regression($\alpha = 0$), lasso regression ($\alpha = 1$), or a linear combination ($0 < \alpha < 1$).

The package glmnet (Friedman et al., 2010) solves the minimization problem

$$\min_{\beta_0,\beta} \frac{1}{N} \sum_{i=1}^{N} w_i l \left(y_i, \beta_0 + \beta^T x_i\right) + \lambda \left[(1-\alpha)\|\beta\|_2^2/2 + \alpha\|\beta\|_1\right],$$

over a grid of values of tuning parameter λ controlling the strength of the penalty. The negative log-likelihood contribution for observation i is given by $l(y_i, \eta)$. The parameter w_i is the weight of observation i and is normally set to 1 for all observations.

There is a deep link between actuarial credibility and penalized regression, which was demonstrated and proved by Miller (2015). In this work, we decide to only penalize the variables related to time. That is, we penalize the time variable in the random walk process and the interaction of time with the categorical segment variable. The goal of the penalization is to apply credibility weighting to the variables related to time and segment. When we would apply penalty factors to the ratemaking variables, we would distort the value of λ. Most likely, λ would turn out to be significantly lower and in that case, we would not be able to appropriately apply credibility weighting.

Tuning hyperparameters α and λ

We perform k-fold cross-validation using the glmnet package to determine the hyperparameters α and λ that maximize the AUC of the elastic net models. We first run an elastic net model on all the folds, to determine an appropriate sequence of λ's. After that, we determine k different models to run them on each of the folds omitted and calculate the average error over the folds. Since the folds are selected at random, the average error will differ each time. This difference can be reduced by performing repeated cross-validation, where the errors are averaged. To perform cross-validation on α, we supply a grid of values while keeping the folds the same.

2.3.4 Formulation of the proposed model

We show in Table 1 how time can be included as a random walk process by changing the dummy encoding. Table (b) shows how the coefficient for week 2 not only affects that week, but also subsequent weeks. This structure allows for the fitted value of each week to be used as the complement of credibility for the following week. This is called 'credibility weighting' and it can be applied when fitting a penalized regression.

Table 1: Dummy encoding for the week variable.

(a) Dummy encoding for week as a categorical variable.

	week 2	week 3	week 4
week 1	0	0	0
week 2	1	0	0
week 3	0	1	0
week 4	0	0	1

(b) Dummy encoding for a random walk.

	week 2	week 3	week 4
week 1	0	0	0
week 2	1	0	0
week 3	1	1	0
week 4	1	1	1

We formalize the time-dependent structure of the coefficient of the time variable as follows. Let X_t be the true coefficient of the random walk process for the full portfolio at time t, such that

$$X_t = X_{t-1} + \eta_t, \quad X_0 = 0$$

where η_t is the shock in the random walk process for the full portfolio at time t. Similarly, let $X_{t,k}$ be the true coefficient of the random walk process for segment k at time t, such that

$$X_{t,k} = X_{t-1,k} + \eta_t + \eta_{t,k}, \quad X_{0,k} = 0$$

where $\eta_{t,k}$ is the segment-specific shock in the random walk process for segment k at time t.

The model assumes that the full portfolio shock η_t affects all segments equally. The segment-specific shock $\eta_{t,k}$ models changes in competitiveness that are specific to segment k.

Again, the true coefficients X_t and $X_{t,k}$ are observed with some error. Let Y_t be the observed coefficient for the full portfolio at time t, such that

$$Y_t = X_t + \epsilon_t,$$

where error term ϵ_t is the difference between the observed coefficient Y_t and the true coefficient X_t for the full portfolio at time t.

Similarly, let $Y_{t,k}$ be the observed coefficient for segment k at time t, such that

$$Y_{t,k} = X_{t,k} + \epsilon_{t,k},$$

where error term $\epsilon_{t,k}$ is the difference between the observed coefficient $Y_{t,k}$ and the true coefficient $X_{t,k}$ for segment k at time t.

The goal of this framework is to infer η_t and $\eta_{t,k}$ by fitting an elastic net model that penalizes the coefficients of the shock terms. The modified dummy encoding makes it easier to model not only changes of the full portfolio, but also the changes by segment, by including an interaction term between the random walk structure and the segmentation. This produces a model that credibility weights segment changes using the full portfolio's changes as a complement. This hierarchical structure should better capture the underlying structure of changes in the competitiveness level.

2.3.5 Evaluation of binary classification models

The proposed models classify the decision of each customer as either 'yes' or 'no', based on the probability of conversion predicted by the model. We introduce two metrics that are used in this work to evaluate the out-of-sample predictive performance of the different binary classifiers. The accuracy rate is simply the sum of the number of true positives and true negatives divided by the total number of observations. The accuracy rate is not a perfect measure when there is a class imbalance, where one of the classes is more prevalent in the data. In our case, the class imbalance is not severe and therefore it is appropriate to consider the accuracy rate. However, there is sufficient reason to introduce another metric that is even more appropriate.

An ROC curve can be created by plotting the true positive rate (sensitivity) against the false negative rate (1-specificity) at various threshold levels. The threshold level determines the cutting point for positive predictions. After plotting, the area under the ROC curve (AUC) can be computed, with a higher value indicating better performance. The AUC can be used to assess the discriminatory capacity of an individual model. The AUC takes values between 0 and 1, where a perfect predictor gets 1 and a perfectly incorrect predictor gets 0. A realistic model will generally score between 0.5 and 1, since a randomly guessing predictor would score exactly 0.5. The AUC metric is threshold-invariant, since it measures the quality of predictions irrespective of what classification threshold is chosen.

2.4 Software

Throughout our analysis, we have used R, a free software environment for statistical computing and graphics. The following packages proved to be particularly helpful and deserve special recognition:

- Multiple tidyverse packages: tidyr (Wickham and Henry, 2020) for tidying the data, dplyr (Wickham et al., 2020) for data manipulation, and ggplot2 (Wickham, 2016) for creating graphics.
- HiddenMarkov (Harte, 2017) for analysis of discrete-time hidden Markov models. It includes functions for simulation, parameter estimation, and the Viterbi algorithm.
- TraMineR (Gabadinho et al., 2011) for mining, describing and visualizing sequences of states, and more generally discrete sequence data.
- KFAS (Helske, 2017) for Kalman filtering and smoothing for exponential family state-space models.
- rpart (Therneau and Atkinson, 2019) for constructing decision trees used for customer segmentation.
- glmnet (Friedman et al., 2010) for fitting lasso and elastic net regularized generalized linear models.
- caret (Kuhn, 2020) for training and plotting classification models.

Furthermore, Python was used for the collection of transition data.

3 Inferring competitiveness from transition data

In this section, we show the results of the different methods we have proposed to infer quarterly competitiveness levels of insurance policies offered by insurers active in the Portuguese market. Both methods attempt to recover the true incoming transition probability, by applying a state-stace modelling approach. The hidden Markov model assumes a discrete state-space, while the Kalman filter is a continuous state-space method. We explain the reasoning and intuition behind both approaches and compare the results. We limit our analysis to the insurers in the Portuguese insurance market that had the highest number of incoming transitions in the period 2007-2020. The names of the insurers are anonymized for confidentiality purposes by renaming them to insurer A, B, C, D, E, and F.

3.1 Descriptive summary of observed transitions

As discussed in Section 2.2.3, we are interested in the true incoming transition probability p_i^t as we aim to infer the competitiveness level of insurers for policies offered to new customers. First, we visualize and discuss the observed transition data to find the observed incoming transition probability $\hat{p}_i^t$. Figure 3 shows the number of incoming transitions on the left side and the number of outgoing transitions on the right side. We show the outgoing transitions for all insurers over time to determine the exposure e_i^t. The exposure of an insurer is given by all outgoing transitions minus the outgoing transitions from the insurers itself. Similarly, we show the incoming transitions y_i^t for insurer i in period t, which are used to calculate $\hat{p}_i^t$. We observe in Figure 3 that for Fidelidade, the outgoing transitions outnumber the incoming transitions for almost the entire observation period, although the effect flattens out in most recent years.

