Received: 16 October 2022 | Revised: 24 May 2023 | Accepted: 26 May 2023

DOI: 10.1111/jori.12436

ORIGINAL ARTICLE

Journal of Risk and Insurance

Improving risk classification and ratemaking using mixture-of-experts models with random effects

Spark C. Tseung[1] | **Ian Weng Chan**[1] | **Tsz Chai Fung**[2] | **Andrei L. Badescu**[1] | **X. Sheldon Lin**[1]

[1]Department of Statistical Sciences, University of Toronto, Toronto, Ontario, Canada

[2]Department of Risk Management and Insurance, Georgia State University, Atlanta, Georgia, USA

Correspondence
Spark C. Tseung, Department of Statistical Sciences, University of Toronto, Ontario Power Bldg, 700 University Ave, 9th Floor, Toronto, ON M5G 1Z5, Canada.
Email: spark.tseung@mail.utoronto.ca

Funding information
Natural Sciences and Engineering Research Council of Canada, Grant/Award Numbers: RGPIN 284246, RGPIN-2017-06684

Abstract

In the underwriting and pricing of nonlife insurance products, it is essential for the insurer to utilize both policyholder information and claim history to ensure profitability and proper risk management. In this paper, we apply a flexible regression model with random effects, called the *Mixed Logit-weighted Reduced Mixture-of-Experts*, which leverages both policyholder information and their claim history, to categorize policyholders into groups with similar risk profiles, and to determine a premium that accurately captures the unobserved risks. Estimates of model parameters and the posterior distribution of random effects can be obtained by a stochastic variational algorithm, which is numerically efficient and scalable to large insurance portfolios. Our proposed framework is shown to outperform the classical benchmark models (Logistic and Lognormal GL(M)M) in terms of goodness-of-fit to data, while offering intuitive and interpretable characterization of policyholders' risk profiles to adequately reflect their claim history.

KEYWORDS

mixture-of-experts, random effects, ratemaking, risk classification, variational inference

1 | INTRODUCTION

In the underwriting and pricing of nonlife insurance products, policyholders' information, or *covariates*, is typically a useful indicator of their risk level. Using automobile insurance as an example, driver's age has been empirically observed as an important factor on both accident rates and frailties (see, e.g., Kelly & Nielson, 2006; Zhang et al., 1998). Similarly, there usually exists certain dependence structure between claim frequency/severity and a driver's annual mileage (the average distance driven per year), see, for example, Bailey and Simon (1960), Vickrey (1968), Edlin (1999), and Lemaire et al. (2016). Such information is often leveraged by the insurer to make better decisions in *risk classification* and *ratemaking*: the former categorizes policyholders into relatively homogeneous groups with similar risk profiles, while the latter determines a premium to be charged for insurance protection.

For new policyholders, risk classification and ratemaking are usually done on an a priori basis, whereby only policyholder covariates are utilized. A widely used approach is to incorporate the covariates as regressors in the Generalized Linear Models (GLMs) for modeling claim frequency and/or severity, see, for example, McCullagh and Nelder (1989), De Jong and Heller (2008), and Ohlsson and Johansson (2010). However, a priori risk classification and ratemaking may fail to capture certain unmeasurable or unobserved risk factors, for example, the aggressiveness when driving, which cannot be reflected by covariates alone (see, e.g., Antonio & Beirlant, 2007; Antonio & Valdez, 2012; Denuit et al., 2007). These unmeasurable or unobserved risk factors are an additional source of heterogeneity among seemingly homogeneous policyholders who have very similar (or even exactly the same) covariate information. Still, one may reasonably assume that such latent risk factors can be reflected by and observed from policyholders' claim history, that is, riskier policyholders tend to have a higher number of claims and/or larger severities. As time goes by, the insurer gains additional, up-to-date insights into the policyholder's risk profile by observing their claim history, including both frequency and severity. At policy renewal, the insurer may decide to update their decision of risk classification and ratemaking on an a posteriori basis, whereby both policyholder covariates and their individual claim history are utilized. A widely known and used approach for a posteriori risk classification and ratemaking is the Bonus-Malus System (BMS), which categorizes policyholders into risk groups with appropriate premia based on their claim counts in the previous policy year (see, e.g., Denuit et al., 2007; Lemaire, 1995, for an introduction to BMS).

