MULTI-STATE MODELLING OF CUSTOMER CHURN

BY

YUMO DONG, EDWARD W. (JED) FREES, FEI HUANG, AND FRANCIS K. C. HUI

ABSTRACT

Customer churn, which insurance companies use to describe the non-renewal of existing customers, is a widespread and expensive problem in general insurance, particularly because contracts are usually short-term and are renewed periodically. Traditionally, customer churn analyses have employed models which utilise only a binary outcome (churn or not churn) in one period. However, real business relationships are multi-period, and policyholders may reside and transition between a wider range of states beyond that of the simply churn/not churn throughout this relationship. To better encapsulate the richness of policyholder behaviours through time, we propose multi-state customer churn analysis, which aims to model behaviour over a larger number of states (defined by different combinations of insurance coverage taken) and across multiple periods (thereby making use of readily available longitudinal data). Using multinomial logistic regression (MLR) with a second-order Markov assumption, we demonstrate how multi-state customer churn analysis offers deeper insights into how a policyholder's transition history is associated with their decision making, whether that be to retain the current set of policies, churn, or add/drop a coverage. Applying this model to commercial insurance data from the Wisconsin Local Government Property Insurance Fund, we illustrate how transition probabilities between states are affected by differing sets of explanatory variables and that a multi-state analysis can potentially offer stronger predictive performance and more accurate calculations of customer lifetime value (say), compared to the traditional customer churn analysis techniques.

KEYWORDS

Commercial insurance, multinomial logistic regression, higher-order Markov, longitudinal data, transition probability.

Astin Bulletin 52(3), 735–764. doi:10.1017/asb.2022.18 © The Author(s), 2022. Published by Cambridge University Press on behalf of The International Actuarial Association.

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

1. Introduction

Customer churn (also known as customer attrition, dropout, turnover, or defection) refers to the non-renewal of existing customers for an insurance company. In general insurance, customer churn poses a major challenge to insurers because contracts are renewed periodically (commonly after one year), and with substantial competition and diverse contracts, it is easy for customers to switch insurance providers to satisfy their needs at the time. Customer churn also poses a challenge in life insurance and other industries such as banking and telecommunications; keeping a current customer rather than acquiring a new one is often a more profitable strategy for a company (Simchi-Levi *et al*., 2008), because long-term consumers typically take up less of the company's resources compared to new customers (Min and Zhou, 2002). As such, analyses to determine the drivers and customer behaviours involving churn are critical to a company's customer retention programme (Okongwu *et al*., 2016); identifying those customers with high intention to churn can help insurers secure continued loyalty.

This paper is motivated by customer churn analysis for commercial insurance, which plays an important role in the general insurance industry. According to the direct premium written reported by Brian Briggs (2021), the commercial market represents 49% of the total direct premium written in the US property & casualty insurance industry. Commercial insurance contracts can include several coverages such as building, equipment, and auto coverage, and it is common for customers to have multiple contracts with the same company or (alternatively) have multiple coverages under a single bundled contract (Dong *et al*., 2022). Additionally, customers add or drop coverages, or switch providers entirely, on a regular basis to meet their benefits and budgets. Indeed, as revealed in a survey for Australian small commercial insurance consumers (Brant Carson, 2017), in the small and medium enterprise insurance market, around three-quarters of small business owners are dissatisfied with their commercial insurance providers. Approximately 19% of small business owners choose to switch insurance providers every year.

Since commercial insurance contracts are typically of short duration (so that customer churn is more about the non-renewal of a contract rather than breaking an existing contract), traditional customer churn analysis has focused on static models applied to data from a single period. This is despite the fact that commercial insurers can trace information about their customers over multiple periods, resulting in readily available longitudinal data. Moreover, with single period data, various methods such as logistic regression, decision trees, and neural networks have been proposed to predict the probability of customer churn for the next contract period (e.g., Yeo *et al*., 2001; Guillen *et al*., 2003; He *et al*., 2020). However, with multi-period or longitudinal data, customers often traverse a larger number of states beyond simply deciding whether to churn or not. That is, as discussed above, during a multi-period contract term,

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

commercial policyholders may regularly adjust their coverages combination by adding and dropping coverages. For example, an organisation policyholder taking both building and equipment coverages as a bundled contract in one year may decide to drop equipment coverage but add auto coverage in the following year. Put simply, multi-period data more accurately reflect the dynamic changes in the needs of policyholders and the broader market, for example, the coverage first dropped by a policyholder is usually an important signal to predict subsequent likelihood to churn (Brockett *et al*., 2008). Thus, there is a strong motivation for commercial insurers to break away from traditional customer churn analysis. In this paper, we propose and advocate using multi-period data to perform multi-state customer churn analysis, focusing not only on the customer churn *per-se* but also on the coverage adjustment behaviour of commercial customers over time.

While there are a number of ways by which multi-state customer churn analysis can be performed, in this article, we adopt a method that prioritises interpretability over predictive performance, namely multinomial logistic regression (MLR) with a higher-order Markov framework. To our knowledge, this paper is the first to use a higher-order Markov model for customer churn research in actuarial science. We illustrate the use of this MLR and multi-state customer churn analysis on a longitudinal commercial insurance dataset from the Wisconsin Local Government Property Insurance Fund (LGPIF). Exploratory analysis of this data reveals that transition probabilities tend to depend strongly on the states of the policyholder in their two previous periods, which motivates considering a second-order MLR model to study policyholders' states transition. We demonstrate how different sets of predictors tend to drive the various likelihoods of transition, which we visualise through conditional effect and average marginal effect plots. A comparison of multi-state customer churn analysis to traditional (single-state) customer churn analysis reveals that the former can lead to more accurate calculations of customer lifetime value. Furthermore, the second-order MLR model has consistently strong out-of-sample predictive performance relative to several parametric models of different orders and also compared to more non-parametric techniques such as support vector machines (SVMs) and gradient boosting machines (GBMs).

The remainder of this paper is organised as follows. In Section 2, we review the literature on customer churn and multi-state modelling. In Section 3, we illustrate how to perform multi-state customer churn analysis using MLR models under the first- and second-order Markov assumptions. In Section 4, we illustrate an application of second-order MLR models to study multi-state customer churn in the LGPIF data, visualising the estimated effects along with demonstrating differences in calculating customer lifetime value between multi-state versus traditional customer churn modelling. In Section 5, we compare the out-of-sample prediction performance of the second-order MLR model with several other methods for forecasting transition. Section 6 offers some concluding remarks.

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

2. Literature review

In actuarial science, multi-state models are commonly divided into time-discrete and time-continuous approaches (see Haberman and Pitacco, 2018, for an overview), with the primary difference being whether the transitions between states are assumed to occur at any time or at equally spaced time points during a certain research period. In this paper, motivated by commercial insurance data from the LGPIF and to be consistent with traditional customer churn analysis (e.g., Guillen *et al*., 2003; Brockett *et al*., 2008), we adopt the former and assume that the results of policyholders' renewal decisions are realised at the end of each contract period. However, we acknowledge the vast literature on time-continuous multi-state models used in long-term care insurance (e.g., Pritchard, 2006; Fong *et al*., 2015), life insurance (e.g., Norberg, 2013; Buchardt, 2014; Buchardt *et al*., 2015), and general insurance (e.g., Orr, 2007; Hürlimann, 2015).

In terms of traditional customer churn analysis, particularly in general insurance, there exists an array of different techniques for modelling the likelihood of customer churn (a binary response). By far, the most popular of these is logistic regression, or some variation thereof, to quantify the factors driving customer churn (see Guillen *et al*., 2003; Günther *et al*., 2014; Staudt and Wagner, 2018; Frees *et al*., 2021, among many others). More recently, data mining techniques such as tree-based methods and neural networks have grown in popularity for predicting churn rates (Bolancé *et al*., 2016; He *et al*., 2020). Yet another approach uses survival analysis to study how long a customer will stay with a company after their first policy cancellation (see for instance Brockett *et al*., 2008; Guillen *et al*., 2009, 2012).

It is important to highlight the general trade-offs among these different techniques for traditional customer churn analysis. Logistic regression is more parsimonious in the sense of having coefficients that explicitly quantify the effects of predictors on the response variable (churn rate), on which one can perform statistical inference on. In contrast, machine learning models often exhibit superior predictive capacity due to their relative flexibility, but the parameters of (most) machine learning models are related more to controlling the learning process, without necessarily having a practical meaning. Such trade-offs also exist in the context of multi-state customer churn analysis. In this paper, we adopt MLR, which is a natural extension of logistic regression to the case of categorical responses with more than two levels, as we prioritise interpretability over predictive capacity (although we note that in our application to the LGPIF data, MLR worked reasonably well in predicting transitions compared to some machine learning techniques for multi-state modelling).

MLR has already been used in actuarial science to study various multi-class classification problems. For example, Christiansen *et al*. (2016) employed MLR to study policyholders' decisions to switch insurance coverage in long-term health insurance, Zhu *et al*. (2015) used MLR to model the claim rate, lapse rate, and in-force rate jointly in life and annuity insurance, while

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

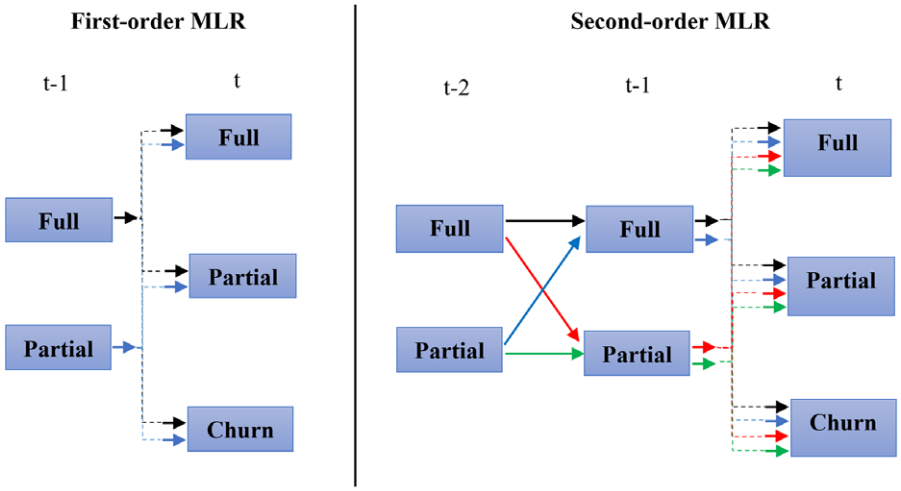


FIGURE 1. Multi-state customer churn analysis: first-order MLR model (left) and second-order MLR model (right).