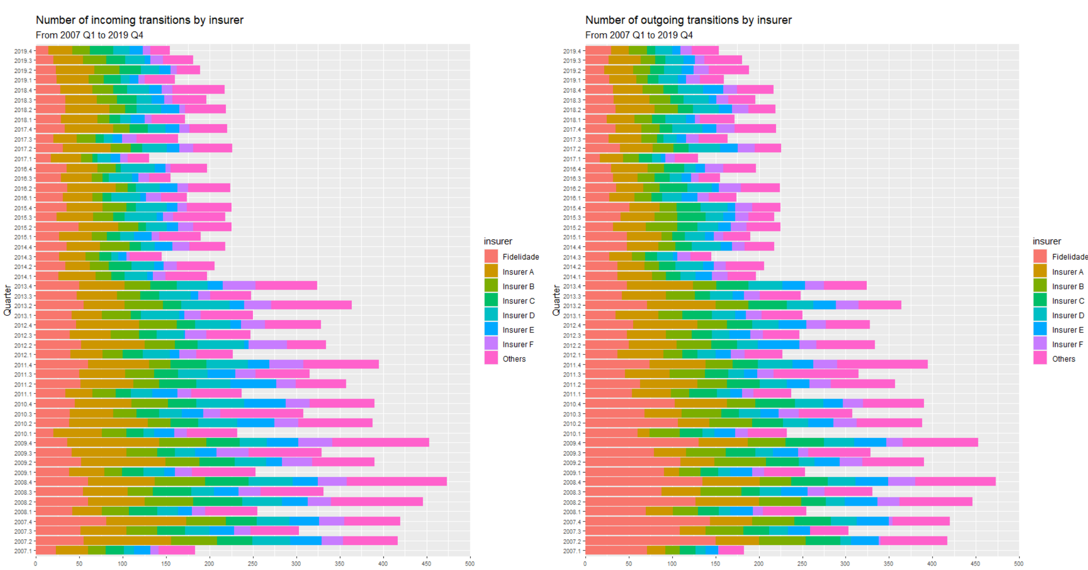

Figure 3: Observed quarterly incoming and outgoing transitions.

Several anomalies are observed in the pattern of incoming and outgoing transitions. Firstly, there is a visible seasonality effect, such that the incoming and outgoing transitions are higher in the second and fourth quarter. This is because the collected license plates are from the second and fourth quarter of 2005 and car owners are more likely to change insurer after a full year of service. Furthermore, we observe that both the number of incoming and outgoing transitions is diminishing over time. This is caused by the fact that part of the license plates does not insure itself anymore after a sufficient number of years, simply because the car is no longer on the road or because the data is missing in the tool.

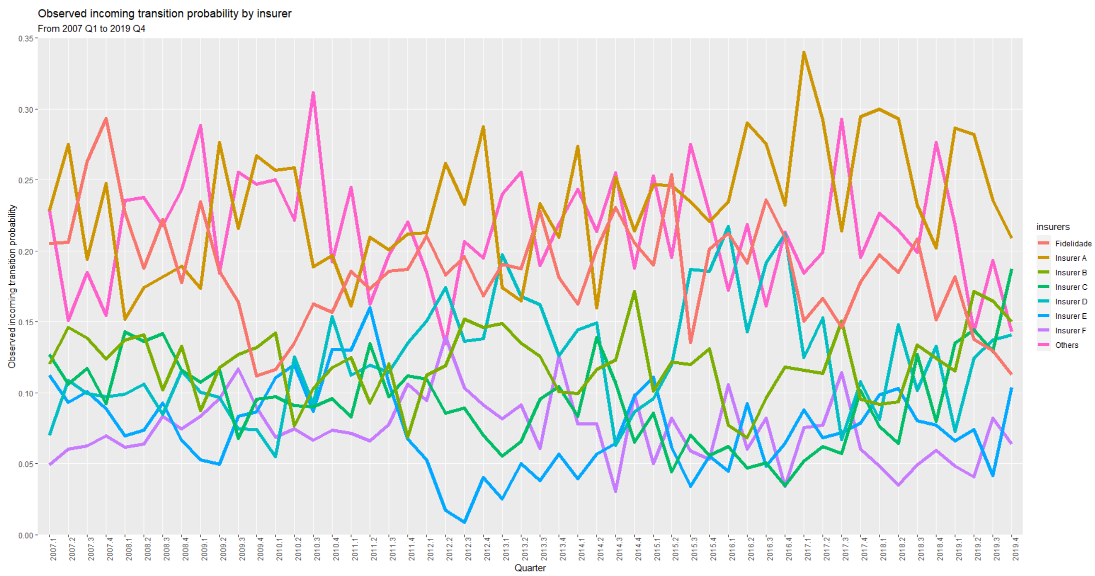

Figure 4: Observed quarterly incoming transition probabilities.

Finally, Figure 4 visualizes the evolution of the observed transition probability $\hat{p}_i^t$ by insurer. Note that the observed transition probabilities are rather volatile, even for adjacent periods. A considerable proportion of this volatility can be attributed to noise in the estimate of the true probability p_i^t. Obviously, it can be challenging to compare the competitiveness of insurers by this figure alone. Therefore, the remaining analysis is dedicated to removing part of the noise from the observed incoming transition probability.

3.2 Model 1: Binomial-HMM

Firstly, we fit a binomial-HMM with $N = 5$ states for all insurers i. This means that we assume that the true incoming transition probability p_i^t can take on five values that differ between insurers. Therefore, this method is not suitable to compare competitiveness between insurers, but rather to see how competitiveness of an individual insurer changes relative to the changes of competitors. The chosen number of states is arbitrary, but has a clear intuition. The middle states 2, 3, and 4 represent a state of low, medium, and high competitiveness. The extreme states 1 and 5 stand for very low and very high competitiveness and are added to capture outliers in the observed transition probability.

Table 2 shows the incoming transition probability levels corresponding to the 5 states of competitiveness for all insurers. The Baum-Welch algorithm is used to find the unknown parameters $\lambda = \{A, B, \pi\}$ of the HMMs. The algorithm finds the parameters that best explain the observed sequence, and therefore the parameters will be different for another set of realizations from the same underlying process. It is important to point out that the Baum-Welch algorithm

Table 2: Incoming transition probabilities corresponding to the 5 states of competitiveness.

Insurer	Very low	Low	Medium	High	Very high
Fidelidade	0.132	0.163	0.190	0.239	0.292
Insurer A	0.178	0.196	0.232	0.253	0.262
Insurer B	0.080	0.103	0.119	0.125	0.142
Insurer C	0.055	0.074	0.101	0.117	0.139
Insurer D	0.073	0.099	0.120	0.154	0.183
Insurer E	0.014	0.048	0.062	0.082	0.122
Insurer F	0.055	0.065	0.073	0.085	0.108
Other insurers	0.153	0.196	0.209	0.229	0.281

is sensitive to initial conditions. We note that the final probabilities corresponding to the five different levels of competitiveness are indeed inconsistent when changing the starting values. Therefore, it can be worthwhile to experiment with these values and to repeat the procedure multiple times. We have aimed to maximize the distance between different transition probability levels, to increase the intuition behind a transition from one state to another. We have done so by looping over a grid of starting conditions, and subsequently choosing the conditions that maximize the minimal distance between two adjacent states of competitiveness.

