Supplementary Material

This is the supplementary material of *Multi-State Modelling of Customer Churn*.

1 Five-State Transition Framework and Data Imbalance

Figure 1 shows the original five-state transition diagram of the application to data from LGPIF.

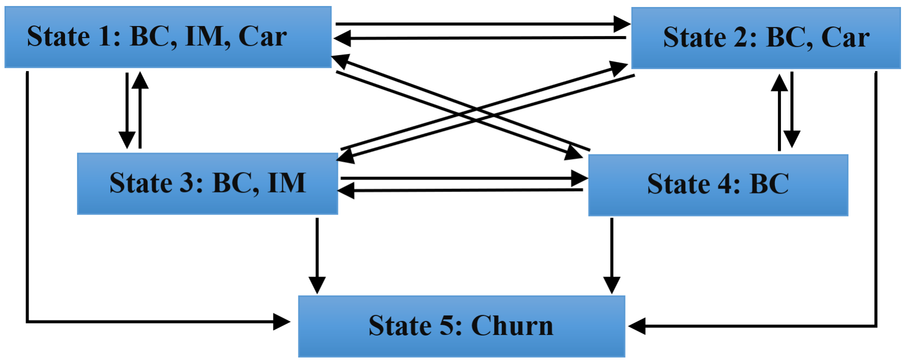

Figure 1: A five-state transition diagram of the application to data from LGPIF. BC refers to building and content insurance, IM refers to contractor's equipment insurance, while Car refers to vehicle coverage.

Table 1 lists the five-state transition counts and empirical transition probabilities from 2006 to 2013. We can see the numbers of several types of transition are very small. For example, the number of transitions from state 3 to state 2 is 0, which means that this type of transition has not occurred from 2006 to 2013 among all policyholders. The transitions in the minority are statistically rare events. Consequently, when a multi-state model is estimated on a random sample of the customer population, the majority transitions will dominate the statistical analysis, which may decrease the predictive accuracy on the minority transitions. This is a typical issue of data imbalance, and has been addressed through "balanced sampling" in the literature on traditional customer churn probability prediction; see Lemmens & Croux (2006). However, in this paper, we focus on using the second-order MLR model to study multi-state customer churn analysis rather than achieving a high degree of predictive accuracy, so we merged states 2, 3, and 4 in Figure 1 into one state in the paper to ensure that there is sufficient data to model all transition probabilities reasonably well.

Table 1: Five-state transition counts and empirical transition probabilities in percent (in parentheses) from 2006 to 2013.

	State of destination				
State of origin	State 1	State 2	State 3	State 4	Churn
State 1	2552 (94.14%)	4 (0.15%)	74 (2.73%)	3 (0.11%)	78 (2.88%)
State 2	11 (4.40%)	217 (87.80%)	1 (0.40%)	9 (3.60%)	12 (4.80%)
State 3	34 (0.94%)	0 (0.00%)	3479 (95.87%)	5 (0.14%)	111 (3.06%)
State 4	3 (0.25%)	7 (0.59%)	29 (2.45%)	1112 (94.08%)	31 (2.62%)

2 Explanatory Variables in Customer Churn Analysis in General Insurance

Premium information is found to be the key driver of customer churn, with the majority of literature finding that policyholders are more likely to churn after they experience the premium increase (Brockett et al. 2008, Haugen & Moger 2016, Jeong et al. 2018, De la Llave et al. 2019, Leiria et al. 2021). Premium information is also a frequently used predictor when applying machine learning techniques to predict customer churn rates (see for instance Bolancé et al. 2016, Paredes 2018, Scriney et al. 2020).