Alai *et al*. (2015) and Kwon and Nguyen (2019) applied MLR to model cause-specific mortality. A more relevant work is that of Antonio *et al*. (2016), who used MLR with a first-order Markov assumption to build a multi-state model for the development of a non-life insurance claim process. However, to the best of our knowledge, this paper is the first to propose its use with a second-order Markov assumption in customer churn analysis.

3. MULTI-STATE CUSTOMER CHURN MODELLING

For constructing time-discrete multi-state models of customer churn, there are two strategies we can adopt in terms of fitting MLR models; see Chapter 11.3 of Frees (2004). First, we can split the data into separate subsets based on each state of origin and build a separate MLR model for each subset. For example, suppose there is a total of three possible states that a policyholder can be in, namely full-coverage, partial-coverage, and churn, and we want to model the probabilities of transitioning into these states based (only) on the current state. Then, we could split the data based on the policyholder's current state, such that we have a subset comprising all observations whose current state is full-coverage, and another subset comprising all observations whose current state is partial-coverage (note we ignore the subset comprising all observations whose current state is churn; see also Figure 1 later on). Then, we fit a separate MLR model to each subset, thereby modelling the transition probabilities from these two states of origin.

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

Alternatively, we can fit a single MLR model to the whole data and use indicator variables to denote the states of origin when calculating the likelihood contribution. That is, we can construct interaction terms between each of the explanatory variables and a factor variable indicating the state of origin (in the above example, either full-coverage or partial-coverage) and then include this expanded set of predictors in the MLR model when calculating the likelihood contribution.

Both the above strategies have their advantages, with the former being computationally more efficient (especially using parallel computing), while the latter is more unified and offers a convenient means of (say) testing the equality of parameters; that is, the effect of a particular explanatory variable is the same irrespective of the state of origin and thus simplifying the regression structure. Importantly though, both methods of fitting the MLR model will produce the same estimates, statistical inference, and predictions. With this in mind, we will adopt the second strategy in this paper and use an expanded set of covariates in what follows.

To trace the path of a policyholder's states over time, let $y_{i,t}$ denote the state of i-th policyholder at current time t, and $\{y_{i,1}, \ldots, y_{i,t-1}\}$ denote the history of i-th policyholder's states from time 1 up to the previous time point $(t-1)$. Throughout the paper, we assume independence among individuals i for $i = (1, 2, \ldots, N)$, which is standard and widely used in the actuarial studies (see e.g., Frees *et al*., 2016; Okine *et al*., 2022, among others for the LGPIF data specifically) as well as more broadly in the longitudinal data analysis literature; see also our discussion in Section 6. To build a MLR model (or any general approach for customer churn analysis as a matter of fact), a key decision to be made is how much of the policyholder's state history is associated with the likelihood of being in their current state. That is, what order a Markov assumption should be assumed for transition probabilities. In actuarial science, by far the most common assumption is that of the first-order Markov property (Kwon and Jones, 2006, 2008; Bettonville *et al*., 2021), meaning that the probability of being in the current state depends only on the state at the previous time point, and not on earlier states. That is, $P(y_{i,t}|y_{i,1}, \ldots, y_{i,t-1}) = P(y_{i,t}|y_{i,t-1})$. For multi-state customer churn analysis, this leads to the following first-order MLR model. Assume there are a total of Q states that the policyholder can transition to, that is, states of destination. Then for the i-th policyholder, the transition probability from state r at time $(t-1)$ to state s at time t is given by

$$P(y_{i,t} = s|y_{i,t-1} = r) = \frac{\exp\left(\boldsymbol{x}_{i,t-1,r}^{\top}\boldsymbol{\beta}_{rs}\right)}{\sum_{s'=1}^{Q}\exp\left(\boldsymbol{x}_{i,t-1,r}^{\top}\boldsymbol{\beta}_{rs'}\right)}, \tag{3.1}$$

where $\boldsymbol{x}_{i,t-1,r}$ is a vector of explanatory variables for the i-th policyholder who was in state r at time $(t-1)$, with its first element set equal to one to reflect an intercept term, and the quantity $\boldsymbol{\beta}_{rs}$ denotes the vector of regression coefficients quantifying the effect of the explanatory variables on the transition probability

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

from state r to state s. Note the dependence of explanatory variables on the state of origin reflects the idea that one may wish to, whether based on *a-priori* knowledge or exploratory data analysis, include different sets of explanatory variables for the transitions with different states of origin. Indeed, we adopt this approach in the LGPIF application, for example, see Table 3. At the same time, the dependence on r also reflects the construction of an expanded set of covariates, that is, interaction terms between each of explanatory variable and a factor variable indicating the state of origin. In traditional customer churn analysis, it is assumed that the probability of churn from $(t-1)$ to t depends only on the explanatory variables $\boldsymbol{x}_{t-1}$ at time $(t-1)$ (see Guillen *et al*., 2003; Brockett *et al*., 2008). In this paper, we further allow $\boldsymbol{x}_{i,t-1,r}$ in Equation (3.1) to include (values of) explanatory variables previous to $(t-1)$, for example, in the LGPIF application these include time-variant predictors such as the ratio of total premium of this to previous year minus one $(\boldsymbol{x}_{i,t-1,r}/\boldsymbol{x}_{i,t-2,r}-1)$; see Table 2.

When building a MLR model as in Equation (3.1), we need to decide on a reference state or reference category among all states of destination $s=1,\ldots,Q$. All the coefficients corresponding to this reference state will then be set to zero for reasons of parameter identifiability. For instance, if we choose $s=1$ to be the reference state, then we set $\boldsymbol{\beta}_{r1}=\mathbf{0}$ for all states of origin r, and as such Equation (3.1) for this particular state of destination simplifies to $P(y_{i,t}=1|y_{i,t-1}=r)=\left(\sum_{s'=2}^{Q}\exp\left(\boldsymbol{x}_{i,t-1,r}^{\top}\boldsymbol{\beta}_{rs'}\right)+1\right)^{-1}$. Note this constraint is consistent with logistic regression, and it is also consistent with the interpretation of the coefficients in an MLR model in terms of log-odds. Following on from the above example, we have $\log\left(P(y_{i,t}=s|y_{i,t-1}=r)/P(y_{i,t}=1|y_{i,t-1}=r)\right)=\boldsymbol{x}_{i,t-1,r}^{\top}\boldsymbol{\beta}_{rs}$, for $s=2,\ldots,Q$, meaning $\boldsymbol{\beta}_{rs}$ explicitly quantifies changes in the log-odds of transitioning (from state r) to state s relative to transitioning to state 1.

As discussed above, the first-order Markov assumption is a common one made in the context of (traditional) customer churn analysis. In this paper however, motivated specially by our exploration of the LGPIF data, we consider a second-order MLR model where the probability of being in the current state now depends upon the states in the previous two time points. More generally, a MLR model of order k implies that the probability of being in the current state depends on the k earlier states that is, $P(y_{i,t}|y_{i,1},\ldots,y_{i,t-1})=P(y_{i,t}|y_{i,t-k},y_{i,t-k+1},\ldots,y_{i,t-1})$. Mathematically, a second-order MLR model expands on Equation (3.1) as follows. For the i-th policyholder, the transition probability from states (q,r) at time points $(t-2,t-1)$ to state s at time t is given by

$$P(y_{i,t}=s|y_{i,t-2}=q,y_{i,t-1}=r)=\frac{\exp\left(\boldsymbol{x}_{i,t-1,qr}^{\top}\boldsymbol{\beta}_{qrs}\right)}{\sum_{s'=1}^{Q}\exp\left(\boldsymbol{x}_{i,t-1,qr}^{\top}\boldsymbol{\beta}_{qrs'}\right)}. \tag{3.2}$$

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

The state of origin is now defined by two prior states (q,r) together, while $\boldsymbol{\beta}_{qrs}$ denotes the vector of regression coefficients quantifying the effect of the explanatory variables on the transition from state of origin (q,r) to state s. We use the second-order MLR model given by Equation (3.2) as an example to calculate the likelihood. Following the second-order Markov assumption, that is, $P(y_{i,t}|y_{i,1},\ldots,y_{i,t-1}) = P(y_{i,t}|y_{i,t-2},y_{i,t-1})$, we can write the likelihood function as

$$L = \prod_{i=1}^{N} P(y_{i,1},y_{i,2}) \times P(y_{i,3}|y_{i,1},y_{i,2}) \times \ldots \times P(y_{i,T_i}|y_{i,T_i-2},y_{i,T_i-1}), \quad (3.3)$$

where T_i is the total number of observed states for i-th policyholder. Note that if a policyholder churned during the study period, then we treated the churn state as the last observed state (y_{i,T_i}) for this policyholder. For the second-order MLR model, we focus on those policyholders who have entered their contracts for more than one year and ignore the new policyholders; in practice, these new policyholders are treated differently compared to existing policyholders. Hence, we treat the first two states of each policyholder as given, that is, $P(y_{i,1},y_{i,2}) = 1$. Then, we can replace the second-order conditional probabilities in Equation (3.3) with the second-order MLR formula on the right-hand side of Equation (3.2). The likelihood of the second-order MLR model can be written as

$$L(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{x}) = \prod_{i=1}^{N}\prod_{t=3}^{T_i}\prod_{s=1}^{Q}\prod_{r=1}^{Q-1}\prod_{q=1}^{Q-1}\left(\frac{\exp\left(\boldsymbol{x}_{i,t-1,qr}^{\top}\boldsymbol{\beta}_{qrs}\right)}{\sum_{s'=1}^{Q}\exp\left(\boldsymbol{x}_{i,t-1,qr}^{\top}\boldsymbol{\beta}_{qrs'}\right)}\right)^{\mathbb{I}(y_{i,t-2}=q)\mathbb{I}(y_{i,t-1}=r)\mathbb{I}(y_{i,t}=s)}, \quad (3.4)$$

where indicator variables $\mathbb{I}(y_{i,t-2}=q)$, $\mathbb{I}(y_{i,t-1}=r)$, and $\mathbb{I}(y_{i,t}=s)$ identify the states of origin and destination separately of the transition from $(t-2,t-1)$ to t for i-th policyholder. Note the number of states of origin for q and r is $(Q-1)$ because they excludes the churn (absorbing) state. In Equation (3.4), three successive observations $(y_{i,t-2},y_{i,t-1},y_{i,t})$ of a policyholder can contribute to the likelihood once, which is the reason that fitting a higher-order MLR model requires more data as the order increases. Note that similar to its first-order counterpart, we again need to select a reference state and set the coefficients corresponding to this reference state to zero for reasons of parameter identifiability, for example, $s=1$ and $\boldsymbol{\beta}_{qr1} = \mathbf{0}$ for all (q,r).