For the sake of profitability and risk management, it is essential for the insurer to design a good framework for a posteriori risk classification and ratemaking. As mentioned above, a priori information alone may be insufficient to accurately capture latent, heterogeneous risks of individual policyholders, which could lead to mispricing (i.e., overpricing safer policyholders while underpricing riskier ones) and potentially large losses for the insurer. By combining information from both covariates and claim history for a posteriori risk classification and ratemaking, the insurer may be able to offer competitive pricing for low-risk policyholders, while also appropriately charging high-risk policyholders so that their claims are expected to be covered to ensure the insurer's profitability in the long run. Besides, on a portfolio level, the insurer gains additional insights into the risk segmentation and the categorization of policyholders with similar risk profiles, which may be particularly helpful for risk management purposes such as identifying and ceding losses from high-risk policyholders to avoid tail risks (see, e.g., Chapados et al., 2008).

Consequently, the problem of a posteriori risk classification and ratemaking has led to an abundance of literature from actuarial researchers (see Section 2 for a detailed review). Due to the inherent differences between a priori information and latent risk factors (e.g., observable vs. unobservable, typically fixed vs. potentially time-varying, etc.), they are usually analyzed and modeled

with different methodologies. While regression models have been the predominant method for incorporating a priori information such as policyholder covariates, *random/mixed effects* seem to be a popular approach for modeling heterogeneous latent risk factors which are manifested in the policyholder's claim history. In short, latent risk factors of individual policyholders are assumed to be random variables, rather than fixed values, generated from some unknown distribution which is typically chosen to be normal. In contrast, policyholder information is treated as *fixed effects* because they are typically known in advance and are usually fixed. In this formulation, the most important assumption is that policyholder-level random effects are shared across multiple policy years for the same individual, which creates certain dependency structure between past and future claim experiences. As a result, the claim history of policyholders can be utilized by the insurer for a posteriori risk classification and ratemaking in the upcoming policy year. More specifically, the problem of a posteriori risk classification and ratemaking is effectively transformed into the following two problems to be solved simultaneously: (i) the estimation of regression coefficients of the fixed effects, and (ii) the inference of the posterior distribution of random effects given the policyholder's claim history. In the case of normal random effects, this is equivalent to finding the posterior mean and standard deviation. A classical example is the addition of normal random effects into GLM which results in the Generalized Linear Mixed Models (GLMMs), which has been widely used in statistical problems such as longitudinal data analysis (see, e.g., Diggle et al., 2002; Fitzmaurice et al., 2012), as well as previous works on a posteriori risk classification and ratemaking (see also Section 2).

In this paper, we propose to apply a flexible regression modeling framework, called the *Mixed Logit-weighted Reduced Mixture-of-Experts (Mixed LRMoE)*, to the problem of a posteriori risk classification and ratemaking. On a high level, our work builds upon the *Logit-weighted Reduced Mixture-of-Experts (LRMoE)* framework which uses policyholder covariates as regressors in a flexible and interpretable finite mixture structure for modeling insurance losses (see Section 3.1 for an overview). The Mixed LRMoE is first introduced Fung and Tseung (2022) to incorporate random effects into the LRMoE model. As a general modeling framework, the Mixed LRMoE is flexible enough to resemble any complex characteristics inherited from any mixed effects models, including the joint distribution, the regression pattern, the random intercept, and the random slope, to an arbitrary degree of accuracy (see the discussion of *denseness* in Section 3.3 and Fung & Tseung, 2022). Such theoretical flexibility renders the Mixed LRMoE a powerful tool for modeling complex underlying structures typically observed in insurance data sets (e.g., multimodality and heavy tails), which is essential for accurately describing different policyholders' claim behaviors.