The claim experience is a special feature in general insurance. In life insurance, when a policyholder makes a claim, it often means that the policyholder has died and the life insurance contract is terminated. However, in general insurance, it is common to see a policyholder makes several claims in a single period, in which case customers will normally face a higher premium for the next period after they had made one or more claims in the previous period. Conversely, customers may receive a bonus which reduces the premium for the next period if they did not make any claim in the previous period. Jeong et al. (2018) and Frees et al. (2021) provide evidence of a strong association between claim occurrence and customer churn. Furthermore, numerous studies have shown that making a claim increases the probability of customer churn and reduces the overall lifetime of a contract (see for example Guillen et al. 2003, Brockett et al. 2008, Guillen et al. 2009, Haugen & Moger 2016).

Finally, contract information refers to a broad range of information that can be found in signed contracts between policyholders and insurers, including the geographic location of the insured object (e.g., Haugen & Moger 2016, Paredes 2018, Staudt & Wagner 2018, De la Llave et al. 2019, Frees et al. 2021), policyholder characteristics (e.g., Paredes 2018, Staudt & Wagner 2018, Frees et al. 2021), and other useful information such as the date of renewal.

3 Endogeneity and Exogeneity

In economic models, variables are commonly divided into two main categories: endogenous variables, i.e., variables that a model tries to explain, and exogenous variables, i.e., variables that a model takes as given (Mankiw 2003). Identifying whether a variable is exogenous versus endogenous poses a challenge for insurers. There are two main forms of exogeneity, depending on the level of independence shown by the variable. A strictly exogenous variable is completely unaffected by the output of a model in the past, present, and future. A sequentially exogenous (also called predetermined) variable is not affected by past instances of the model's output, but future instances may be affected by current or future instances of the model's output. If a variable is neither strictly exogenous nor sequentially exogenous, it is called endogenous. In the context of multi-state customer churn analysis, the endogenous variable is the transition among different states of a contract held by a policyholder, i.e., the values of y_{it} in the paper. In order to decide if a new variable is strictly exogenous or sequentially exogenous, we have to decide if the current or future customers' transition decision would cause the new variable to change in the future.

As we have seen in the existing literature, premium information is known to be statistically related to the customer churn, e.g., the higher the premium, the higher is the churn rate, and there is a causal relationship among the claim, the premium, and the probability of customer churn. To determine whether claims occurrence and frequency and premiums are strictly exogenous or sequentially exogenous in a multi-state transition model, we need to consider the moral hazard and adverse selection under the insurance context. Adverse selection is the tendency of policyholders in high-risk positions to purchase and renew insurance contracts. Moral hazard occurs if policyholders act in a more risky way after they enter insurance contracts. In the absence of moral hazard and adverse selection, a premium can be simply regarded as the insurance company actuary's summary measure of several risk factors with realistic considerations, and a claim is only affected by strictly exogenous factors such as climate and economic change. In this case, a customer's current or future purchasing decision will not affect his or her future premiums and claims, so both claims and premiums are strictly exogenous. On the other hand, if we take into account moral hazard and adverse selection, then policyholders who keep purchasing insurance contracts are more likely to make claims in the future, and high insurance claims in one year lead to high premiums in subsequent years. From this perspective, both premiums and claims evolve over time, and so both of them should be regarded as sequentially exogenous. For the purposes of this paper, we treat all variables as strictly exogenous, and we do so to be consistent with the existing literature on traditional customer churn analysis (see for instance Guillen et al. 2003, Brockett et al. 2008, among others). We also refer the reader to Pinquet (2000) for a discussion of exogeneity in an insurance rating context, and Chapter 6 of Frees (2004) for an overview of exogeneity in longitudinal models.

4 Robust Standard Errors of the Second-Order MLR Model

We provide robust standard errors using Stata 17; see Table 2. Generally, the formula for the robust estimator of variance is given in Stata (2022) as

$$\hat{V} = \hat{\mathrm{V}} \left(\sum_{j=1}^{M} \mathrm{u}'_{\boldsymbol{j}} \mathrm{u}_{\boldsymbol{j}} \right) \hat{\mathrm{V}}$$

where M is the total number of observations, $\hat{\mathrm{V}} = (-\partial^2 \ln L/\partial \boldsymbol{\beta}^2)^{-1}$ (the conventional estimator of variance), and $\mathrm{u}_{\boldsymbol{j}}$ (a row vector) is the contribution from the j-th observation to $\partial \ln L/\partial \boldsymbol{\beta}$.