By comparing Equations (3.1) and (3.2), we observe that a second-order MLR model greatly expands the states of origin and subsequently accounts for more of policyholders' past history when determining the likelihood of moving to different states or staying in the same state. If such an assumption is appropriate, then Equation (3.2) will generally improve model fitting and predictive power compared to a first-order MLR model. On the other hand, fitting a

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

higher-order MLR model requires more data as the order increases, for example, policyholders who are observed for less than three time points will not be useful for fitting a second-order MLR model, and we require a bigger sample size overall to ensure we have sufficient data within each state of origin to model the transition probabilities effectively.

To better illustrate the difference between using first- and second-order MLR models for multi-state customer churn analysis, we consider an insurance company that offers auto insurance (obliged) and homeowners insurance (optional). A policyholder of this company can reside in three possible states: full-coverage (state 1) where a customer is covered by both auto insurance and homeowners insurance, partial-coverage (state 2) where a customer is covered by auto insurance only, and churn (state 3) meaning a customer drops all insurance coverage and leaves the company. If we adopt the first-order MLR model, then we need only consider two states of origin at time $(t-1)$, namely full and partial-coverages (Figure 1 left panel). Since churn is an absorbing state, we do not model transition probabilities from this state. Following the notation in Equation (3.1), we have $r=\{1,2\}$, $s=\{1,2,3\}$, and $Q=3$. On the other hand, if we adopt the second-order MLR model, then we will need to consider four potential states of origin, depending on whether the policyholder had full or partial-coverage at time points $(t-2)$ and $(t-1)$ (Figure 1 right panel). Following Equation (3.2), we have $q=\{1,2\}$, $r=\{1,2\}$, and $s=\{1,2,3\}$. It is clear from Figure 1 that a larger number of parameters need to be involved in the second-order MLR model: after choosing the reference state, under a first-order Markov assumption, we have four sets of coefficients $\boldsymbol{\beta}_{rs}$ to model, while under a second-order Markov assumption we have eight sets of coefficients $\boldsymbol{\beta}_{qrs}$ to model.

4. Application to data from Wisconsin local government property insurance fund

We illustrate an application of multi-state customer churn analysis to commercial insurance data, using a data set publicly available from the Wisconsin Local Government Property Insurance Fund (LGPIF); see Frees *et al*. (2016), Shi and Yang (2018), Dong *et al*. (2022) among others for previous uses of the LGPIF data in actuarial science. The LGPIF is administered by the Wisconsin Office of the Commissioner of Insurance, with its overarching purpose being to make property insurance available for local government units. The LGPIF offers six coverages for policyholders to choose building and content (BC), contractor's equipment (IM), comprehensive new vehicles (PN), comprehensive old vehicles (PO), collision new vehicles (CN), and collision old vehicles (CO). Since the last four coverages all represent the vehicle coverage, and also to ensure that there is sufficient data to model the transition probabilities reasonably well, we choose to combine the four vehicle coverages as a single coverage that we refer to simply as "Car" coverage. Only BC coverage is compulsory, while IM and Car coverages are optional for a policyholder.

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

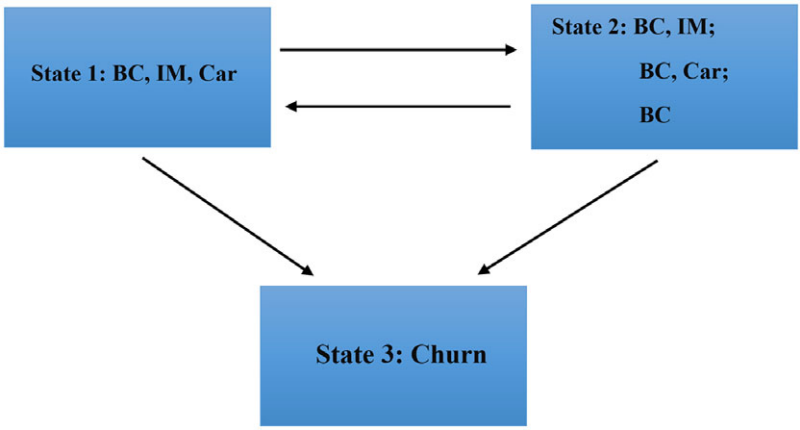


FIGURE 2. A three-state transition diagram of the application to data from LGPIF. BC refers to building and content insurance, IM refers to contractor's equipment insurance, while Car refers to vehicle coverage.

Based on the above, we set up a three-state transition framework for multi-state customer churn analysis, as exemplified by Figure 2. When a policyholder enters a contract, they can take either a full-coverage contract (state 1, which includes BC, IM and Car coverage) or a partial-coverage contract (state 2, which includes BC plus optional IM or Car coverage). Afterwards, if a policyholder does not adjust their combination of coverages, then the contract will be in the same state in the following period, that is, they stay in the same state. Alternatively, a policyholder may drop or add certain coverages without churning, which indicates a transition from state 1 to state 2 or vice versa. Finally, if a policyholder decides to cancel all the coverages in a contract, then the state of the contract will move to state 3, that is, will churn, which is the absorbing state. Note state 2, that is, partial-coverage, actually includes three types of contracts (Figure 2); again, the reason we collect these three forms of contracts into one state is to address data imbalance (a very small number of observations for some transitions) and subsequently to ensure that there is sufficient data to model all transition probabilities reasonably well. We refer the reader to Section 1 of Supplementary Material for a discussion of an expanded five-state transition framework as well as the issue of data imbalance in the LGPIF data.

As part of our exploratory data analysis, Table 1 lists the second-order transition data and corresponding empirical transition probabilities from 2006 to 2013. As discussed in Section 3, although we will not model new policyholders' behaviour, we list these new policyholders' transitions (referred to as state "0" at time $(t-2)$) in Table 1 to offer a complete picture of the data. From time $(t-2)$ to $(t-1)$, there are more observations in the states of origin (1, 1) and (2, 2), that is, staying in the same state, compared to changing states. More importantly, we notice that observations staying in the same states over the past *two* years have higher empirical transition probabilities of staying in

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

TABLE 1.

SECOND-ORDER TRANSITION COUNTS AND EMPIRICAL TRANSITION PROBABILITIES IN PER CENT (IN PARENTHESES) FROM 2006 TO 2013. IN EACH ROW, THE SUM OF COUNTS REPRESENTS THE TOTAL OBSERVATIONS IN THE CORRESPONDING STATE OF ORIGIN.

	State of destination (t)		
State of origin ($t-2, t-1$)	State 1 (full-coverage)	State 2 (partial-coverage)	State 3 (churn)
(0, 1)	416 (92.24%)	22 (4.88%)	13 (2.88%)
(1, 1)	2098 (94.63%)	54 (2.44%)	65 (2.93%)
(2, 1)	38 (88.37%)	5 (11.63%)	0 (0.00%)
(0, 2)	14 (1.72%)	780 (95.59%)	22 (2.70%)
(1, 2)	7 (9.33%)	63 (84.00%)	5 (6.67%)
(2, 2)	27 (0.65%)	4015 (96.31%)	127 (3.05%)

the same state next year. For example, one can compare the empirical transition probabilities based on state of origin (1, 1), that is, having full-coverage in the past two years, with those based on state of origin (2, 1), that is, having partial-coverage and then full-coverage in the past two years. The empirical transition probability of moving from state of origin (1, 1) to state 3, that is, churn, is 2.93%, while the empirical transition probability of moving from state of origin (2, 1) to churn is 0%. Furthermore, the empirical transition probability of moving from state of origin (1, 1) to state 2, that is, partial-coverage, is 2.44%, while the empirical transition probability of moving from (2, 1) to partial-coverage is 11.63%. That is, policyholders who stayed with full-coverage over the past two years are more likely to churn and less likely to move to partial-coverage, relative to policyholders who moved from partial to full-coverage in the past two years. Overall, these initial results suggest that a policyholder's coverage at time $(t-2)$ plays an important role in affecting their behaviour, beyond that seen at time $(t-1)$. With this in mind, we decide to build a second-order MLR model, as exemplified by Equation (3.2), to analyse multi-state customer churn in the LGPIF data.

Finally, it is important to highlight that we did not consider orders higher than two in our application, primarily because the LGPIF did not contain sufficient information to do so, that is, the data did not contain enough time points or observations such that we should feasibly fit a higher-order MLR model effectively. For life insurance companies that do have long-term policyholders' data available (say), then an exploration of subsequent higher-order analysis is warranted.

4.1. Review of variables in customer churn analysis

In Table 2, we summarise the explanatory variables that we will consider in our multi-state customer churn analysis of the LGPIF data based on a combination of our background knowledge and exploratory data analysis (see also

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

TABLE 2.