In the context of a posteriori risk classification and ratemaking, we propose to add policyholder-level random effects which are shared across different policy years. By calibrating our model on both policyholder information and their claim history, we gain insights into the impacts of both fixed and random effects on the distribution of claim frequency and/or severity, which are then leveraged for a posteriori risk classification and ratemaking. While our treatment of the fixed and random effects is similar to that in many previous works on the same problem, our proposed methodology provides an alternative approach to this classical problem and offers additional insights into policyholders' risk profiles given their claim history. Using a real automobile insurance data set, we empirically investigate and compare the performance of Mixed LRMoE against various benchmark models for the problem of a posteriori risk classification and ratemaking (see Section 5). The Mixed LRMoE is shown to outperform the benchmark models (e.g., Logistic and Lognormal GL(M)M) in terms of goodness-of-fit to data, while offering fair and interpretable risk classification and ratemaking which adequately reflect policyholders' claim history. We also provide a brief

discussion on the economic and business implications of applying the Mixed LRMoE model in practice. Besides addressing the problem of a posteriori risk classification and ratemaking from a practical perspective, we also provide technical details of a stochastic variational Expectation–Conditional–Maximization (ECM) algorithm for the simultaneous estimation of model parameters and inference of the posterior distribution of random effects (see Section 4), which complements the theoretical development in Fung and Tseung (2022). The algorithm is also numerically efficient and scalable, which greatly boosts the applicability of our proposed Mixed LRMoE framework to large, multiyear insurance portfolio, as well as to more general modeling problems with one or multiple random effects.

The remainder of this paper is organized as follows. Section 2 contains a short literature review for previous works on a posteriori risk classification and ratemaking, as well as estimation methods for random effects models. Section 3 provides an overview of the LRMoE model and introduces the Mixed LRMoE in the general formulation and its adaptation for the problem of a posteriori risk classification and ratemaking. Then, Section 4 develops a stochastic variational ECM algorithm for estimating model parameters and inferring the posterior distribution of random effects. Next, Section 5 contains an application of our proposed framework on a real insurance data set. Finally, Section 6 concludes with a brief discussion and outlook for future research directions. In the Supporting Information, Appendix A contains technical details of the stochastic variational ECM algorithm, and Appendix B presents two simulation studies which aim to numerically illustrate and examine the proposed estimation algorithm.

2 | LITERATURE REVIEW

In this section, we review the existing literature on two fronts: the methodological development on a posteriori risk classification and ratemaking, and various algorithms for parameter estimation in the presence of random effects. We also briefly address how the present paper relates to and differs from previous works.

2.1 | A posteriori risk classification and ratemaking

The use of claim history for a posteriori risk classification and ratemaking is a classical problem which has been studied in depth in the actuarial literature. Early works in credibility theory, such as Bühlmann (1967), Norberg (1979), and Bühlmann and Gisler (2005), assume some common parameters underlying the distribution of insurance losses. One uses the observed claim history to infer the posterior distribution of the parameters, which then yields the posterior distribution of future losses given the history. As for the widely used BMS mentioned in Section 1, a comprehensive introduction can be found in, for example, Lemaire (1995) and Denuit et al. (2007). On the basis of the claim history (typically the number of claims in the year before policy renewal), policyholders are (re)classified into one of a number of prespecified risk classes according to certain transition rules, whereby each risk class corresponds to a premium relativity which reflects the level of risk. However, classical formulations of credibility theory (e.g., greatest accuracy credibility) and BMS (e.g., rate tables based solely on the claim counts in the previous policy year) do not consider covariate information, which is usually deemed as important indicators of policyholders' risk characteristics. To this end, there has been an abundance of literature that aims to apply more sophisticated statistical models, which typically involve a regression component, to the problem of a