For the second-order MLR model, we need to select a reference state and set the coefficients corresponding to this reference state to zero for reasons of parameter identifiability e.g., in the paper, $s = 1$ and $\beta_{qr1} = 0$ for all (q, r). For the transition with state of origin (q, r) and state of destination s, we define

$$\eta_{qrs} = \begin{cases} \frac{\exp\left(\boldsymbol{x}_{qr}^{\top} \boldsymbol{\beta}_{qrs}\right)}{1+\sum_{s'=2}^{Q} \exp\left(\boldsymbol{x}_{qr}^{\top} \boldsymbol{\beta}_{qrs'}\right)}, & 1 < s \leq Q, \\ \frac{1}{1+\sum_{s'=2}^{Q} \exp\left(\boldsymbol{x}_{qr}^{\top} \boldsymbol{\beta}_{qrs'}\right)}, & s = 1. \end{cases}$$

To calculate the score vector and Hessian matrix later, we will use the fact that for all $1 < l, s \leq Q$

$$\frac{\partial}{\partial \beta_{qrs}} \eta_{qrl} = \eta_{qrl} [\mathbf{1}(s = l) - \eta_{qrs}] \boldsymbol{x}_{qr}.$$

We write the score $= (\mathrm{sc}_{1,1,2}, \ldots, \mathrm{sc}_{1,1,Q}, \mathrm{sc}_{1,2,2}, \ldots, \mathrm{sc}_{1,2,Q}, \ldots, \mathrm{sc}_{Q-1,Q-1,2}, \ldots, \mathrm{sc}_{Q-1,Q-1,Q})$. For example, in the application to data from LGPIF of the paper, score $= (\mathrm{sc}_{1,1,2}, \mathrm{sc}_{1,1,3}, \ldots, \mathrm{sc}_{2,2,3})$. Any $\mathrm{sc}_{q,r,s}$ can be calculated as

$$\begin{aligned} \mathrm{sc}_{q,r,s} &= \frac{\partial}{\partial \boldsymbol{\beta_{qrs}}} \ln L \\ &= \sum_{j=1}^{M_{qr}} [\mathbf{1}(y_j = s) - \eta_{j,qrs}] \boldsymbol{x}_{j,qr}, \end{aligned}$$

where M_{qr} is the observed number of transitions with states of origin (q, r), y_j is the state of destination for the j-th transition among M_{qr} transitions, and $\boldsymbol{x}_{j,qr}$ is the vector of corresponding covariates used in the second-order MLR model.

The Hessian matrix is a block diagonal matrix, with (r, s)-th block given by

$$
\begin{aligned}
\mathrm{bl}_{rs} &= \frac{\partial^2}{\partial \boldsymbol{\beta}_{qrs} \partial \boldsymbol{\beta}_{qrl}} \ln L \\
&= -\sum_{j=1}^{M_{qr}} \eta_{j,qrs} [\mathbf{1}(y_j = l) - \eta_{j,qrl}] \boldsymbol{x}_{j,qr} \boldsymbol{x}_{j,qr}^{\top}.
\end{aligned}
$$

If we compare the coefficients in Table 2 below with those in Table 3 of the paper, we can see the coefficients are slightly different in several cases, which is due to the use of different softwares (R in the paper and Stata in the supplementary material).

Table 2: Summary of the second-order MLR model with robust standard errors (in parentheses.)