VARIABLE DESCRIPTION FOR OUR APPLICATION TO THE LGPIF DATA.

Variable name	Description
RateBC	Rate of BC, that is, premium per million BC coverage ($1000 USD)
RateIM	Rate of IM, that is, premium per million IM coverage ($1000 USD)
RateCar	Rate of Car, that is, premium per million Car coverage ($1000 USD)
RatioPremium	Ratio of total premium of this to previous year minus one
TotalCoverage	Value of policy coverage (hundred millions USD)
FrequencyBC	Frequency of BC claim
FrequencyIM	Frequency of IM claim
FrequencyCar	Frequency of Car claim
$\mathbb{I}$(ClaimBC)	Indicator of BC claim occurrence (0 = no claim)
$\mathbb{I}$(ClaimIM)	Indicator of IM claim occurrence (0 = no claim)
$\mathbb{I}$(ClaimCar)	Indicator of Car claim occurrence (0 = no claim)
$\mathbb{I}$(Entity: types)	Indicator of types of entity: Ci = city; Co = county; Sc = school; Mi = misc; To = town; Vi = village, for example, $\mathbb{I}$(Entity: CiSc) = 1 means the entity is either city or school
$\mathbb{I}$(FinaCris)	Indicator of 2008 financial crisis (1 means in 2008)

Note: All the variables take the values at time $(t-1)$, when considering the transition probability from time $(t-1)$ to t.

the discussion at the beginning of Section 4.2). These explanatory variables are commonly used in customer churn analysis in general insurance, and we refer the reader to Section 2 of the Supplementary Material for a discussion on these variables. Note in particular that we included an indicator variable for the 2008 financial crisis, because it was one of the most important global financial events within the training set and statistically highly significant. As part of our exploratory data analysis for the LGPIF data, we examined several other continuous economic growth variables such as GDP and federal rate, but found that none of them were statistically significant predictors for the transition probabilities. Also, note that since the LGPIF is a commercial insurance fund targeting government units as policyholders, then variables commonly seen and used in personal insurance such as age and gender are not present for inclusion in our analysis.

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 refer the reader to Section 3 of the Supplementary Material for a discussion on exogeneity in the context of general insurance and leave the issue of treating certain variables as sequentially exogenous/endogenous for future research. It is also important to point out that not all of the variables will be used in our final model. Moreover, we will allow different sets of explanatory variables to drive the transition probabilities with different states of origin, that is, the dependence on the subscripts (q,r) in the covariates $\boldsymbol{x}_{i,t-1,qr}$ in Equation (3.2). On the explanatory variables themselves, we calculate and use rates (premium per coverage) instead of using premiums directly for each type of coverage, as we found that the premium of a contract

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

TABLE 3.

THE SECOND-ORDER MLR MODEL ESTIMATES WITH 90% CONFIDENCE INTERVAL IN PARENTHESES.

	State of destination			
	Partial-coverage ($s=2$)		Churn ($s=3$)	
	Estimate	CI	Estimate	CI
State of origin: ($q=1, r=1$)				
Intercept	-23.23	(–25.18, –21.29)*	-21.52	(–23.16, –19.87)*
𝕀(Entity: ScToVi)	19.73	(18.74, 20.72)*	17.01	(16.16, 17.87)*
𝕀(Entity: CiMi)	20.66	(19.63, 21.68)*	16.21	(15.25, 17.18)*
𝕀(ClaimBC)	0.20	(–0.34, 0.74)	–0.64	(–1.26, –0.02)*
𝕀(ClaimIM)	-0.47	(–2.18, 1.24)	-27.55	(–27.55, –27.55)*
RateBC	–2.86	(–4.67, –1.05)*	0.30	(–0.76, 1.35)
RateIM	1.37	(0.16, 2.57)*	0.76	(–0.25, 1.77)
RateCar	–0.39	(–0.61, –0.16)*	–0.04	(–0.25, 0.17)
RatioPremium	0.81	(–0.49, 2.12)	0.93	(0.02, 1.83)*
TotalCoverage	–0.88	(–1.48, –0.27)*	–0.66	(–1.31, –0.00)*
State of origin: ($q=2, r=1$)				
Intercept	–15.98	(–16.38, –15.60)*	–16.74	(–16.74, –16.74)*
𝕀(Entity: ScToVi)	14.29	(13.89, 14.69)*	–3.66	(–3.66, –3.66)*
𝕀(ClaimCar)	–11.78	(–11.78, –11.78)*	–2.82	(–2.82, –2.82)*
State of origin: ($q=1, r=2$)				
Intercept	20.87	(19.45, 22.30)*	20.84	(19.43, 22.26)*
𝕀(Entity: ScToVi)	–21.24	(–22.68, –19.81)*	–21.01	(–22.43, –19.59)*
RateIM	1.19	(0.07, 2.31)*	–0.07	(–1.23, 1.09)
State of origin: ($q=2, r=2$)				
Intercept	23.72	(23.41, 24.03)*	20.26	(19.74, 20.58)*
𝕀(Entity: CiScToVi)	–18.72	(–19.03, –18.40)*	–18.60	(–18.94, –18.27)*
𝕀(FinaCris)	–0.90	(–1.58, –0.20)*	–0.42	(–1.19, 0.35)
FrequencyBC	0.02	(–0.21, 0,25)	–0.71	(–1.16, –0.27)*
RatioPremium	3.25	(0.70, 5.80)*	3.56	(1.00, 6.13)*

Note: *confidence intervals excluding zero.

was highly affected by the coverage size of the contract. Since the size of coverage in commercial insurance is relatively large and varies from year to year, then we want to separately study the effect of rate variation and coverage variation on the transition behaviour of policyholders.

4.2. Second-order MLR model: Estimated effects

We fitted second-order MLR models to the LGPIF data from years 2006 to 2012, and left year 2013 as a holdout set that we will use later on to assess out-of-sample performance in Section 5. We performed exploratory data analysis to assess which explanatory variables to include as part of $\boldsymbol{x}_{i,t-1,qr}$. In particular, we started with all variables listed in Table 2 and chose the final set of covariates as the set of explanatory variables which were statistically

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

significant at the 10% level for at least one transition probability (either to partial-coverage or to churn, where full-coverage was set to be the reference state). Note from Table 2, there were only a handful of observations with states of origin equal to (2, 1) and (1, 2). As such, we restricted the number of explanatory variables to two for the transitions with these two states of origin, so as to avoid (excessive) overfitting. We used the `multinom` function from `R` package `nnet` (Venables and Ripley, 2002) to fit the second-order MLR model. Note the standard errors calculated are non-robust, as the function does not support calculating robust standard errors currently; we refer the reader to Section 4 of Supplementary Material for the same analysis using Stata 17, which also returns robust standard errors.

Table 3 presents the estimated coefficients and corresponding 90% confidence intervals for the explanatory variables included in the fitted second-order MLR model. From the estimated MLR model, we observe that both the set of important explanatory variables and the resulting effect sizes differed greatly depending on the state of origin. This in turn reflected a notion that the second-order states of origin play an important role in distinguishing between different transition probabilities, and we discuss the results for each of the four states of origin separately below.

For policyholders with full-coverage in the previous two time points, that is, state of origin equal to (1, 1), there was statistically clear evidence of an association between the likelihood of transitioning to partial-coverage and the three coverage rates. In particular, a higher rate of IM was related to a higher transition probability to partial-coverage, while higher rates of BC and Car were associated with a lower transition probability to partial-coverage. Both the occurrence of BC claim and the occurrence of IM claim significantly reduced the likelihood of churn, while the occurrence of BC claim suggested that the customer was likely to drop some coverage (moved to the partial state). That is, we found that different claim types exhibited different effects on transition probabilities, which is an interesting new insight relating to the role of claims in customer churn analysis. The slopes for TotalCoverage suggested that big buyers (policyholders with large coverages) were less likely to move to either partial-coverage or churn compared to small buyers. Finally, policyholders who experienced larger total premium increases in the previous two years (i.e., positive values of RatioPremium) were significantly more likely to churn the next year, although there was no statistically clear evidence of a similar effect on the likelihood of transitioning to partial-coverage.

For policyholders who went from partial-coverage to full-coverage over the previous two time points, that is, state of origin equal to (2, 1), the occurrence of Car claim was found to significantly increase the probability of staying in full-coverage. Also, the three coefficients of the transition to churn were all negative, which suggested the predicted probabilities of this type of transition would always be zero. This result was not surprising given such transitions were not found in the training set in the first place; see the transition from (2, 1) to churn in Table 1.

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

For policyholders who dropped from full-coverage to partial-coverage over the previous two time points, that is, state of origin equal to (1, 2), there was statistically clear evidence of an effect of a higher rate of IM on the transition probability to partial-coverage, and moreover, the magnitude of this effect was similar to that of policyholders with state of origin (1, 1).

Finally, for policyholders with partial-coverage over the previous two time points, that is, state of origin equal to (2, 2), a large number of BC claims was found to be associated with a significantly lower probability of churn (but no effect on the likelihood of transitioning to partial-coverage). Also, policyholders who experienced total premium increases in the previous two years were significantly less likely to obtain full-coverage in the following year. Finally, we found that the 2008 financial crisis only presented a statistically clear negative effect for this particular group of policyholders, as it strongly decreased their likelihood of retaining partial-coverage in the following year.

For all policyholders, we found that entity type played a statistically significant role in driving the transition probabilities, although this differed depending on the precise state of origin. For example, the entity being equal to either school, town, or village, that is, $\mathbb{I}(\text{Entity: ScToVi}) = 1$ was found to be associated with a significantly higher probability of churn when the state of origin was (1, 1), but a significantly lower probability of churn when the state of origin was (2, 1), compared to the corresponding reference entity type.

Finally, it should be noted that several of the statistically significant effects estimated from the MLR model were distinctly large (Table 3). This potentially reflected a level of overfitting by the second-order MLR model to the LGPIF data. For instance, the estimated pair of effects for Entity(ScToVi) when the state of origin was (1, 2) was approximately -21. This large negative coefficient was attributed to the fact that in the LGPIF training data, when the entity type was equal to a school, town, or village, all policyholders with a state of origin (1, 2) retained full-coverage.