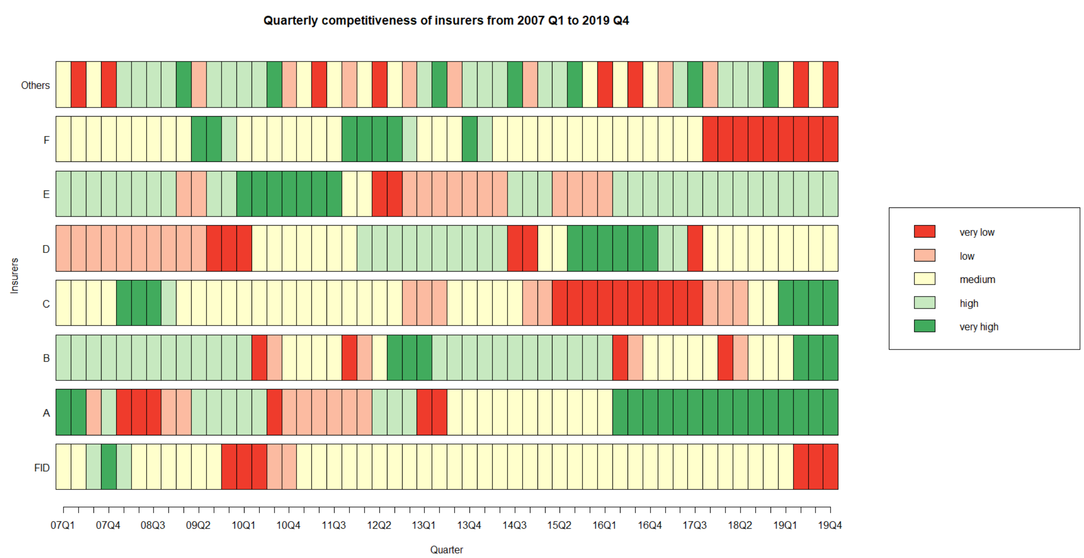

Figure 5: Most likely sequence of the quarterly states of competitiveness.

Figure 5 visualizes the most likely sequence of the quarterly states of competitiveness for all insurers as obtained from the Viterbi algorithm. The corresponding competitiveness levels are as shown in Table 2. A transition from one state to another is indicated by a change in color. For example, we observe that Fidelidade ('FID') transitions first from medium to high and then from high to very high competitiveness in the span of two periods, from 2007 Q2 to 2007 Q4. Similarly, their competitiveness worsens from medium to very low competitiveness in the second quarter of 2019. At and around that time, we observe that insurer B and C see an increase in competitiveness from high to very high. It is possibly that those insurers changed their price policies in this time frame, which in due course affected the incoming transition probability of Fidelidade. We observe that the transition sequence of ‘others’ is relatively chaotic, because this sequence is directly affected by all changes in price policies of the insurers not included in main ones. In general, the figure does well to point out long-term evolution of competitiveness of insurers. Moreover, the figure can be used to speculate about changes in price policies and its effect on competitors.

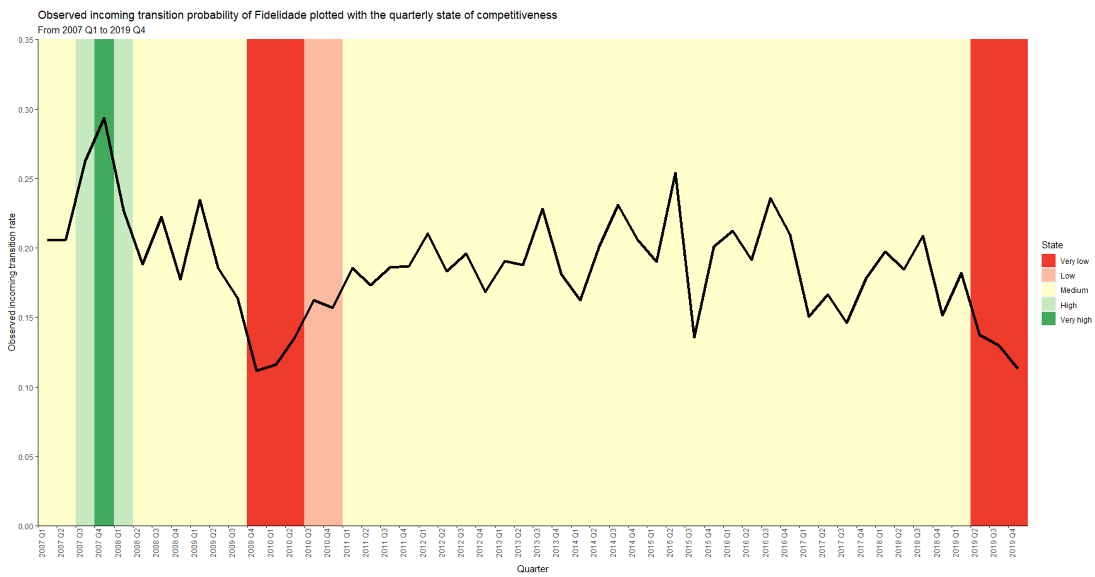

Figure 6: Observed incoming transition probability of Fidelidade with the state of competitiveness on the background.

Figure 6 projects the most likely sequence of hidden states for Fidelidade on the background. The observed incoming transition probability is plotted to show which observations are generated by the true transition probabilities, as given by the results from the Viterbi algorithm. For example, the entire period between the first quarter of 2011 and the second quarter of 2019 is supposedly generated from the probability corresponding to medium competitiveness, which is 0.190 as can be seen in Table 2. As previously mentioned, the model is helpful to observe the long-term evolution of competitiveness, by removing part of the noise from observations. However, if we are interested in observing smaller changes in competitiveness, we should build a model that extends the discrete state-space to continuous state-space.

3.3 Model 2: Kalman smoother

We extend the discrete state-space binomial-HMM by proposing a Kalman smoother, which assumes continuous state-space. In the previous model, the true incoming transition probability could take on five values that were different for every insurer. Now, the true transition probability can take on any value between 0 and 1. We apply Kalman smoothing to the observed transition rates of all insurers using the following state-space model:

$$
\begin{aligned}
Y^t &\sim B(e^t,\ \pi^t) \\
\pi_t &= \mathrm{logit}(p_t) \\
p^t &= p^{t-1} + \eta_t
\end{aligned}
$$

where $\eta_t \sim N(0, \sigma_t)$ and σ_t is unknown. We refer to Section 2.2.3 for details on the methodology.