State of destination	to partial-coverage ($s = 2$)	to churn ($s = 3$)
State of origin: (1,1)		
Intercept	-19.76^{***} (1.88)	-19.88^{***} (1.36)
Entity(ScToVi)	16.25^{***} (0.34)	15.36^{***} (0.25)
Entity(CiMi)	17.18^{***} (0.39)	16.25^{***} (0.53)
RateBC	-2.86^{***} (1.03)	0.30 (0.60)
RateIM	1.37^{*} (0.79)	0.76 (0.61)
RateCar	-0.39^{***} (0.13)	-0.04 (0.13)
I(ClaimBC)	0.20 (0.34)	-0.64^{*} (0.38)
I(ClaimIM)	-0.47 (1.02)	-14.66^{***}(0.31)
TotalCoverage	-0.88^{**} (0.31)	-0.66^{*}(0.51)
RatioPremium	0.81 (1.07)	0.93^{*}(0.51)
State of origin: (2,1)		
Intercept	-18.02^{***} (0.58)	
Entity(ScToVi)	16.34^{***} (0.76)	
I(ClaimCar)	-16.63^{***} (0.67)	
State of origin: (1,2)		
Intercept	16.68^{***} (0.90)	16.65^{***} (1.12)
Entity(ScToVi)	-17.06^{***} (1.01)	-16.82^{***} (0.95)
RateIM	1.19^{*} (0.77)	-0.07 (0.66)
State of origin: (2,2)		
Intercept	19.75^{***} (0.18)	16.29^{***} (0.33)
Entity(CiScToVi)	-14.75^{***} (0.28)	-14.63^{***} (0.37)
FrequencyBC	0.02 (0.09)	-0.72^{***} (0.22)
RatioPremium	3.25^{**} (1.39)	3.56^{**} (1.40)
I(FinaCris)	-0.90^{**} (0.41)	-0.42 (0.46)
Note:		*p<0.1; **p<0.05; ***p<0.01

5 Goodness-of-Fit Tests

Due to the fact that multinomial models have more than one groups of coefficients, assessing Goodness-of-Fit (GoF) is more challenging, and is still an area of intense research. The most approachable method to assess model confidence is the Hosmer-Lemeshow (HL) test (Hosmer & Lemesbow 1980), which is extended in Fagerland et al. (2008) for MLR models. In a MLR model setting, we have data of the form $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, where y_i is a k-vector that indicates which class the i-th observation belongs to (exactly one entry contains a one and the rest are zero), x_i is the vector of explanatory variables for the i-th observation, and n is the number of observations. After we fit a MLR model that estimates a vector of predicted probabilities for k classes $\hat{p}(x_i) = \big(\hat{p}_1(x_i), \hat{p}_2(x_i), \ldots, \hat{p}_k(x_i)\big)$, the predicted values $(1 - \hat{p}_1(x_i)$ assuming the class 1 is the reference level) are arrayed from the lowest to the highest, and then separated into several groups of approximately equal size. Let O_{hj} and E_{hj} denote the sum of the observed frequencies and estimated frequencies for j-th class in h-th group. The HL statistic is calculated as

$$C_g = \sum_{h=1}^{g} \sum_{j=1}^{k} \frac{(O_{hj} - E_{hj})^2}{E_{hj}},$$

where g is number of groups (normally is set to be 10). Under the null hypothesis that the fitted model is the correct model and the sample is sufficiently large, we expect C_g to have an approximate χ^2 distribution with $(g-2) \times (k-1)$ degrees of freedom.

Besides the HL test, the deviance and Pearson chi-square tests for the binary logistic regression can also be extended to MLR models. The deviance can be written as

$$D = -2 \sum_{i=1}^{n} \sum_{j=1}^{k} y_{ij} \log \left(\frac{\hat{p}_j(x_i)}{y_{ij}} \right),$$

where y_{ij} is an indicator for j-th class in vector y_i, and we treat zero times the log of anything as zero. Under the null hypothesis that the fitted model is the correct model, the distribution of D is chi-squared with $(n-p) \times (k-1)$ degrees of freedom, where p is the number of explanatory variables including intercept.

The Pearson chi-square is calculated as

$$\chi^2 = \sum_{i=1}^{n} \sum_{j=1}^{k} \frac{\big(y_{ij} - \hat{p}_j(x_i)\big)^2}{\hat{p}_j(x_i)}.$$

Under the null hypothesis that the fitted model is the correct model, the distribution of χ^2 is chi-squared with $(n-p) \times (k-1)$ degrees of freedom.