The large estimated effects in this application arise more as a feature of the data rather than an actual problem with the MLR model. That is, the sample size of the LGPIF data was not overly large, and its time horizon (2006–2012) was not long enough to observe more actual transitions. Importantly, we expect this data problem to be neither common nor serious for most multi-state customer churn analyses, given the majority of insurance companies commonly have (much) more accumulated years of longitudinal data. From a modelling perspective, it is possible to apply regularisation penalties, for example, the lasso, ridge or bias-correction-type penalties, to shrink these coefficients away from large values and hence potentially avoid an excessive degree of overfitting (Madigan *et al*., 2005), while also ensuring that overall conclusions in terms of statistically significant effects are retained. We leave such extensions of the MLR model for future research, as the main focus of this paper is on illustrating MLR models as a modelling technique for multi-state customer churn analysis. For readers interested in the goodness-of-fit tests

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

of the fitted MLR model, we provide the results of these in Section 5 of the Supplementary Material.

In summary, for the LGPIF data, we have found that premium information, claim experience, and contract information play statistically significant roles in driving multi-state customer transition probabilities. It is important to highlight that the marketplace we examined in the article has not been previously examined in the actuarial customer churn literature, that is, the LGPIF is a special line of commercial insurance with its overarching purpose being to make property insurance available for local government units. Government units differ from personal customers, this is a smaller marketplace with fewer insurance providers, and there may not be readily available alternatives which government customers can choose from (the original motivation for establishing the fund). This lack of availability will certainly affect policyholders' decisions to leave the fund, suggesting that the results from this paper may not be consistent with other studies that have focused on personal insurance.

At the same time, our analysis offers a more complete picture of the different roles that factors play in determining customer behaviour beyond simply whether policyholders churn or not. Indeed, understanding the drivers of customer transitions behaviour can assist insurance companies with marketing campaigns, for example, pushing customers to concentrate all their insurance policies in the same company. For example, the estimation result of RatioPremium in Table 3 suggests that a small discount (leading to a negative value of RatioPremium) may attract policyholders to stay with the insurance company or even purchase additional coverages. An open question for future research is how to use the multi-state customer churn analysis to identify target customers for marketing campaigns and design optimal marketing strategies for insurance companies. Note also that the effects and scope of exploratory variables considered depends on the second-order state of origin, which in turn sheds light on the importance of a higher-order analysis.

4.3. Second-order MLR model: Visualisation

We examine two types of figures which can be used to better understand the effects of explanatory variables in MLR models, namely conditional effect figures and average marginal effect figures. Briefly, a conditional effect plot quantifies the estimated effect for an individual policyholder (micro-perspective), while an average marginal effect plot quantifies the effect for an entire population of policyholders (macro-perspective).

A conditional effect is based on varying a single predictor (the focal variable) while fixing all the other variables at some values, such that it quantifies the effect of the focal variable and ignores how it correlates (and thus might hence) with other variables in practice. As a result, in the context of multi-state customer churn analysis, conditional effect plots are useful for interpreting the model results from the perspective of an individual policyholder. The insurer can draw conditional effect figures by setting the "condition" to reflect the

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

characteristics of a target policyholder, and thus understand the role a focal variable plays in triggering this policyholder's transition behaviour. This in turn can assist the insurer, say, in developing a targeted strategy to maintain a policyholder's loyalty. By default, when varying the focal variable in a conditional effect plot, we can fix all numerical variables at their mean values, and fix all factor variables at their reference levels (say).

To illustrate the use of a conditional effect plot, we consider the second-order MLR model fitted above to the LGPIF data and target a city entity (policyholder) with partial-coverage over the past two years, that is, state of origin equal to (2,2). In this case, there are only four exploratory variables affecting transition probabilities for this policyholder; see Table 3. Suppose we want to specifically study how the ratio of total premiums affects this policyholder's transition behaviour. From the resulting conditional effect plot (Figure 3 left panel), we observe that, generally speaking, if that target policyholder experiences a larger total premium increase in the previous two years (i.e., positive values of RatioPremium), then the likelihood of transitioning to full-coverage and even maintaining partial-coverage decreases, as that targeted policyholder is more likely to churn. This full picture is not immediately obvious when we only examine the corresponding estimated effects in Table 3: the estimated coefficients of RatioPremium to partial-coverage and to churn are 3.25 and 3.56, respectively. However, as discussed in Section 3 these coefficients explicitly quantify the effect of the ratio of premiums on the log-odds relative to full-coverage and thus only offer an idea of the *relative* changes in transition probabilities. By contrast, from the conditional effect plot we see that, in fact, the probability of transitioning to partial-coverage is concave in shape as a function of RatioPremium, while the probability of churn is monotonically increasing.

As an alternative to conditional effects, the average marginal effect calculates each policyholder's effect of transition probabilities from varying a focal variable and then averages this across the resulting effect estimates among all policyholders. Consequently, this takes into account how variables are naturally correlated with each other as based on the data (Leeper, 2017). Average marginal effect figures are useful for interpreting the fitted MLR model at a population level, that is, macro-perspective for all policyholders. That is, the insurance company can construct average marginal effect plots with different focal variables to understand how the changes affect the transition probabilities across all policyholders currently holding contracts and in turn develop general as opposed to personalised strategies to maintain customer loyalty. As an example of this, and to contrast with the conditional effect plot, we again consider the ratio of premiums as a focal variable, but this time study its average marginal effect among all policyholders in the LGPIF data. The resulting plot is shown in the right panel of Figure 3, from which we see that the probability of transitioning to partial-coverage is (also) concave in shape as a function of premium ratio, while the probability of churn is (also) monotonically increasing. However, the probability of transitioning to partial-coverage

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

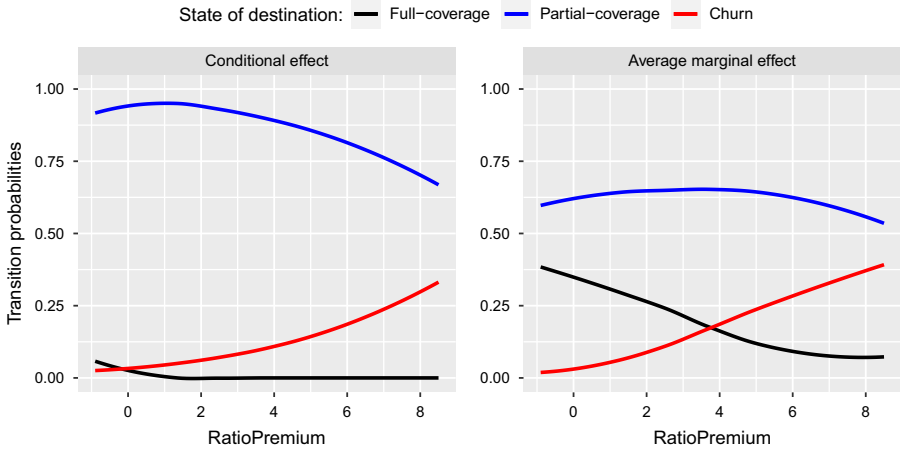


FIGURE 3. The effect of RatioPremium on transition probabilities: conditional effect (left panel) for a city entity with state of origin equal to (2, 2), compared to an average marginal effect (right panel).

(relative to churn) in the average marginal effect plot is consistently lower compared to that in the conditional effect plot. This difference in probability curves indicates the difference of effects from a focal variable to an individual policyholder versus all policyholders. Notice also that in Table 3, the premium ratio was *only* included to model the transition probabilities from states of origin (1, 1) and (2, 2), that is, policyholders who had the same level of coverage across the two previous years. Interestingly, this alone suffices to induce an effect across the broader population of policyholders with the insurance fund, and the average marginal effect plot of Figure 3 reflects this.

4.4. Comparing multi-state and traditional customer churn analysis

As a comparison between multi-state and traditional customer churn analysis for data from the LGPIF, we use an example based on calculating customer lifetime value (CLV). At the end of each year before the renewal date, an insurer may construct models as part of their customer churn analysis, which are then used to calculate policyholders' CLV, the discounted future income stream derived from acquisition, retention, and expansion projections and their associated costs (Gupta *et al*., 2004). As a simple illustration of this, we consider an "average" policyholder for a two-year time horizon, defined as a customer whose numerical explanatory variables are set equal to the mean values from the data, and whose factor variables are set equal to their modal categories. All exploratory variables are then fixed for two years. We further assume that the policyholder has had full-coverage over the past two years, that is, state of origin equal to (1, 1), and we use customer churn

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

TABLE 4.

CLV CALCULATION: STATE-SPECIFIC REVENUE, STATE-SPECIFIC EXPENSE, AND INTEREST RATE.

	E_t	A_t	i_t
Full-coverage	\$34,556	\$28,283	1.78%
Partial-coverage	\$12,268	\$12,904	1.78%

models to assess their behaviour over the next two years. For readers interested in the precise values of exploratory variables, we refer to Section 6 of Supplementary Material. We also point out that all settings in this section were purposefully designed to make the illustration relatively simple yet insightful. In future applications of the proposed MLR model for multi-state customer churn modelling, it is perhaps recommended to adopt richer techniques such as generalised linear models or machine learning techniques to predict claims and interest rates for the CLV calculation.

We calculate the CLV of this average policyholder for the next two years via multi-state customer churn modelling and traditional customer churn modelling. The formula for the CLV of a policyholder is given by $\text{CLV} = \sum_{t=1}^{T} \text{P}_t(E_t - A_t)/(1+i_t)^t$, where E_t and A_t denote the revenue and expenses respectively for a customer at time t, P_t denotes the probability that the customer still has a valid contract up to time t, and i_t is interest rate at time t. For illustrative purposes, we assume E_t is the premium paid in dollar value by the policyholder at time t, and that it is fixed at the state-specific average value in the data set. Moreover, A_t is the expected claim cost in dollar value from the policyholder after paying deductibles, and it is also fixed at the state-specific average value in the data set. We fix i_t at 1.78%, that is, the annual average federal funds rate from 2006 to 2012 in the LGPIF data. Table 4 lists the values of these quantities according to the state of the policyholder. For a policyholder in full-coverage, the revenue is higher than the expense. However, the revenue is lower than the expense for a policyholder in partial-coverage.