posteriori risk classification and ratemaking. Most notably, random effects have been a popular choice for modeling the temporal dependence between past and future claim behavior. For example, many authors have considered adding random effects in GLM which results in GLMM, see, for example, Dionne and Vanasse (1989), Dionne and Vanasse (1992), Pinquet (1998), Frangos and Vrontos (2001), Boucher and Denuit (2006), and Antonio and Beirlant (2007), whereby the posterior distribution of random effects given claim history is used for prediction. Another important consideration is the dependence structure between multiple coverages which is common in automobile insurance, see, for example, Pinquet (1998), Gómez-Déniz et al. (2008), Boucher et al. (2009), Gómez-Déniz (2016), and Tzougas and di Cerchiara (2021) for using shared random effects to model such dependence. Other researchers have also investigated the potential dependence random effects on policyholder covariates (e.g., Boucher & Denuit, 2006, or dynamic random effects, e.g., Bolancé et al., 2007). Besides, while some works mainly focus on claim frequency alone, many researchers have also attempted to incorporate claim severity and its dependence structure with frequency, for example, Ni et al. (2014), Park et al. (2018), Oh et al. (2020), and Oh et al. (2022). Furthermore, to overcome certain restrictive assumptions in GLM, finite mixture models have recently become popular in a posteriori risk classification and ratemaking for more flexible and accurate modeling of claim frequency and severity, as used in Bermúdez and Karlis (2012), Tzougas et al. (2014), Tzougas et al. (2018), and Tzougas and di Cerchiara (2021).

Similar to many papers cited above, the Mixed LRMoE model uses policyholder covariates as fixed effects in a regression framework. The addition of random effects introduces dependence between observations across multiple policy years of the same policyholder, from which the posterior distribution of random effects is inferred and then utilized for a posteriori risk classification and ratemaking. Our work also intersects with mixture model-based approaches, such as Tzougas and di Cerchiara (2021), in that the Mixed LRMoE model allows for more flexible and accurate modeling of the loss distribution compared with classical regression models, such as GLM. In the broader class of general mixture-of-experts (MoE) models, our work is closely related to Yau et al. (2003), Ng and McLachlan (2007), and Ng and McLachlan (2014), where random effects are also incorporated to account for heterogeneity observed in real data. However, the Mixed LRMoE presented in this paper has an arguably simpler model structure.

2.2 | Estimation algorithms

Under certain assumptions such as the classical conjugate pairs of prior-posterior distributions, there exist closed-form solutions for model parameters and the posterior distribution of random effects, which also yields nice, closed-form results for a posteriori premium, see, for example, Chap. 13 of Klugman et al. (2012) for the Gamma-Poisson case, and Denuit and Lu (2021) for the Wishart-Gamma case. However, in many moderately complex regression modeling frameworks with random effects, parameter estimation and inference may be challenging due to typically intractable likelihood functions. As a classical approach, one may consider applying the Best Linear Unbiased Predictor (BLUP) procedure for obtaining the realization of random effects, combined with Restricted/Residual Maximum Likelihood for estimating the model parameters, see, for example, Henderson (1973), Henderson (1975), McLean et al. (1991) for Linear Mixed Models, McGilchrist (1994) and McGilchrist and Yau (1995) for GLMMs, and Yau et al. (2003) and Ng and McLachlan (2007) for MoE models. Alternatively, one may choose to estimate the parameters from the marginal likelihood by numerically integrating out the random effects using, for example, the Gauss–Hermite Quadrature (e.g., Pechon et al., 2019;

Pinheiro & Bates, 1995) or the Laplace approximation (e.g., Breslow & Clayton, 1993; Raudenbush et al., 2000). One may also apply Markov Chain Monte Carlo (MCMC) methods (e.g., Booth & Hobert, 1999; Brooks et al., 2011; Zeger & Karim, 1991) for generating samples of random effects from their posterior distribution given the observed data, based on which the posterior of model parameters can also be obtained. A comparison of these methods for models with random effects can be found in Browne and Draper (2006). However, the aforementioned methods may not be suitable for the problem of a posteriori risk classification and ratemaking. For example, when working with large insurance portfolios, it is desirable to develop an algorithm which scales with the number of random effects and the size of data sets, which may be difficult for numerical integration or MCMC methods. Also, it is desirable to obtain posterior distributions, rather than point estimates, of certain quantities of interest (e.g., a posteriori premium based on different premium principles), which are not produced by either BLUP or numerical integration methods.