In Table 3, we show the results of the three tests for the second-order MLR model in the application to data from LGPIF. In all cases, p-values are high, so there is no evidence showing that the second-order MLR model is a poor fit.

Table 3: Goodness-of-fitting tests for the second-order MLR model.

	Pearson Chi-Square			Deviance			HL (10 groups)		
	Statistic	df	p-value	Statistic	df	p-value	Statistic	df	p-value
Second-order MLR	9258.0	10846	1	2191.6	10846	1	15.5	16	0.49

6 Exploratory Variables of an "Average" Policyholder

The values of exploratory variables of an average policyholder are listed in Table 4.

Table 4: The state-specific values of exploratory variables of an "average" policyholder.

Exploratory variable	**Full-coverage**	**Partial-coverage**
Entity(ScToVi)	1	1
Entity(CiMi)	0	0
RateBC	0.537	0.652
RateIM	2.214	1.660
RateCar	3.600	0.169
I(ClaimBC)	0	0
I(ClaimIM)	0	0
I(ClaimCar)	0	0
TotalCoverage	0.666	0.285
RatioPremium	0.034	0.018

7 One-versus-All Binary Logistic Regression

It is worth comparing MLR models to another, an obvious approach that could be used for multi-state customer churn analysis, namely fitting multiple sets of binary logistic regression (BLR) models based on either a one-versus-all (also known as one-versus-rest) or one-versus-one approach. Recall that in traditional customer churn analysis, BLR models are fitted to a binary response for identifying the explanatory variables driving the probability of customer churn. With multi-state data, one can extend this approach by fitting separate BLR models for each transition, including the transition to churn.

We illustrate this with a one-versus-all, second-order BLR model approach as follows. Consider the example discussed in the right panel of Figure 1 of the paper, where there are three possible states a policyholder can transition to at time t (full-coverage or state 1, partial-coverage or state 2, churn or state 3). Then for each state of origin (q, r) where $q = \{1, 2\}$, $r = \{1, 2\}$, a one-vs-all, second-order BLR model would construct the following three separate logistic regression models:

$$\begin{aligned} P(y_{i,t} = 1 | y_{i,t-2} = q, y_{i,t-1} = r) &= \text{logit}^{-1}\left(\boldsymbol{x}_{i,t-1,qr1}^{\top}\boldsymbol{\beta}_{qr1}\right), \\ P(y_{i,t} = 2 | y_{i,t-2} = q, y_{i,t-1} = r) &= \text{logit}^{-1}\left(\boldsymbol{x}_{i,t-1,qr2}^{\top}\boldsymbol{\beta}_{qr2}\right), \\ P(y_{i,t} = 3 | y_{i,t-2} = q, y_{i,t-1} = r) &= \text{logit}^{-1}\left(\boldsymbol{x}_{i,t-1,qr3}^{\top}\boldsymbol{\beta}_{qr3}\right), \end{aligned} \tag{1}$$

where $\text{logit}^{-1}(x) = \exp(x)/(1 + \exp(x))$ denotes the inverse link function. It is important to emphasise that in equation (1), three sets of BLR models are fitted for each state of origin. This contrasts to the MLR model in equation (2) of the paper, i.e., for each state of origin, only a single model needs to be fitted for all transition probabilities. Indeed, this reflects a fundamental difference between a MLR modelling approach and a one-versus-all BLR modelling approach for multi-state analysis: in the latter, transition probabilities from the same state of origin are independently modelled, and thus need not and in general do not sum to one. That is, in equation (1) there is no guarantee that $P(y_{i,t} = 1 | y_{i,t-2} = q, y_{i,t-1} = r) + P(y_{i,t} = 2 | y_{i,t-2} = q, y_{i,t-1} = r) + P(y_{i,t} = 3 | y_{i,t-2} = q, y_{i,t-1} = r) = 1$. One consequence of this is that the transition probabilities from BLR models do not properly reflect the conditional or marginal effects of covariates in multi-state customer analysis, and we discuss this point further in Section 4.3 of the paper. Also, it is clear that, conditional on a state of origin (q, r), the coefficients in a one-versus-all BLR model only affect that particular corresponding transition probability. In contrast, for the MLR modelling approach in equation (2) of the paper, the coefficients $\boldsymbol{\beta}_{qrs}$ affect all transition probabilities, as they occur in both the numerator and the denominator of the regression function. This is again nothing more than a direct consequence of the sum to one constraint.