For traditional customer churn analysis, we assume the insurance company will build a second-order binary logistic regression (BLR) model to predict the probability of churn for this policyholder in the next two years, while for multi-state customer churn analysis, we use the second-order MLR model fitted in Section 4.2. We refer the reader to Section 7 of Supplementary Material for the illustration of the second-order one-versus-all BLR models in the context of multi-state customer modelling. Importantly, multi-state customer churn analysis focuses on all types of transition among states (see Figure 2), while traditional customer churn analysis only focuses on the transition to churn. Consequently, there are three scenarios to consider if the insurer performs traditional customer churn analysis, but seven scenarios to consider if the insurer performs multi-state customer churn analysis. We refer to Section 8 of Supplementary Material for the visualisation of all scenarios for traditional and multi-state customer churn analysis.

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

TABLE 5.

SCENARIOS OF FUTURE PATH AND CLV CALCULATION: TRADITIONAL CUSTOMER CHURN ANALYSIS VERSUS MULTI-STATE CUSTOMER CHURN ANALYSIS.

Traditional customer churn analysis (second-order BLR)		
Scenario	*Probability (%)*	*Expected present value*
(full, full)	92.60	$2835.7
(full, churn)	3.63	$81.8
(churn)	3.77	$0.0
Sum	100	$2917.4
Multi-state customer churn analysis (second-order MLR)		
Scenario	*Probability (%)*	*Expected present value*
(full, full)	89.15	$2730.3
(full, partial)	1.68	$36.5
(full, churn)	3.58	$80.7
(partial, full)	0.27	$1.5
(partial, partial)	1.32	$−4.1
(partial, churn)	0.20	$−0.5
(churn)	3.80	$0.0
Sum	100	$2844.5

Table 5 presents results of CLV calculations comparing traditional and multi-state customer churn analysis. In the case of the former, the CLV over the three possible scenarios was $2917.4. In the latter, we notice that when the (average) policyholder's future path is to partial-coverage for both years or to partial-coverage followed by churn, the expected present value is negative. This is because when the policyholder takes partial-coverage (in state 2), the revenue is less than the expense; see Table 4. Across the seven possible scenarios, the CLV of the policyholder calculated by multi-state customer churn analysis was $2844.5.

Through Table 5, we can see the difference in calculating CLV between traditional versus multi-state customer churn analysis, which in case ends up being $72.9. In practice, customer lifetime is commonly longer than two years, and so it is likely that an even larger difference in CLV between the two approaches would be observed. Furthermore, while we have considered an "average" policyholder, insurers are typically more interested in those who pay a lot of premiums annually, that is, big buyers. The difference in CLV between the two approaches will again likely be larger for this select group, relative to what we saw in this example. To summarise, because multi-state customer churn analysis considers all possible transitions among states when calculating CLV, and since policyholders in different states naturally have different values for insurers, then we believe such an analysis will generally improve the accuracy of CLV calculations. Ultimately, this is relevant to the profitability of insurance companies (Neslin *et al*., 2006), as it helps them better distinguish valuable customers and trivial customers.

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

5. Assessing out-of-sample performance

Using the LGPIF data, we assessed the out-of-sample prediction performance of the proposed second-order MLR model, and compared it with several other models for predicting customer transition/churn rates. In particular, we also considered a first-order MLR model as formulated in Equation (3.1), along with a first-order and a second-order one-versus-all BLR models (see Section 7 of the Supplementary Material). We also included two non-parametric methods that are widely used in traditional customer churn analysis (He *et al*., 2020), namely gradient boosting machines (GBMs) and support vector machines (SVMs). As both can be applied for multi-class classification problems also, then they can be straightforwardly adapted for use in multi-state customer churn analysis. We refer to Section 9 of Supplementary Material for details of applying GBMs and SVMs to the LGPIF data.

All methods were fitted to LGPIF data from years 2006 to 2012, and out-of-sample validation was performed on data in year 2013. In terms of assessing performance, a widely used evaluation metric in traditional customer churn analysis is error rate $= (1 - \text{accuracy})$, which is defined as the percentage of incorrectly classified observations in the test set (see for example Bolancé *et al*., 2016; Loisel *et al*., 2019; Scriney *et al*., 2020). However, a major disadvantage of error rate is that it is not well suited to imbalanced data with rare events (with imbalanced classes, it is easy to get a low error rate without actually making useful predictions, Morrison, 1969), and this is the case here with the LGPIF data, as exemplified in Table 1. As such, we require alternatives to error rate, and here we consider the area under the receiver operating characteristic curve (AUC, Cortes and Mohri, 2003). The AUC is better suited for assessing classification accuracy in imbalanced data, since it is independent of the choice of classification threshold and avoids choosing the largest estimated probability for the multi-class classification; see Hand and Till (2001) and He and Ma (2013). In this paper, we calculate both multi-class AUCs and the standard binary-class AUCs for model assessment. Additionally, we calculate the top-decile lift, which is a well-established evaluation metric in traditional customer churn research of marketing, and is noted for its ability to target critical customers (Pendharkar, 2009; De Bock and Van den Poel, 2011). We demonstrate how to make use of the top-decile lift for multi-state customer churn analysis.

Note we compare different models' predictive power using widely accepted metrics, namely AUCs and the top-decile lift. But we also acknowledge they are not always the best choices depending on the specific setting. For instance, Guelman *et al*. (2012) suggest that insurers should develop models and metrics to instead identify the target customers who are likely to respond positively to a customer retention campaign. Designing and choosing appropriate metrics in multi-state customer churn analysis for insurers, in order to maximise their utility, is an interesting topic that we leave for future research.

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

5.1. AUC

While AUC is well-established for assessing binary classification, here we consider two other types of AUC for multi-class classification problems, namely micro-AUC and macro-AUC (Wu and Zhou, 2017). Macro-AUC is derived from the receiver operating characteristic curve based on macro-average true positive and false positive rates (TPR and FPR), where the macro-average is defined by calculating TPR and FPR independently for each class and then taking the average. Because macro-average quantities treat all classes equally, then they better reflect the statistics of the smaller classes and so are more appropriate when performance on all the classes is equally important. By contrast, micro-AUC is derived from the receiver operating characteristic curve based on micro-average TPR and FPR. The micro-average pools the contributions of all classes before constructing TPRs and FPRs. As such, micro-average quantities are dominated by the more frequent class.

Table 6 presents the binary AUCs for each type of transition, along with the micro- and macro-AUCs for all six methods considered. Additionally, we consider applying the methods both to the original training set and to a balanced training set constructed through oversampling. Since the main aim of this paper is not about optimising a model's predictive performance, then we used the `upSample` function from `R` package `caret` (Kuhn, 2021) to oversample the training set based on the state of destination in order to deal with data imbalance issue. We ensured the numbers of transitions with different states of destination (full-coverage, partial-coverage, and churn) were equal by adding additional samples to the minority transitions with replacement. In contrast to oversampling, undersampling is also widely used in customer churn analysis when a data set includes a large number of records; see Scriney *et al*. (2020). However, as discussed in Section 4, the LGPIF data do not contain enough time points or observations for this to be relevant. Note for each of the BLR and MLR models, we used explanatory analysis to determine the appropriate regression forms based on the original training set (explanatory variables were chosen if they were statistically significant at the 10% level for at least one transition probability). We then used these forms in the models fitted to the balanced training set. For GBMs and SVMs, we tuned the relevant hyperparameters using the original training set and the balanced training set separately.

For methods fitted to the original training set (Table 6 top half), there is no single-best method across all AUCs. The second-order BLR model had the highest AUC to churn and macro-AUC, which indicates that it performed well at predicting the likelihood of customer churn specifically. If we compare the two second-order models with the two first-order models, then the former was consistently better than the latter in terms of AUC. This further supports the notion that, for MLR models, a second-order Markov assumption leads to more accurate classifications of customer transition. The GBM had the highest AUC to full-coverage, AUC to partial-coverage, and micro-AUC. When we

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

TABLE 6.

OUT-OF-SAMPLE VALIDATION: AUCS OF SIX MODELS (TOP: ORIGINAL TRAINING SET; BOTTOM: BALANCED TRAINING SET).

		AUC (to full)	AUC (to partial)	AUC (to churn)	Macro-AUC	Micro-AUC
Original training set	Second-order MLR	0.990	0.974	0.690	0.885	0.987
	First-order MLR	0.985	0.971	0.651	0.869	0.984
	Second-order BLR	0.990	0.979	**0.691**	**0.887**	0.987
	First-order BLR	0.985	0.977	0.651	0.871	0.984
	SVM	0.993	0.976	0.676	0.882	0.988
	GBM	**0.994**	**0.980**	0.654	0.876	**0.988**
Balanced training set	Second-order MLR	0.989	**0.980**	**0.700**	**0.890**	0.980
	First-order MLR	0.985	0.976	0.648	0.870	0.984
	Second-order BLR	0.989	0.979	0.696	0.888	0.978
	First-order BLR	0.985	0.977	0.645	0.869	**0.984**
	SVM	0.972	0.880	0.532	0.795	0.907
	GBM	**0.991**	0.975	0.643	0.870	0.982

compare methods fitted in the balanced training set (Table 6 bottom half), the second-order MLR model had the highest AUC to partial-coverage, AUC to churn, and macro-AUC, while GBM had the highest AUC to full-coverage. Again, if we compare the two second-order models with the two first-order models, then the former had higher AUCs than the latter in all cases except micro-AUC.

Lastly, when comparing methods across the original and balanced training sets, we observe that the second-order models tended to achieve higher AUC to churn in the balanced training set. However, the overall prediction performance of the other four models became worse in the balanced training set compared to the original training set. This suggests that, at least for the LGPIF data, balancing the data set for second-order models had the potential to improve their prediction performance.