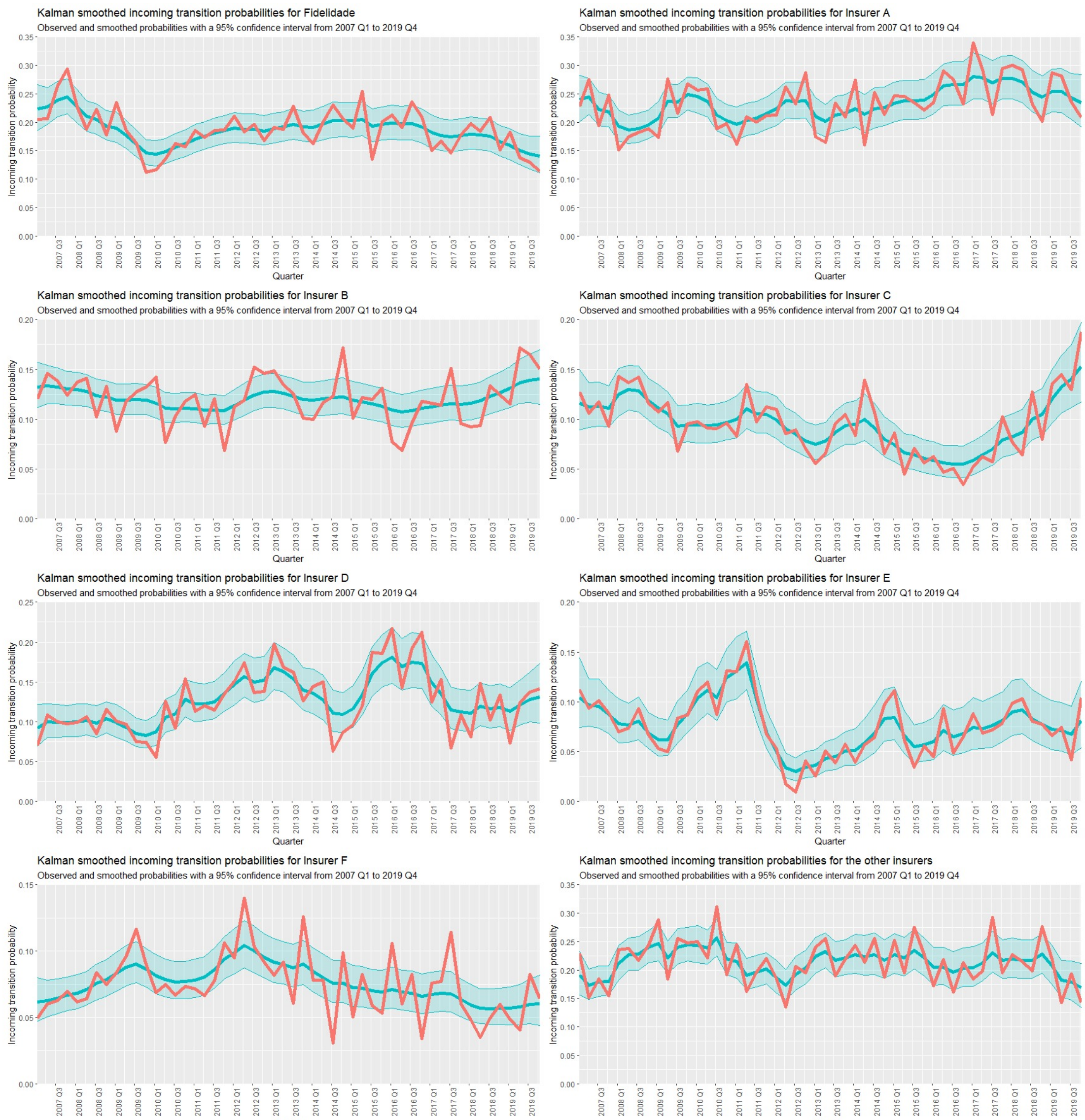

Figure 7: Comparison of the Kalman smoothed transition probabilities with the observed transition probabilities.

Figure 7 shows for all insurers the Kalman smoothed transition probability and its 95% confidence interval (blue line and area), together with the observed transition probability sequence (red line). The Kalman smoother indeed manages to smoothen the observed probabilities and gives a more interpretable and realistic sequence of estimated true coefficients. We stress that the 95% confidence interval is under the assumption that the true incoming transition probability follows a random walk process, as stated by the model. The random walk process assumes normally distributed shocks, which could be somewhat unrealistic. This can be one of the reasons that there are many instances where the observed probability falls outside the confidence interval of the Kalman smoother. However, it is is in no way a testament to the failure of the Kalman smoother. On the contrary, it underlines that the Kalman smoother is doing well at removing noise from the observed sequence of probabilities.

An example can be seen in the figure of insurer B, where the observed probability goes back-and-forth to be above and below the Kalman smoothed probability. In this case, the Kalman smoothed probability removes noise by being more stable than the observed rate. On the other hand, we see in the figure of insurer E that the Kalman smoother adapts quickly in case of a big shock in competitiveness, when the observed probability does not return to its previous level afterwards.

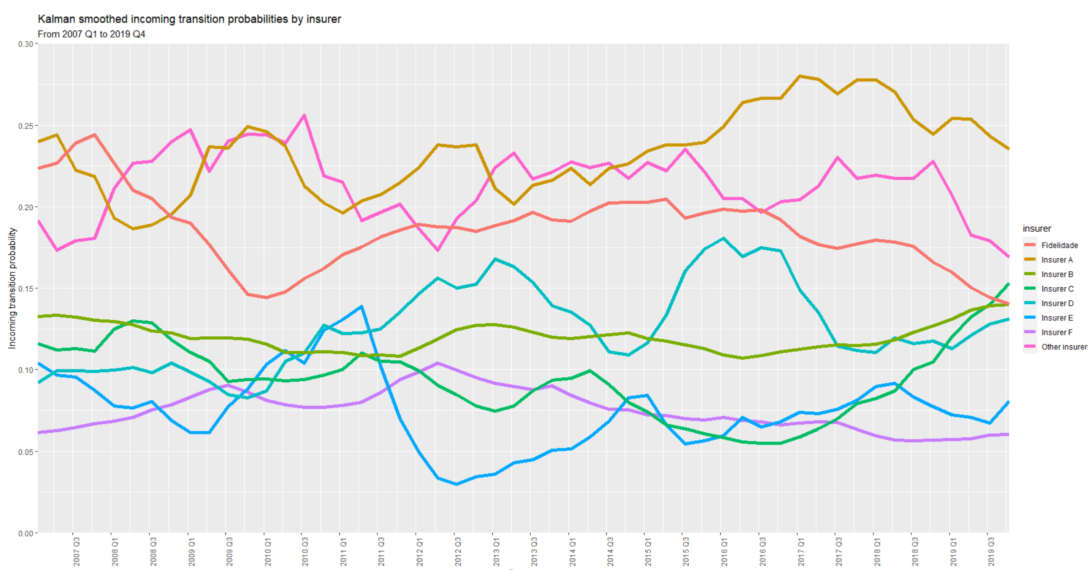

Figure 8: Kalman smoothed incoming transition probabilities.

Figure 8 shows the evolution of the Kalman smoothed transition probability for the major insurers in the Portuguese car insurance market. The method is helpful to compare long-term trends of competitiveness changes between insurers. For example, we observe that the gap in competitiveness between Fidelidade and insurer B has been increasing in recent years, in favor of insurer B. In the same period, insurer C has increased its competitiveness in a remarkable way. The Kalman smoother provides the opportunity to create better insight in how changes in competitiveness of one insurer affect another. Those insights can be incorporated into the price decision-making process.

4 Inferring competitiveness from conversion data

In this section, we propose a new method to infer weekly competitiveness levels of insurers. We use conversion data from Fidelidade's motor insurance portfolio gathered between January 2018 and September 2019. We add a time variable as a random walk process to capture the changes in competitiveness over time. By fitting an elastic net model, we aim to apply credibility weighting between periods and between segments and the full portfolio. We compare the out-of-sample predictive performance and interpretability of the new model with the more common generalized linear model where time is included as a categorical variable.

4.1 Customer segmentation

We perform customer segmentation to divide customers into homogeneous groups of similar conversion behavior. For details on the methodology, we refer to Section 2.3.1. A brief overview of the variables considered in the analysis of conversion data is provided in Table 5 in appendix A. We fit a decision tree model for customer segmentation using the following formula:

$$\begin{aligned} \text{convert} \sim\ & \text{brand} + \text{engine} + \text{cylinder} + \text{power} + \text{weight} + \text{RPP} + \text{weight2} + \text{value\_vehicle} + \\ & \text{tariff\_zone2} + \text{age\_license2} + \text{age\_vehicle2} + \text{birthyear} + \text{intermediary} + \text{bonusmalus}. \end{aligned}$$

We stress that we do not aim to maximize predictive performance, in which case the idea would be to minimize the cross-validated error and select the associated complexity parameter. Rather, we would like to segment the customers in such a way that the segmentation makes practical sense from a business perspective, by creating segments that are easily distinguishable by a few simple rules. Too many segments would diminish the interpretability of the segmentation, while too few segments would not do justice to the diversity of the portfolio. For these reasons, we decide not to use a mathematically objective measure for determining the number of clusters. We set the minimum number of customers in a segment to be 2% of the total number of customers in the portfolio to avoid creating segments that are very small. Moreover, we use a complexity parameter of 0.002 to avoid creating segments that are too similar to each other.