In this paper, we choose to use MLR models instead of a set of separate BLR models for the purposes of multi-state customer churn analysis, as the former naturally considers the dependence among the transition probabilities. Indeed, MLR models precisely quantify the effect of covariates on transitions to a multitude of states, and thus offers a more complete picture of dynamic changes in the needs of policyholders over time. That being said, it is important to acknowledge that if the primary question of interest is related to studying *specific* transitions instead of all transitions jointly, then the one-versus-all BLR modelling approach is suited for this. For example, if the primary focus was studying the state of destination as churn, then building a BLR model based on the third line of equation (1) would be a more direct way of answering this question since it explicitly and only models the probability of customer churn versus not churn. However, it is also possible to use the MLR model to answer the same question, albeit with more mathematical calculations.

8 Scenarios by Conducting Traditional Customer Churn Analysis and Multi-state Customer Churn Analysis

We visualise the scenarios of future path for traditional and multi-state customer churn analysis in Figure 2.

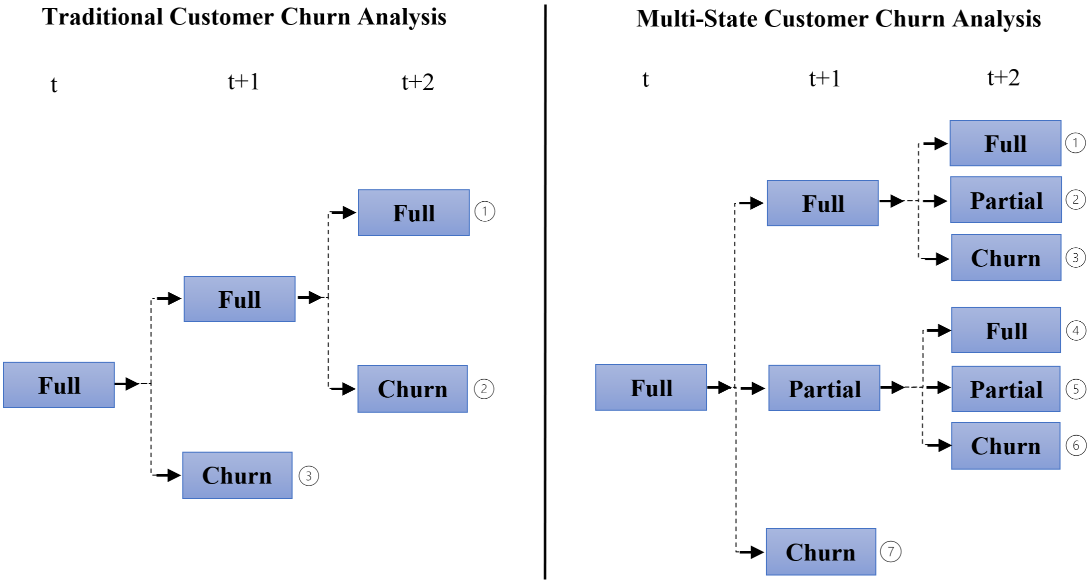

Figure 2: Scenarios of future path for CLV calculation: three scenarios for traditional customer churn analysis (left panel) versus seven scenarios for multi-state customer churn analysis (right panel).