5.2. Top-decile lift

Top-decile lift is commonly used in customer churn analysis to focus on the most critical group of customers and their churn risk (see e.g., Neslin *et al*., 2006; Holtrop *et al*., 2017; Devriendt, Berrevoets and Verbeke, 2021). Broadly speaking, those with the highest 10% predicted churn rates in a test set present the riskiest customers and subsequently an ideal segment for targeting a customer retention programme. Top-decile lift is then calculated as the proportion of churners in this riskiest segment, divided by the proportion of churners in the whole test set. It thus acts as a practical measure to compare different methods for predicting customer churn by assessing their relative abilities to identify the set of riskiest customers. Put another way, a predictive method with a high

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

TABLE 7.

OUT-OF-SAMPLE VALIDATION: TOP-DECILE LIFTS OF SIX MODELS (TOP: ORIGINAL TRAINING SET; BOTTOM: BALANCED TRAINING SET).

		TDL (to full)	TDL (to partial)	TDL (to churn)
Original training set	Second-order MLR	**3.09**	1.50	0.87
	First-order MLR	**3.09**	1.51	0.87
	Second-order BLR	**3.09**	1.51	0.87
	First-order BLR	**3.09**	1.51	0.87
	SVM	3.03	1.51	**1.74**
	GBM	**3.09**	**1.53**	**1.74**
Balanced training set	Second-order MLR	**3.06**	1.51	**2.61**
	First-order MLR	**3.06**	1.51	1.74
	Second-order BLR	**3.06**	1.51	2.17
	First-order BLR	**3.06**	1.51	1.30
	SVM	2.97	1.50	0.00
	GBM	**3.06**	**1.53**	1.74

top-decile lift gives confidence to insurers that they can still reach a majority of all potential churners by targeting the top 10% riskiest segment only (Neslin *et al*., 2006).

We extend the idea of top-decile lift to multi-state customer churn analysis. That is, we calculate top-decile lift and target the corresponding top 10% riskiest segments for transitions with different states of destination. In detail, for a given state of destination s, we firstly identify the 10% riskiest segment, that is, the policyholders with the highest 10% of predicted transition probabilities to that state of destination. The top-decile lift is then given by $\text{TDL}_s = \hat{\pi}_s^{10\%}/\hat{\pi}_s$, where $\hat{\pi}_s^{10\%}$ is the proportion of the actual transitions to state s in the corresponding 10% riskiest segment, and $\hat{\pi}_s$ is the proportion of the true transitions to state s across the entire test set. A higher value of top-decile lift reflects a better capacity to predict the riskiest policyholders for a particular type of transition.

Table 7 presents the results from calculating top-decile lifts to full-coverage, partial-coverage, and churn in both the original and balanced training sets of the LGPIF data. GBM and SVM had the highest top-decile lift to churn in the original training set, while the second-order MLR model had the highest top-decile lift to churn in the balanced training set. Both first- and second-order MLR models performed equivalently in terms of top-decile lifts to churn in the original training set, but when fitted in the balanced training set, the second-order MLR model performed noticeably better than its first-order counterpart. A similar conclusion can be drawn if we compare first- and second-order BLR models. This again supported the use of the second-order Markov assumption for customer churn analysis in the LGPIF data.

Finally, when we compare Tables 6 and 7, we can see that balancing the training set had the potential to improve a model's out-of-sample predictive

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

performance, especially for predicting transitions to churn. This reflected the fact that, in the original training set, the transitions to churn were in the minority and dominated by transitions to full- and partial-coverage (Table 1). When we balanced the original training set through oversampling, we oversampled the transitions to churn, and so their corresponding weight was increased when fitting models, thus generally improving out-of-sample predictions for this minority class.

6. Conclusion

In this paper, we have proposed the concept of multi-state customer churn analysis for commercial insurance. To study policyholders' behaviour beyond just churn/not churn and across multiple periods, we develop MLR models with a second-order Markov assumption. Through the use of such models, which favour interpretability over predictive power, insurance companies can gain a complete picture of how a policyholder's past states affect their decision to move to the next state, as well as the complex relationships between transition probabilities and various explanatory variables related to premium information, claim information, and contract type. To the best of our knowledge, this is the first paper to employ a second-order MLR model for customer churn research in actuarial science.

We present an application of multi-state customer churn analysis by fitting second-order MLR models to investigate policyholders' behaviour on data from the Wisconsin LGPIF. Empirical analysis demonstrates that explanatory variables play differing roles in triggering different types of transitions and this in turn leads to a more nuanced understanding of how a policyholder's previous history and attributes are associated with their decision to future behaviour. We present conditional and average marginal effect plots to interpret the estimated effects on the probability scale. Insurers can intuitively understand the effect of a focal variable on different types of transitions for a specific policyholder using the former, while the latter allows the study of a broader customer group or population of policyholders. We use the example of an average policyholder to illustrate the difference in calculating customer lifetime value between multi-state and traditional customer churn modelling, the latter based on one-versus-all binary logistic regression (BLR) models. Because policyholders in different states naturally have different values for insurers, we believe that multi-state customer churn analysis will improve the accuracy of customer lifetime value calculation for insurers as a whole.

Although prediction is not the primary aim of this paper as a whole, the second-order MLR model was shown to predict relatively well. Specifically, compared to a first-order MLR model, BLR models, GBMs and SVMs, the proposed second-order models performed strongly in terms of out-of-sample

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

AUCs and top-decile lift. We also find that balancing the data through oversampling the minority groups tended to improve the predictive performance of most methods, although the improvement for second-order models was more substantial compared to their first-order counterparts in the LGPIF data. Overall, our findings open up a new way of perceiving and understanding customer behaviour in the insurance market and shed light on the importance of multi-state customer churn modelling.

The multi-state, multi-period modelling used in this paper presents a new approach to churn modelling, and so naturally, there are several variations that warrant further investigation. First, we have implicitly treated all variables as strictly exogenous, which means that we assume all variables are completely unaffected by the customers' state transitions in the past, present, and future. For future research, variables can be treated as sequentially exogenous, which indicates that current or future customers' state transitions would cause variables to change in the future. For example, premium and claim could be treated as sequentially exogenous variables, an approach that may provide insights into contractual moral hazard and adverse selection. Second, we assume that the effect of each covariate is fixed across all time points. Future research could examine the use time-varying coefficients instead for multi-state customer churn analysis, perhaps building on the dynamic survival model approach of Guillen *et al*. (2012). Allowing for time-varying coefficients would enhances the flexibility of the model and potentially improve prediction performance. However, it would generally require more data to estimate than fixed coefficients models, noting also that the records in our motivating LGPIF data are in discrete time, in contrast to continuous time approach of Guillen *et al*. (2012). Third, the assumption that all policyholders are independent is not only questionable in our application but also in studies of personal homeowners and automobiles where things being insured that is, houses and cars, are close to one another, sharing common environmental hazards such as hail, windstorm, and so forth. Some research has been being done to accommodate these dependencies, such as the spatial autocorrelation approach (Brechmann and Czado, 2014) and the more recent marked point process approach (Shi *et al*., 2022). Both of these, among others, may serve as valuable reference points for relaxing such assumption in future developments of multi-state customer churn modelling.

Fourth, if prediction is the primary goal of a customer churn analysis, then a better-performing classifier can be obtained from a balanced sample with appropriate sampling techniques. Future research could examine the use of bias correction methods such as intercept correction and weighting correction (Lemmens and Croux, 2006), along with regularisation and feature selection techniques (e.g., extending the work of Devriendt *et al*., 2021, to support higher-order MLR models). Fifth, for higher-order MLR models, the value of order can be regarded as a hyperparameter that (also) needs to be tuned for optimising the predictive performance. In practice though, building

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

higher-order models (whether they be parametric or more machine-learning based techniques) ideally requires a large data set with many time points, and demands close collaboration between academia and industry. Finally, there are a plethora of advanced multi-class classification models available in the machine learning literature, such as convolutional neural networks and XGBoost, which have already been used in traditional customer churn analysis (Loisel *et al*., 2019; He *et al*., 2020). Applying these to build higher-order models for multi-state customer churn analysis presents an important and challenging avenue for future research.

Acknowledgments

FKCH was supported by an Australian Research Council Discovery Early Career Research Award DE200100435.

Supplementary Material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/asb.2022.18.

References

Alai, D.H., Arnold, S. and Sherris, M. (2015) Modelling cause-of-death mortality and the impact of cause-elimination. *Annals of Actuarial Science*, **9**(1), 167–186.

Antonio, K., Godecharle, E. and Van Oirbeek, R. (2016) A multi-state approach and flexible payment distributions for micro-level reserving in general insurance. Available at SSRN 2777467.

Bettonville, C., d'Oultremont, L., Denuit, M., Trufin, J. and Van Oirbeek, R. (2021) Matrix calculation for ultimate and 1-year risk in the semi-Markov individual loss reserving model. *Scandinavian Actuarial Journal*, **2021**(5), 380–407.

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

Brant Carson, Christine Korwin-Szymanowska, O. L. (2017) Mckinsey Australia small commercial insurance consumer survey. URL: https://www.mckinsey.com/industries/financial-services/our-insights/sme-insurance-in-australia-a-market-ripe-for-change

Brechmann, E. and Czado, C. (2014) Spatial modeling. In Predictive Modeling Applications in Actuarial Science: Volume 1, Predictive Modeling Techniques, p. 260.

Brian Briggs, E.C. (2021) U.S. property & casualty and title insurance industries - 2020 full year results, Technical report, National Association of Insurance Commissioners. URL: https://content.naic.org/

Brockett, P.L., Golden, L.L., Guillen, M., Nielsen, J.P., Parner, J. and 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.

Buchardt, K. (2014) Dependent interest and transition rates in life insurance. *Insurance: Mathematics and Economics*, **55**, 167–179.

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

BUCHARDT, K., MØLLER, T. and SCHMIDT, K.B. (2015) Cash flows and policyholder behaviour in the semi-Markov life insurance setup. *Scandinavian Actuarial Journal*, **2015**(8), 660–688.