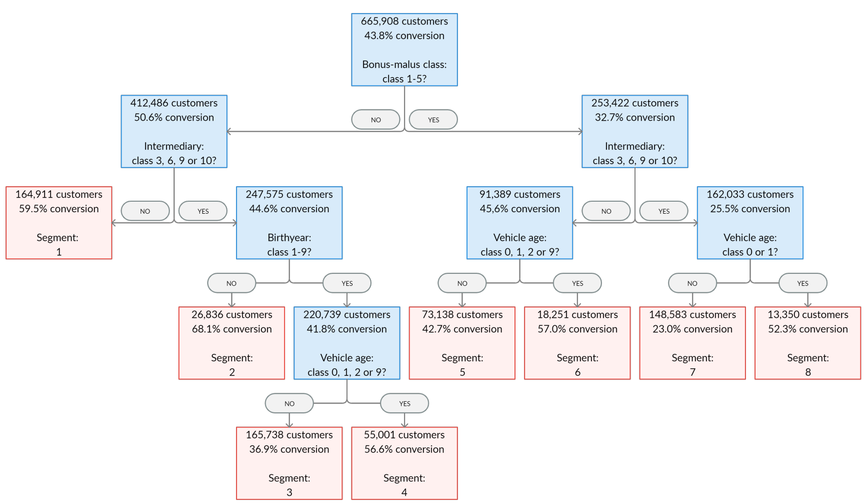

Figure 9: Visualization of the decision tree used for segmentation.

The resulting decision tree is shown in Figure 9. We note that some sensitive information about the factor levels is left out in this version of the report. The blue nodes represent the decision nodes where a decision rule is applied to split the customers of the node into two parts. The red nodes represent the leaf nodes, which do not have any additional nodes coming off them. Each leaf node represents one segment that covers a different set of customers. The first division is based on the bonus-malus class of the customer. Bad drivers have a lower conversion rate, because they are offered a price that is less competitive. After the first split, both nodes are split by intermediary, who play an important role in the conversion behavior of customers. The good drivers with bad intermediary are further divided by birthyear. Young drivers turn out to have a higher conversion rate than older drivers. Finally, the older drivers are split by the age of their vehicle. The data shows that there is a quadratic relationship between conversion rate and vehicle age. That is, the conversion rate is relatively high for very young and very old cars. On the other side of the tree, we see that the bad drivers are separated by the age of their vehicles. The final decision tree consist of 8 segments.

4.2 Evaluation and comparison of the different models

The following models will be compared based on out-of-sample predictive performance and interpretability:

1. A generalized linear model without time structure (**Model 1**).
2. A generalized linear model with time included as a categorical variable (**Model 2**).
3. A generalized linear model with time included as a categorical variable, interacting with the segment variable (**Model 3**).
4. An elastic net model with time included as a random walk process (**Model 4**).
5. An elastic net model with time included as a random walk process, interacting with the segment variable (**Model 5**).

Model 1 is the baseline model that does not include any time structure whatsoever. Model 2 introduces time as a categorical variable, which is the most common way to include time in actuarial work. Model 3 adds an interaction effect between time and the segment variable to Model 2, to capture segment-specific competitiveness changes over time. All of these models do not assume a specific time-dependent structure for the weekly competitiveness levels. Model 4 is the first elastic net model and includes time as a random walk process, so that we can apply credibility weighting among time periods. Model 5 adds an interaction effect between time and the segment variable to Model 4. This interaction effect enables us to also apply credibility weighting among segments and the full portfolio.

We stress that all models include the basic ratemaking variables and the segment variable. We choose to include these variables consistently, because they are recognized to have an impact on the conversion behavior of customers. Moreover, we want to ensure that the effects of ratemaking variables are captured to avoid disturbing the weekly competitiveness level. Similarly, we choose to include a segment variable in all models to capture the effect of relevant ratemaking variables that are currently not included. Evidently, the segment variable also partly captures the competitiveness level of a segment. However, the variable has the same impact on all time periods and therefore we are still able to infer weekly changes in the competitiveness level.

Table 3 shows the performance of the models in terms of area under the ROC curve (AUC) and accuracy rate of predictions. Note that the first three models have $\lambda = 0$ and irrelevant α, since no penalty is applied for generalized linear models. For the elastic net models, we have performed 10-fold cross-validation repeated 3 times to determine the hyperparameters α and λ that maximize the AUC, as explained in Section 2.3.3. We have used a grid for α such that $\alpha \in \{0, 0.1, 0.2, ..., 1\}$ is chosen. We find that $\alpha = 0.3$ maximizes the AUC for Model 4 and $\alpha = 0.2$ for Model 5.

Table 3: Summary of the predictive performance of the proposed models.

Model	λ	α	AUC	Accuracy
1	0	-	0.7210530	0.6618007
2	0	-	0.7226328	0.6630194
3	0	-	0.7224989	0.6628840
4	0.0004263	0.3	0.7226877	0.6630258
5	0.0008470	0.2	0.7232174	0.6636384

We observe that Model 2 outperforms the other generalized linear models. Apparently, adding time as a categorical variable increases the predictive performance compared to no time structure at all. However, the interaction effect between time and segment in Model 3 seems to cause overfitting. This additional term cannot adequately capture the underlying structure of the data, when used in a traditional setting.

Both elastic net models have a higher predictive performance than all the generalized linear models. The random walk process better captures the underlying time structure than simply adding time as a categorical variable. We observe that Model 5 outperforms Model 4 in terms of AUC and accuracy rate. The difference between those models is that we have added a penalty applied to the coefficients of the interaction effect, avoiding the possibility of overfitting.

We acknowledge that the inclusion of a time structure only slightly increases the predictive power of the model. However, we remind the reader that the aim is not to build a model that improves predictions, but rather to improve the interpretability of the coefficients of weekly competitiveness changes to aid decision-making. The remaining analysis shows how time structure is indeed able to significantly improve interpretability.

Figure 10 compares the evolution of competitiveness of the full portfolio as inferred from the coefficients of the time variables in Model 2 and 5. For Model 2, the figure simply shows the coefficients of the categorical variables at time t. For Model 5, we show the cumulative shocks up to week t. We observe that both models show a similar evolution of competitiveness, as expected. Model 2 produces estimates that can vary greatly in a short period of time. Moreover, the direction of the changes often alternates in subsequent weeks such that no clear short-term trend is visible. Those alterations are likely to be a result of noise in the estimates. On the other hand, Model 5 shows a much more smooth evolution of competitiveness due to the inclusion of a time trend in the model. The application of credibility weighting has removed part of the noise that caused the estimates of Model 2 to be disturbed. The estimates of Model 5 are more interpretable and reliable. Therefore, the decision-making process can be improved by using this method.

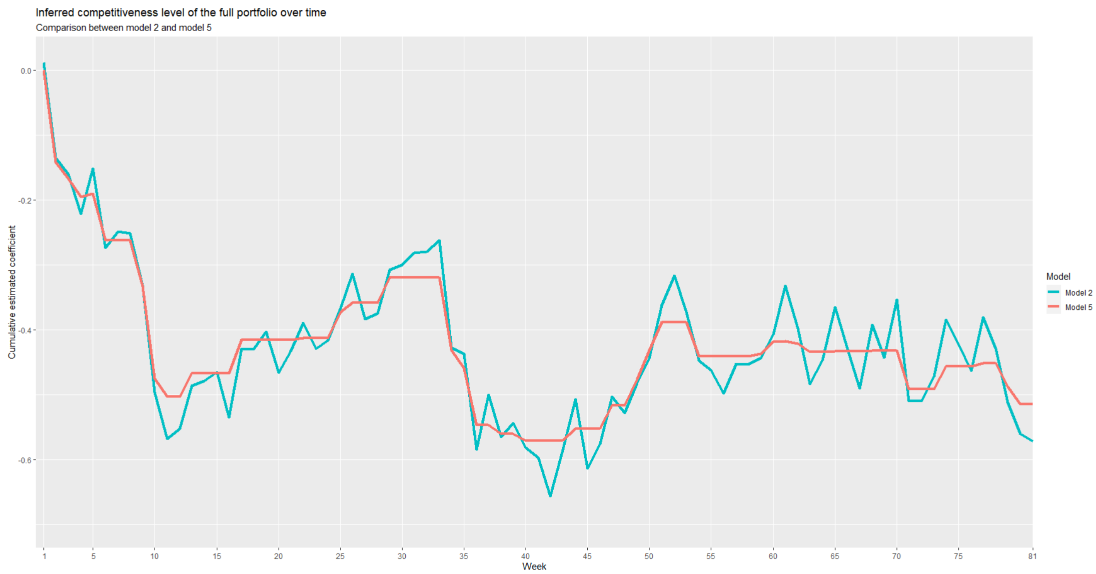

Figure 10: Comparison between Model 2 and Model 5 regarding the estimates of competitiveness for the full portfolio.