9 Models in Out-of-Sample Validation

9.1 Support Vector Machine

Support vector machine (SVM) is one of the most popular supervised learning algorithms, which is used for classification as well as regression problems (Noble 2006). However, primarily, it is used for classification problems in machine learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that people can easily put the new data points in the correct category in the future. However, one disadvantage of SVM is that it does not output probabilities natively, but probability calibration methods can be used to convert the output to class probabilities. Various methods exist, including Platt scaling (particularly suitable for SVMs) and isotonic regression. In this paper, we use the default probabilities provided in `svm` function in `e1071` package (Meyer et al. 2021). The default probability model for SVM classification fits a logistic distribution using maximum likelihood to the decision values of all binary classifiers, and computes the a-posteriori class probabilities for the multi-class problem using quadratic optimisation.

The SVM is an approximate implementation of a theoretical bound on the generalisation performance that is independent of the dimensionality of the feature space. To build a non-linear SVM classifier, we can use either the polynomial kernel or the radial kernel function. In this paper, we use the radial kernel function. There are two hyperparameters to be tuned when we use the radial kernel function. `C`, the "cost" of the radial kernel, controls the complexity of the boundary between support vectors. The radial kernel also requires setting a smoothing parameter, `sigma`. The `caret` (Kuhn 2021) package can be used to tune the radial hyperparameters of SVM radial kernel function models by setting `method = "svmRadial"`. We use 10-fold cross validation to find the optimal hyperparameters that can maximise accuracy in the original training set and balanced training set. Accuracy is the ratio of the number of correctly classified instances to the total number of instances. The result of optimal hyperparameters is summarised in Table 5.

Table 5: Tuning results of SVM classifiers.

	Original training set	**Balanced training set**
C	1	4096
sigma	0.0326	0.0337

9.2 Gradient Boosting Machine

Gradient boosting machine (GBM) combines the predictions from multiple decision trees (weak learners) to generate the final predictions (Friedman 2001). Additionally, each new tree takes into account the errors or mistakes made by the previous trees. So, every successive decision tree is built on the errors of the previous trees. This is how the trees in a gradient boosting machine algorithm are built sequentially.

The hyperparameters of a GBM model to be tuned are `interaction.depth`, `n.trees`, `shrinkage`, and `n.minobsinnode`. The definition of these hyperparameters can be found in gbm (Greenwell et al. 2020) package. `interaction.depth` specifies the maximum depth of each tree (i.e., the highest level of variable interactions allowed). `n.trees` specifies the total number of trees to fit. `shrinkage` is a shrinkage parameter applied to each tree in the expansion (also known as the learning rate or step-size reduction). `n.minobsinnode` specifies the minimum number of observations in the terminal nodes of the trees. GBM model is tuned by 5-fold CV to minimise logLoss using `caret` (Kuhn 2021) package by setting `distribution="multinomial"`, `method="gbm"`, and `metric="logLoss"`. `logLoss` computes the minus log-likelihood of the multinomial distribution (without the constant term):

$$-\text{logLoss} = \frac{-1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k} y_{ij}\log(p_{ij}),$$

where the y values are binary indicators for the classes and p are the predicted class probabilities, n is number of observations, and k is number of classes.

The result of optimal hyperparameters in the original training set and balanced training set is summarised in Table 6.

Table 6: Tuning results of GBMs.

	Original training set	**Balanced training set**
interaction.depth	3	2
n.trees	400	5000
shrinkage	0.01	0.01
n.minobsinnode	10	10

References

Bolancé, C., Guillen, M. & Padilla-Barreto, A. E. (2016), Predicting probability of customer churn in insurance, *in* 'International Conference on Modeling and Simulation in Engineering, Economics and Management', Springer, pp. 82–91.

Brockett, P. L., Golden, L. L., Guillen, M., Nielsen, J. P., Parner, J. & Perez-Marin, A. M. (2008), 'Survival analysis of a household portfolio of insurance policies: how much time do you have to stop total customer defection?', *Journal of Risk and Insurance* **75**(3), 713–737.