Therefore, in place of these classical methods, we opt to use variational inference (VI) primarily for its superior speed and scalability for large insurance portfolios. Besides estimating model parameters with computational efficiency, our stochastic variational ECM algorithm also directly produces the approximated posterior distribution of random effects for each individual policyholder, which is key for a posteriori risk classification and ratemaking for future policy years. Further, while VI methods have been widely used in the machine learning community as an alternative to computationally more expensive methods, such as MCMC (Blei et al., 2017), there have been few use cases of VI in the actuarial literature (see, e.g., Gomes et al., 2021; Kim et al., 2022; Kuo, 2020). We hope our paper serves as another example to showcase the potentials of VI methods for analyzing the ever-growing amount of data available for insurance applications.

3 | MODELING FRAMEWORK

In this section, we first give an overview of the LRMoE modeling framework, including model formulation, theoretical properties, implementation, and application in actuarial contexts. Then, we extend the LRMoE model with random effects to account for the temporal dependence across different policy years. Finally, we provide some discussion on the Mixed LRMoE specifically for the application of a posteriori risk classification and ratemaking.

3.1 | Overview of LRMoE

The LRMoE model first introduced in Fung et al. (2019b) is formulated as follows. Let $\boldsymbol{x}_i$ denote a P-dimensional vector of covariates of policyholder i, such as demographic information and vehicle specification. Given $\boldsymbol{x}_i$, the policyholder is classified into one of g latent risk classes by the logit *gating function*

$$\pi_j(\boldsymbol{x}_i; \boldsymbol{\alpha}) = \frac{\exp\left(\boldsymbol{\alpha}_j^T \boldsymbol{x}_i\right)}{\sum_{j'=1}^{g} \exp\left(\boldsymbol{\alpha}_{j'}^T \boldsymbol{x}_i\right)}, \quad j = 1, 2, ..., g, \tag{1}$$

where $\boldsymbol{\alpha}_j$ is a vector of regression coefficients for latent class j. Within each latent class j, a D-dimensional vector of response variable(s) $\boldsymbol{y}_i$ such as claim frequency and severity is modeled

by an *expert function* $f_j(\boldsymbol{y}_i; \boldsymbol{\psi}_j)$, where $\boldsymbol{\psi}_j$ denotes the parameters of the expert function. Consequently, the likelihood function for a portfolio of n policyholders is given by

$$L(\boldsymbol{\alpha}, \boldsymbol{\Psi}; \boldsymbol{X}, \boldsymbol{Y}) = \prod_{i=1}^{n} \left[\sum_{j=1}^{g} \pi_j(\boldsymbol{x}_i; \boldsymbol{\alpha}) f_j(\boldsymbol{y}_i; \boldsymbol{\psi}_j) \right], \tag{2}$$

where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_2^T, ..., \boldsymbol{\alpha}_g^T)^T$ and $\boldsymbol{\Psi} = \{\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, ..., \boldsymbol{\psi}_g\}$ are the model parameters to estimate given the observed data $(\boldsymbol{X}, \boldsymbol{Y}) = \{(\boldsymbol{x}_i, \boldsymbol{y}_i) : i = 1, 2, ..., n\}$. We assume conditional independence among all dimensions in $\boldsymbol{y_i}$ given the latent class j such that $f_j(\boldsymbol{y}_i; \boldsymbol{\psi}_j) = \prod_{d=1}^{D} f_{jd}(y_{id}; \boldsymbol{\psi}_{jd})$ for $d = 1, 2, ..., D$, where y_{id} is the dth dimension in $\boldsymbol{y}_i$ and f_{jd} is the expert function for y_{id} with parameters $\boldsymbol{\psi}_{jd}$.