We briefly discuss the interpretation of the sign and magnitude of the coefficients. The coefficients of a (penalized) logistic regression are in terms of the log-odds. For example, a customer that has a 40% probability of converting in week t, increases to a 42.42% probability of converting in week $t+1$ after an increase of the competitiveness level coefficient by 0.1. On the other hand, if the competitiveness level coefficient decreases by 0.1, the probability of converting drops to 37.63% for the same customer. In fact, we observe several instances in Figure 10 where Model 5 indicates a weekly change in competitiveness greater than 0.1.

Figure 11 compares the evolution of segment-wise competitiveness as inferred from the coefficients of the time variables in Model 3 and 5. The segment-wise competitiveness level is determined by the competitiveness of the full portfolio plus the sum of segment-wise shocks in previous periods. The results are similar to those seen before. The estimates for Model 3 are irregular and unrealistic, since time is included as a categorical variable without a time structure. On the other hand, model 5 gives estimates that are smoother and more realistic, because time is included as a random walk process and as an interaction effect with the segment variable.

The contrast between Model 3 and Model 5 is more prevalent in the segment-wise competitiveness levels, because there is less data available for an individual segment. For segments with relatively many customers, the estimated coefficient tends to be closer to the coefficients in Model 3, because the observed coefficients are more reliable in these segments. Table 4 indeed shows that for bigger segments, the correlation between the estimated coefficients of Model 3 and Model 5 is higher. On the other hand, Model 5 considers the deviations in smaller segments mostly as noise that arises from a lack of data, and therefore we observe a lower correlation in these segments. This is the advantageous effect of applying credibility weighting between segments and the full portfolio.

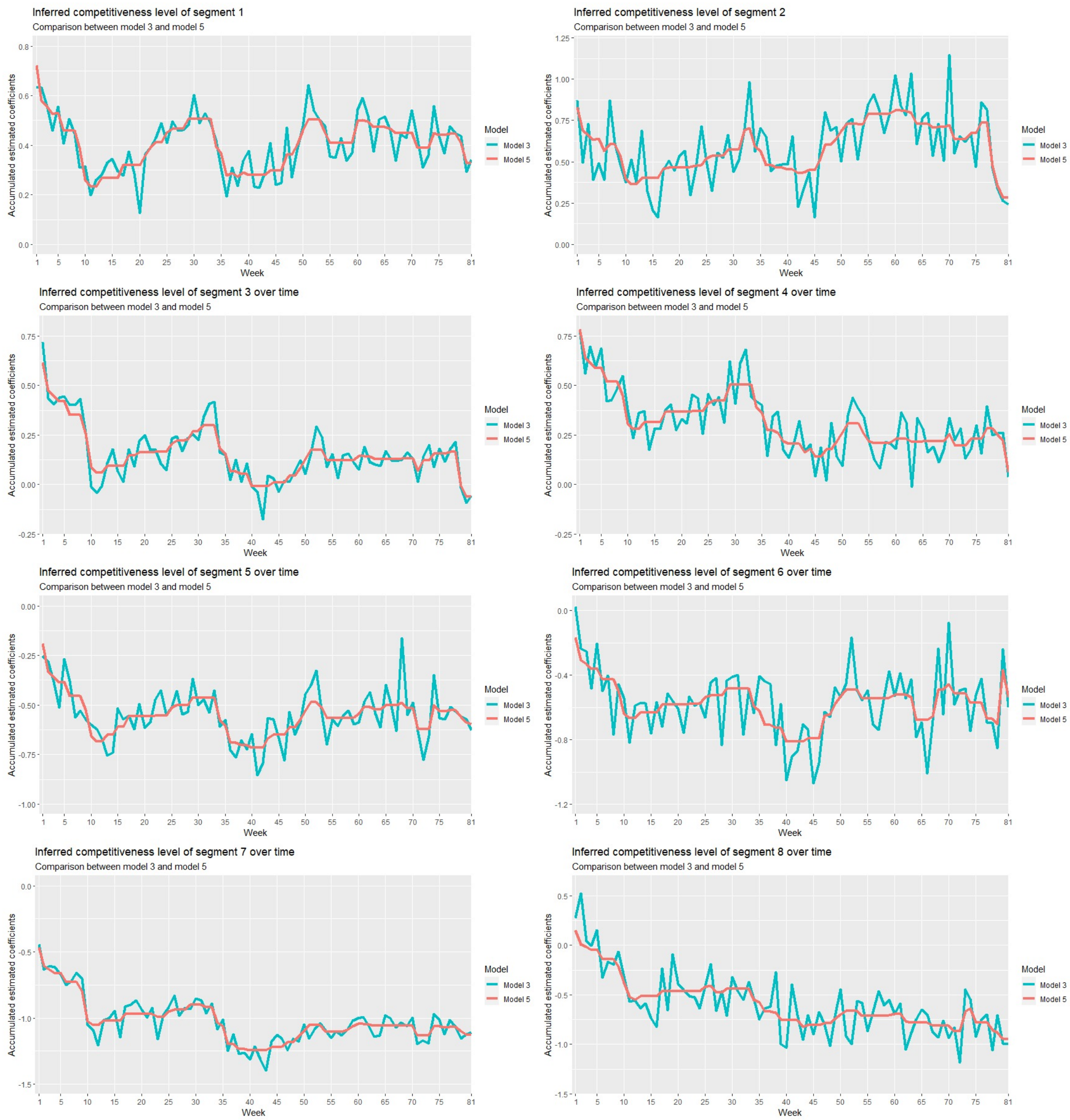

Figure 11: Comparison between Model 3 and Model 5 regarding the estimates of segment-wise competitiveness.

We have previously seen that segments experience segment-specific shocks in their competitiveness level, besides the full portfolio shocks that are common for all segments. Figure 12 shows these segment-specific changes in competitiveness inferred by Model 5. These shocks can happen for example when an insurance company changes their prices for a specific segment, or when a specific group of people is targeted by a marketing campaign. Naturally, a segment will in most periods not experience a segment-specific shock. We indeed observe that the majority of the coefficients are set to zero by the penalty factor. This is an advantage of elastic net regression over ridge regression, where coefficients are never shrunk to zero.