De la Llave, M. Á., López, F. A. & Angulo, A. (2019), 'The impact of geographical factors on churn prediction: an application to an insurance company in madrid's urban area', *Scandinavian Actuarial Journal* **2019**(3), 188–203.

Fagerland, M. W., Hosmer, D. W. & Bofin, A. M. (2008), 'Multinomial goodness-of-fit tests for logistic regression models', *Statistics in medicine* **27**(21), 4238–4253.

Frees, E. W. (2004), *Longitudinal and panel data: analysis and applications in the social sciences*, Cambridge University Press.

Frees, E. W., Bolancé, C., Guillen, M. & Valdez, E. A. (2021), 'Dependence modeling of multivariate longitudinal hybrid insurance data with dropout', *Expert Systems with Applications* **185**, 115552.

Friedman, J. H. (2001), 'Greedy function approximation: a gradient boosting machine', *Annals of statistics* pp. 1189–1232.

Greenwell, B., Boehmke, B., Cunningham, J. & Developers, G. (2020), *gbm: Generalized Boosted Regression Models*. R package version 2.1.8.
URL: *https://CRAN.R-project.org/package=gbm*

Guillen, M., Nielsen, J. & Pérez-Marín, A. (2009), 'Cross buying behaviour and customer loyalty in the insurance sector', *EsicMarket* **132**, 77–105.

Guillen, M., Parner, J., Densgsoe, C. & Perez-Marin, A. M. (2003), Using logistic regression models to predict and understand why customers leave an insurance company, *in* 'Intelligent and other computational techniques in insurance: Theory and applications', World Scientific, pp. 465–490.

Haugen, M. & Moger, T. A. (2016), ‘Frailty modelling of time-to-lapse of single policies for customers holding multiple car contracts’, *Scandinavian Actuarial Journal* **2016**(6), 489–501.

Hosmer, D. W. & Lemesbow, S. (1980), ‘Goodness of fit tests for the multiple logistic regression model’, *Communications in statistics-Theory and Methods* **9**(10), 1043–1069.

Jeong, H., Gan, G. & Valdez, E. A. (2018), ‘Association rules for understanding policyholder lapses’, *Risks* **6**(3), 69.

Kuhn, M. (2021), *caret: Classification and Regression Training*. R package version 6.0-88.
URL: *https://CRAN.R-project.org/package=caret*

Leiria, M., Rebelo, E. & deMatos, N. (2021), ‘Measuring the effectiveness of intermediary loyalty programmes in the motor insurance industry: loyal versus non-loyal customers’, *European Journal of Management and Business Economics* .

Lemmens, A. & Croux, C. (2006), ‘Bagging and boosting classification trees to predict churn’, *Journal of Marketing Research* **43**(2), 276–286.

Mankiw, N. G. (2003), *Macroeconomics*, eighth edn, New York: Worth Publishers.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A. & Leisch, F. (2021), *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*. R package version 1.7-9.
URL: *https://CRAN.R-project.org/package=e1071*

Noble, W. S. (2006), ‘What is a support vector machine?’, *Nature biotechnology* **24**(12), 1565–1567.

Paredes, M. (2018), A case study on reducing auto insurance attrition with econometrics, machine learning, and a/b testing, *in* ‘2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)’, IEEE, pp. 410–414.

Pinquet, J. (2000), Experience rating through heterogeneous models, *in* ‘Handbook of Insurance’, Springer, pp. 459–500.

Scriney, M., Nie, D. & Roantree, M. (2020), Predicting customer churn for insurance data, *in* ‘International Conference on Big Data Analytics and Knowledge Discovery’, Springer, pp. 256–265.

Stata (2022), Estimation and postestimation commands, Technical report.
URL: *https://www.stata.com/manuals13/u20.pdf*

Staudt, Y. & Wagner, J. (2018), ‘What policyholder and contract features determine the evolution of non-life insurance customer relationships?’, *International Journal of Bank Marketing* .