CHRISTIANSEN, M.C., ELING, M., SCHMIDT, J.-P. and ZIRKELBACH, L. (2016) Who is changing health insurance coverage? Empirical evidence on policyholder dynamics. *Journal of Risk and Insurance*, **83**(2), 269–300.

CORTES, C. and MOHRI, M. (2003) AUC optimization vs. error rate minimization. *In Advances in Neural Information Processing Systems*, Vol. 16.

DE BOCK, K.W. and VAN DEN POEL, D. (2011) An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction. *Expert Systems with Applications*, **38**(10), 12293–12301.

DEVRIENDT, F., BERREVOETS, J. and VERBEKE, W. (2021) Why you should stop predicting customer churn and start using uplift models. *Information Sciences*, **548**, 497–515.

DEVRIENDT, S., ANTONIO, K., REYNKENS, T. and VERBELEN, R. (2021) Sparse regression with multi-type regularized feature modeling. *Insurance: Mathematics and Economics*, **96**, 248–261.

DONG, Y., FREES, E.W. and HUANG, F. (2022) Deductible costs for bundled insurance contracts. Available at SSRN 4020299.

FONG, J.H., SHAO, A.W. and SHERRIS, M. (2015) Multistate actuarial models of functional disability. *North American Actuarial Journal*, **19**(1), 41–59.

FREES, E.W. (2004) *Longitudinal and Panel Data: Analysis and Applications in the Social Sciences*. Cambridge, UK: Cambridge University Press.

FREES, E.W., BOLANCÉ, C., GUILLEN, M. and VALDEZ, E.A. (2021) Dependence modeling of multivariate longitudinal hybrid insurance data with dropout. *Expert Systems with Applications*, **185**, 115552.

FREES, E.W., LEE, G. and YANG, L. (2016) Multivariate frequency-severity regression models in insurance. *Risks*, **4**(1), 4.

GUELMAN, L., GUILLÉN, M. and PÉREZ-MARN, A.M. (2012) Random forests for uplift modeling: an insurance customer retention case. *International Conference on Modeling and Simulation in Engineering, Economics and Management*, pp. 123–133. Springer.

GUILLEN, M., NIELSEN, J.P., SCHEIKE, T.H. and PÉREZ-MARN, A.M. (2012) Time-varying effects in the analysis of customer loyalty: A case study in insurance. *Expert Systems with Applications*, **39**(3), 3551–3558.

GUILLEN, M., NIELSEN, J. and PÉREZ-MARN, A. (2009) Cross buying behaviour and customer loyalty in the insurance sector. *EsicMarket*, **132**, 77–105.

GUILLEN, M., PARNER, J., DENSGSOE, C. and 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*, pp. 465–490. World Scientific.

GÜNTHER, C.-C., TVETE, I.F., AAS, K., SANDNES, G.I. and BORGAN, Ø. (2014) Modelling and predicting customer churn from an insurance company. *Scandinavian Actuarial Journal*, **2014**(1), 58–71.

GUPTA, S., LEHMANN, D.R. and STUART, J.A. (2004) Valuing customers. *Journal of Marketing Research*, **41**(1), 7–18.

HABERMAN, S. and PITACCO, E. (2018) *Actuarial Models for Disability Insurance*, Abingdon, Oxfordshire: Routledge.

HAND, D.J. and TILL, R.J. (2001) A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learning*, **45**(2), 171–186.

HE, H. and MA, Y. (2013) *Imbalanced Learning: Foundations, Algorithms, and Applications*. Hoboken, New Jersey: John Wiley & Sons.

HE, Y., XIONG, Y. and TSAI, Y. (2020) Machine learning based approaches to predict customer churn for an insurance company. *2020 Systems and Information Engineering Design Symposium (SIEDS)*, pp. 1–6. IEEE.

HOLTROP, N., WIERINGA, J.E., GIJSENBERG, M.J. and VERHOEF, P.C. (2017) No future without the past? predicting churn in the face of customer privacy. *International Journal of Research in Marketing*, **34**(1), 154–172.

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

HÜRLIMANN, W. (2015) A simple multi-state gamma claims reserving model. *International Journal of Contemporary Mathematical Sciences*, **10**(2), 65–77.
KUHN, M. (2021) *caret: Classification and Regression Training*. R package version 6.0-88. URL: https://CRAN.R-project.org/package=caret
KWON, H.-S. and JONES, B.L. (2006) The impact of the determinants of mortality on life insurance and annuities. *Insurance: Mathematics and Economics*, **38**(2), 271–288.
KWON, H.-S. and JONES, B.L. (2008) Applications of a multi-state risk factor/mortality model in life insurance. *Insurance: Mathematics and Economics*, **43**(3), 394–402.
KWON, H.-S. and NGUYEN, V.H. (2019) Analysis of cause-of-death mortality and actuarial implications. *Communications for Statistical Applications and Methods*, **26**(6), 557–573.
LEEPER, T.J. (2017) Interpreting regression results using average marginal effects with r's margins. *Reference Manual*, **32**.
LEMMENS, A. and CROUX, C. (2006) Bagging and boosting classification trees to predict churn. *Journal of Marketing Research*, **43**(2), 276–286.
LOISEL, S., PIETTE, P. and TSAI, C.-H.J. (2021) Applying economic measures to lapse risk management with machine learning approaches. *ASTIN Bulletin: The Journal of the IAA*, **51**(3), 839–871.
MADIGAN, D., GENKIN, A., LEWIS, D.D. and FRADKIN, D. (2005) Bayesian multinomial logistic regression for author identification. *AIP Conference Proceedings*, Vol. 803, pp. 509–516. American Institute of Physics.
MIN, H. and ZHOU, G. (2002) Supply chain modeling: Past, present and future. *Computers & Industrial Engineering*, **43**(1–2), 231–249.
MORRISON, D.G. (1969) On the interpretation of discriminant analysis. *Journal of marketing research*, **6**(2), 156–163.
NESLIN, S.A., GUPTA, S., KAMAKURA, W., LU, J. and MASON, C.H. (2006) Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research*, **43**(2), 204–211.
NORBERG, R. (2013) Optimal hedging of demographic risk in life insurance. *Finance and Stochastics*, **17**(1), 197–222.
OKINE, A.N.-A., FREES, E.W. and SHI, P. (2022) Joint model prediction and application to individual-level loss reserving. *ASTIN Bulletin: The Journal of the IAA*, **52**(1), 91–116.
OKONGWU, U., LAURAS, M., FRANÇOIS, J. and DESCHAMPS, J.-C. (2016) Impact of the integration of tactical supply chain planning determinants on performance. *Journal of Manufacturing Systems*, **38**, 181–194.
ORR, J. (2007) A simple multi-state reserving model. *ASTIN, Colloqiua Orlando edition*. URL: https://www.actuaries.org/ASTIN/Colloquia/Orlando/Papers/Orr.pdf
PENDHARKAR, P.C. (2009) Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services. *Expert Systems with Applications*, **36**(3), 6714–6720.
PRITCHARD, D. (2006) Modeling disability in long-term care insurance. *North American Actuarial Journal*, **10**(4), 48–75.
SCRINEY, M., NIE, D. and ROANTREE, M. (2020) Predicting customer churn for insurance data. *International Conference on Big Data Analytics and Knowledge Discovery*, pp. 256–265. Springer.
SHI, P., FUNG, G.M. and DICKINSON, D. (2022) Assessing hail risk for property insurers with a dependent marked point process. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. **185**(1), 302–328.
SHI, P. and YANG, L. (2018) Pair copula constructions for insurance experience rating. *Journal of the American Statistical Association*, **113**(521), 122–133.
SIMCHI-LEVI, D., KAMINSKY, P., SIMCHI-LEVI, E. and SHANKAR, R. (2008) *Designing and Managing the Supply Chain: Concepts, Strategies and Case Studies*. New York: Tata McGraw-Hill Education.
STAUDT, Y. and WAGNER, J. (2018) What policyholder and contract features determine the evolution of non-life insurance customer relationships? *International Journal of Bank Marketing*. **36**(6), 1098–1124.
VENABLES, W.N. and RIPLEY, B.D. (2002) *Modern Applied Statistics with S*, 4th ed. New York: Springer. ISBN 0-387-95457-0. URL: https://www.stats.ox.ac.uk/pub/MASS4/

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press

WU, X.-Z. and ZHOU, Z.-H. (2017) A unified view of multi-label performance measures. *International Conference on Machine Learning*, pp. 3780–3788. PMLR.
YEO, A.C., SMITH, K.A., WILLIS, R.J. and BROOKS, M. (2001) Modeling the effect of premium changes on motor insurance customer retention rates using neural networks. *International Conference on Computational Science*, pp. 390–399. Springer.
ZHU, Z., LI, Z., WYLDE, D., FAILOR, M. and HRISCHENKO, G. (2015) Logistic regression for insured mortality experience studies. *North American Actuarial Journal*, **19**(4), 241–255.

YUMO DONG (CORRESPONDING AUTHOR)
Research School of Finance, Actuarial Studies and Statistics,
College of Business and Economics, Australian National University,
Canberra, ACT 2601, Australia
E-Mail: yumo.dong@anu.edu.au

EDWARD W. (JED) FREES
School of Business, University of Wisconsin-Madison, Madison,
WI 53706, USA
Research School of Finance, Actuarial Studies and Statistics,
College of Business and Economics, Australian National University,
Canberra, ACT 2601, Australia
E-Mail: jfrees@bus.wisc.edu

FEI HUANG
School of Risk and Actuarial Studies, UNSW Business School,
UNSW Sydney, NSW 2052, Australia
E-Mail: feihuang@unsw.edu.au

FRANCIS K. C. HUI
Research School of Finance, Actuarial Studies and Statistics,
College of Business and Economics, Australian National University,
Canberra, ACT 2601, Australia
E-Mail: francis.hui@anu.edu.au

https://doi.org/10.1017/asb.2022.18 Published online by Cambridge University Press