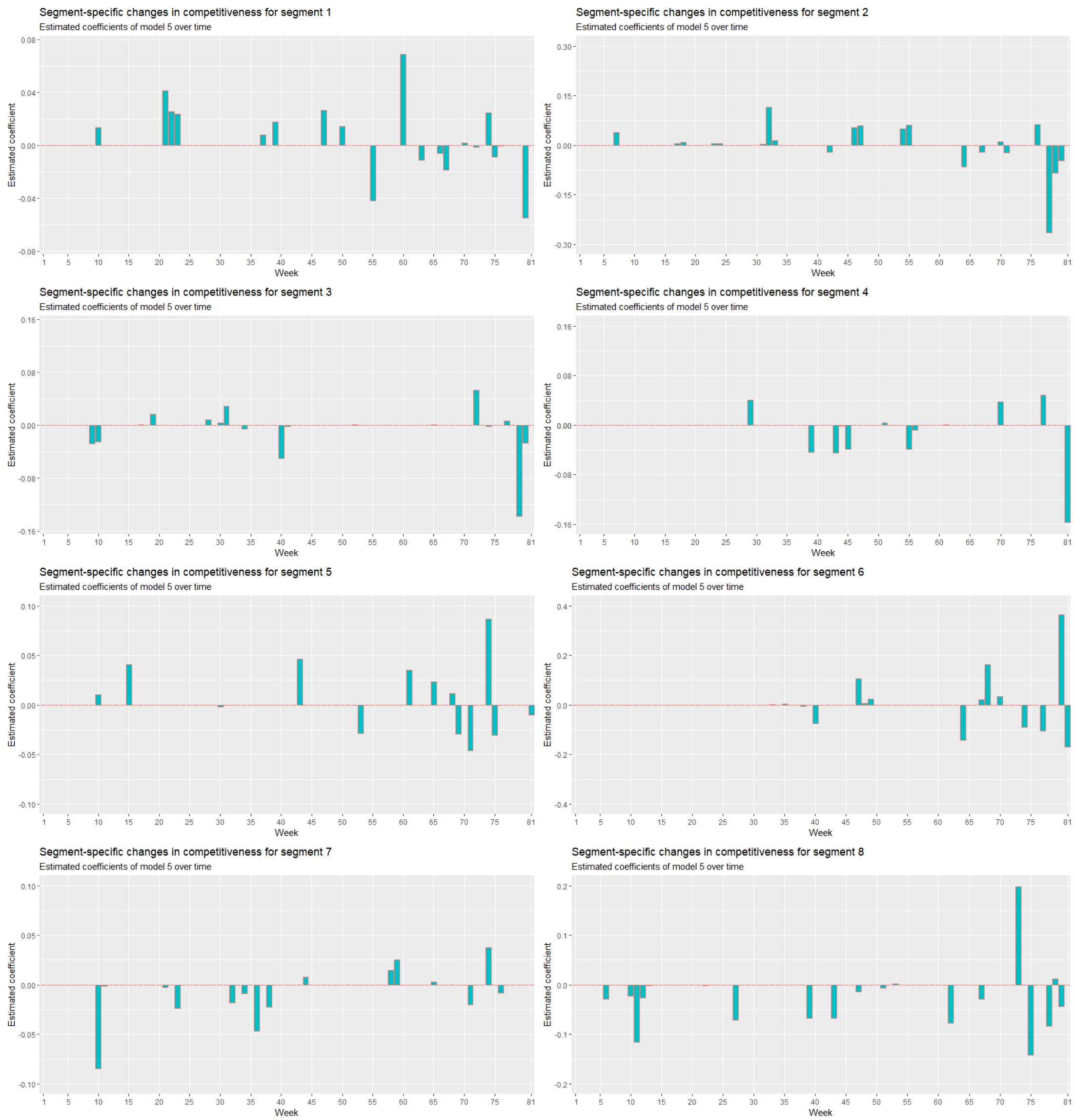

Figure 12: Segment-specific changes in competitiveness.

We observe two effects in the segment-specific changes in Figure 12. Firstly, we note a so-called grouping effect among time periods that are adjacent to each other. This is common for ridge regressions, where correlated variables tend to be shrunk together instead of removed, as would happen for lasso regression. Model 5 is an elastic net model with $\alpha = 0.2$, much closer to a ridge regression. Therefore, it is only natural that we see part of this grouping effect in the estimates of segment-specific changes. An example of this effect can be seen in the plot of segment 1 from week 21 to week 23. In practice, it is likely that the real competitiveness change happened in only one of the three weeks. However, the grouping effect has caused all three weeks to have a non-zero coefficient. This problem can be accommodated by choosing an α closer to 1.

Secondly, we note that there are relatively many shocks in competitiveness in the last weeks of the observation period. This is due to the nature of the model. When a shock in the observed coefficient occurs at the end of the observation period, the model does not have the chance to observe the conversion rate for many weeks to follow. Therefore, it cannot be certain that the decrease in conversion rate was noise, or a true shock in the coefficient. However, Model 5 gives more intuitive estimates of competitiveness in the last periods compared to Model 2 and Model 3, because it applies credibility weighting, and therefore the proposed approach is preferable.

Table 4: The sum of segment-specific changes and the correlation between estimated coefficients from Model 3 and 5

Segment number	Number of customers	Percentage of customers	Correlation of coefficients in Model 3 & Model 5	Sum of absolute value of changes in Model 5
1	164,911	24.76%	0.852	0.4101
2	26,836	4.03%	0.753	1.0190
3	165,738	24.89%	0.936	0.3978
4	55,001	8.26%	0.858	0.4639
5	73,138	10.98%	0.787	0.4008
6	18,251	2.74%	0.736	1.3108
7	148,583	22.31%	0.945	0.3243
8	13,350	2.00%	0.847	1.0096

Table 4 also shows the sum of absolute values of segment-specific changes in competitiveness level. This gives an indication of how much a segment deviates from the full portfolio. We observe that the smaller segments 2, 6, and 8 deviate considerably more from the full portfolio than the segments with more customers, even after applying credibility weighting. This can be considered as an undesirable property of the model. A potential solution is to increase the penalty factor of the segment-specific changes relative to the penalty factor of the full portfolio changes, so that relatively more credibility weighting is applied between segments and the full portfolio.

5 Discussion

We have proposed two data sources that can be used to infer competitiveness in a setting of limited or non-existent price information. Firstly, data of customer transitions between insurers in the Portuguese motor insurance market was collected to follow and compare quarterly competitiveness levels. Secondly, conversion data was employed to model weekly segment-wise competitiveness levels for Fidelidade's portfolio. Those two data sources clearly serve a different purpose. Conversion data is rich in information and is ideal to observe the changes in competitiveness by segment. However, conversion data is available only for our own portfolio and can therefore not be used to track the evolution of competitiveness between insurers, unlike transition data, which can be obtained for all insurers. Moreover, there is a bias in conversion data: The set of customers that disregards an insurer without requesting a quote is neglected. Transition data can be used to attempt to fill up this gap in information.

Firstly, we discuss and compare the binomial-HMM and Kalman smoother that use historical transition data. Both methods provide an approach to observe long-term trends of insurers and to track the evolution of competitiveness between insurers. The binomial-HMM assumes that the transition probability parameter can take on only five values and is therefore somewhat restricted. On the other hand, the Kalman smoother is more realistic because the state-space is continuous and can take on any value between 0 and 1. However, a potential problem of the Kalman smoother is within its specification. The true underlying transition probability is assumed to follow a random walk process, where the one-period competitiveness shock is normally distributed. In practice, the shock might follow a distribution with heavier tails, because of exogenous variables such as price changes. The binomial-HMM better handles these large shocks, because the state transition matrix allows to jump over some in-between states.

The Kalman smoother does remarkably well to filter noise from the observed incoming transition probabilities. The smoothed probabilities remain stable when the observations are hovering around a certain level. On the other hand, the Kalman smoother adapts quickly when the observed probability does not return to its previous level after a big shock. Moreover, the Kalman smoothed probability gives a reasonable estimate of the true probability in the last periods, which makes it especially useful for prediction. Both the binomial-HMM and the Kalman smoother can increase interpretability of transition probability time-series, which is particularly helpful to convey a message to decision-makers with a non-statistical background. Ideally, the methods will be used complementary to each other.

A major limitation of the current set-up is the lack of differentiation in the data. We have collected data only from license plates issued in 2005, knowing that the age of the license plate is highly correlated with other factors such as income of the driver, age of the driver, and geographical area. The current transition data might therefore not give a good overview of the competitiveness of the full portfolio. The data collection process turned out to be rather slow and inefficient. The collection of a more diversified set of license plates remains to be done in the future.

Secondly, we have proposed a penalized regression framework where time is included as a random walk structure. We aim to infer weekly segment-wise estimates of competitiveness changes. First, customer segmentation was performed to divide customers into homogeneous groups of similar conversion behavior. We have used a modified dummy encoding to model a random walk within the generalized linear modelling framework. We are not only able to model the overall changes in the portfolio, but also the changes by segment by including an interaction term between the segment and the random walk variable. This produces a model that credibility weights each segment's changes using the full portfolio's changes as the complement. The powerful hierarchical fashion of this model creates results that are much more interpretable than those of the generalized linear models, since the penalty factor sets most coefficients to zero. Moreover, we observe that the proposed method outperforms the generalized linear models in terms of predictive performance.

We generally observe more substantial segment-specific shocks in the last periods of the observation period. This is a direct effect of the dummy encoding, where the coefficient of the last period only influences that specific period. Therefore, the proposed model is of limited value for prediction of future competitiveness levels. A specific problem for Fidelidade's actuarial team is to predict the conversion rate of segments in the next week. A possible solution is to employ a Kalman filter where the state-dependent distribution is binomial, similar to the solution handling the transition data. The Kalman filter would not be able to handle a hierarchical structure and therefore would lack interpretability of segment-wise competitiveness levels, such as for the proposed model. However, the Kalman filter consistently outperforms a simple average over the last four weeks in terms of prediction for the next week's competitiveness level. The main focus of this work is on creating interpretable results for inferring competitiveness. Hence, we have chosen not to discuss the work that was geared towards prediction.

6 Conclusion

It is often costly or even impossible to collect information about competitor's prices and price changes, which poses a challenge for insurers who want to use objective measures to determine the competitiveness of their policies. In the past, there has been little attention to this problem in actuarial literature. We have proposed several methods to determine the competitiveness levels of insurance policies when price information is limited or non-existent. These methods are capable of handling time-dependent structures and in fact underline the importance of an underlying time structure.

We have collected transition data about the insurance companies of vehicles in the Portuguese auto insurance market between 2007 and 2019. Firstly, we proposed an approach using hidden Markov models to analyze the evolution of competitiveness of the main insurers in the market. The model assumes every insurer to be in one of 5 states of competitiveness, from very low to very high. The method creates interpretable results and succeeds in modelling relatively large shocks in the transition rate. However, the practical use is limited because of the assumption of a discrete state-space of competitiveness. We tackle this problem by considering a continuous state-space model. We apply Kalman smoothing to the observed transition rates to remove noise from the transition data. The approach allows us to observe the long-term development of an insurer's quarterly competitiveness levels and to compare it between insurers. Currently, our analysis is based on license plates from 2005. In the future, the focus should be on collecting data from different years to get a better picture of the evolution of competitiveness for the full portfolio.

Subsequently, we have used conversion data between January 2018 and September 2019 to determine segment-wise competitiveness levels of Fidelidade's portfolio. We have proposed a method for incorporating state-space model functionality into a generalized linear model framework. The method has a hierarchical structure and is able to apply credibility weighting between periods and between segments and the full portfolio. The proposed model has a better predictive performance than the most commonly used generalized linear model with time included as a categorical variable. Moreover, the estimated competitiveness levels are more intuitive and interpretable due to the penalty factor that is applied. The results are suitable for presentation to a non-technical audience. The model is of limited use in prediction of future period's competitiveness levels, although it still does better than the simple generalized linear model, because of the application of credibility weighting.

The current model is flexible enough to be extended in future work. We might be able to collect price data of competitors on a weekly basis. In this case, the cause of a change in competitiveness of our policies is understood and can be related to the price variables of competitors. The price variable can simply be added to the model to improve predictions for future weeks. It should be included as an index so that only price changes affect competitiveness. Similarly, a seasonal component can be incorporated into the model by adding a dummy variable to the penalized regression framework. These extensions can increase the added value of the model to make it an integral part of the price decision-making process.

References

Baggenstoss, P. M. (2001). A modified Baum-Welch algorithm for hidden Markov models with multiple observation spaces. *IEEE Transactions on speech and audio processing*, 9(4):411–416.

De Jong, P. (2005). *State-space models in actuarial science*. Institute of Mathematics and Statistics, University of São Paulo, Brazil.

De Jong, P. and Heller, G. Z. (2008). *Generalized linear models for insurance data*. Cambridge University Press.

De Jong, P. and Zehnwirth, B. (1983). Claims reserving, state-space models and the kalman filter. *Journal of the Institute of Actuaries*, 110(1):157–181.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.

Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state-space methods*. Oxford university press.

Evans, J. P. and Schmid, F. (2007). Forecasting workers compensation severities and frequency using the Kalman filter. In *Casualty Actuarial Society Forum*, pages 43–66. Citeseer.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1.

Gabadinho, A., Ritschard, G., Müller, N. S., and Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, 40(4):1–37.

Harte, D. (2017). *HiddenMarkov: Hidden Markov Models*. Statistics Research Associates, Wellington. R package version 1.8-11.

Helske, J. (2017). KFAS: Exponential family state space models in R. *Journal of Statistical Software*, 78(10):1–39.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied logistic regression*, volume 398. John Wiley & Sons.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*.

Kitagawa, G. (1987). Non-Gaussian state space modeling of non-stationary time series. *Journal of the American statistical association*, 82(400):1032–1041.

Kitagawa, G. and Gersch, W. (1996). Linear Gaussian state space modeling. In *Smoothness priors analysis of time series*, pages 55–65. Springer.

Korn, U. (2018). A simple method for modelling changes over time. In *Casualty Actuarial Society E-Forum*.

Kuhn, M. (2020). *caret: Classification and Regression Training*. R package version 6.0-86.

Miller, H. (2015). A discussion on credibility and penalised regression, with implications for actuarial work. *Actuaries Institute 2015 ASTIN and AFIR/ERM Colloquium.*

Nelder, J. A. and Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135(3):370–384.

Parodi, P. (2014). *Pricing in general insurance.* CRC Press.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.

Ross, S. M. (2014). *Introduction to probability models.* Academic press.

Therneau, T. and Atkinson, B. (2019). *rpart: Recursive Partitioning and Regression Trees.* R package version 4.1-15.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE transactions on Information Theory*, 13(2):260–269.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York.

Wickham, H., François, R., Henry, L., and Müller, K. (2020). *dplyr: A Grammar of Data Manipulation.* R package version 0.8.5.

Wickham, H. and Henry, L. (2020). *tidyr: Tidy Messy Data.* R package version 1.0.2.

Zehnwirth, B. (1996). *Kalman filters with applications to loss reserving.* University of Melbourne, Department of Economics, research paper 35, Centre for Actuarial Studies.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2):301–320.

Zucchini, W., MacDonald, I. L., and Langrock, R. (2017). *Hidden Markov models for time series: an introduction using R.* CRC press.

A Conversion variable explanations

Table 5: Explanation of the variables considered in the analysis of conversion data.

variable	**definition**	**type**
license	License plate number of a customer that performs a simulation (ID variable).	character
sim_date	Date of the first simulation performed by a customer.	date
brand	Brand of the car.	factor (23 levels: *)
engine	Engine class of the car.	factor (2 levels: *)
cylinder	Cylinder class of the car.	factor (7 levels: *)
power	Engine power class of the car.	factor (9 levels: *)
weight	Gross weight class of the car.	factor (9 levels: *)
RPP	Power/weight class of the car.	factor (7 levels: *)
weight2	Net weight class of the car.	factor (6 levels: *)
vehicle_value	Value of the car.	factor (11 levels: *)
tariff_zone	Class representing the claim propensity of the customer by geographical location, from low to high.	factor (12 levels: *)
tariff_zone2	Simplified variable of tariff_zone.	factor (3 levels: *)
age_license	Number of years that the customer has had a driver's license.	integer
age_license2	Simplified variable of age_license, merging similar levels into factors.	factor (5 levels: *)
age_vehicle	Age of the vehicle.	integer
age_vehicle2	Simplified variable of age_vehicle, merging similar levels into factors.	factor (9 levels: *)
birthyear	Birthyear period of the customer.	factor (11 levels: *)
intermediary	Intermediary involved in the simulation of a customer.	factor (10 levels: *)
bonusmalus	Bonus-malus level of the customer, where class 1 represents the level with the worst drivers, and class 9 the level with the best drivers.	factor (9 levels: *)
convert	Decision of the customer to convert at least one of the simulations.	binary

* Sensitive information regarding the factor levels was left out in this version.

B Conversion model results

The conversion model results are left out in this version to protect sensitive information.