The LRMoE model can be viewed as a simplification of the general MoE model (see, e.g., Jordan & Jacobs, 1994), whereby the gating function is restricted to multiple logistic functions and the regression on covariates in the expert functions is eliminated. It is shown in Fung et al. (2019b) that such simplification will not reduce modeling flexibility, provided the expert functions satisfy some mild conditions. In other words, the LRMoE model is capable of achieving the same level of goodness-of-fit as the general MoE with a much simpler model structure. In the meantime, the simplified model structure of LRMoE provides the following intuitive model interpretation in insurance contexts. On the basis of covariates $\boldsymbol{x}_i$ which are indicative of individual risk profiles, policyholders are classified into latent risk groups by a commonly used function for classification problems. Within the same latent group j, the individual risk profiles are naturally assumed to be homogeneous by sharing the same expert function $f_j(\boldsymbol{y}_i; \boldsymbol{\psi}_j)$ whose parameters are independent of policyholder information.

Thanks to its flexibility and interpretability, the LRMoE model has been applied to many actuarial modeling problems. Fung et al. (2019a) used it for modeling correlated claim frequencies of two types of automobile insurance coverage, where the LRMoE mixture of Erlang Count experts is shown to outperform the negative binomial GLM (with and without zero inflation). Fung et al. (2022) discussed fitting LRMoE to censored and truncated data which are commonly encountered when modeling claim severity or reporting delays. The extended model is applied to insurance pricing with policy deductibles and prediction of incurred but not reported claims. In Fung et al. (2022), the LRMoE is further extended to include composite or slicing expert functions which account for multimodal and heavy-tailed distributions. For the implementation of LRMoE, software packages written in R (Tseung et al., 2020) and in Julia (Tseung et al., 2021) are readily available for use, which offer a wide selection of expert functions commonly used for actuarial modeling and utility functions for predictive analysis and model visualization.

As with many mixture models, parameter estimation for LRMoE is done using the ECM algorithm (see, e.g., Dempster et al., 1977; McLachlan & Peel, 2004). Details of the ECM algorithm for LRMoE can be found in the papers cited above. For Mixed LRMoE, we combine the same ECM algorithm with VI methods to deal with intractable marginal likelihood due to the presence of random effects, which will be presented in Section 4.

3.2 | Formulation of Mixed LRMoE

In the context of a posteriori risk classification and ratemaking, it is important to utilize information about policyholders' claim history to make predictions for the upcoming policy years. In effect, one takes advantage of the dependence structure in the claim history across different

policy years generated by the same policyholder. Note that such dependence structure has not been accounted for by the LRMoE model, due to the assumption of independence between observations $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ as indicated by the likelihood function in Equation (2). To incorporate dependence between observations across different policy years, we propose to add policyholder-level random effects to the LRMoE model, which results in the Mixed LRMoE model. In this subsection, we first formulate the Mixed LRMoE in a general setting following Fung and Tseung (2022), and then discuss the special case with only policyholder-level random effects.

Assume each observation $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ is equipped with a vector of random effects $\boldsymbol{w}_i = (w_{i1}, w_{i2}, ..., w_{iL})$, where L is the total number of levels of different random effects. For the lth level of random effect, $l = 1, 2, ..., L$, we assume there are in total S_l factors $\{w_l^{(s)}\}_{s=1,2,\ ...,\ S_l}$, and each observation i is mapped into one of these factors by a known function $c_l(\cdot)$ such that $w_{il} = w_{i'l} = w_l^{(s)}$ if $c_l(i) = c_l(i') = s$ for $s = 1, 2, ..., S_l$. Equivalently, the mapping function $c_l(i)$ can be represented by an S_l-vector $\boldsymbol{t}_{il}$ where exactly the $c_l(i)$th element is one and the others are zero (see also Figure 2 for an example).

Let $\boldsymbol{w} = \{w_l^{(s)}\}_{l=1,2,\ ...,\ L;s=1,2,\ ...,\ S_l}$ denote the collection of random effects across all levels and all factors, which are assumed to be independent across l and s. We also assume their distribution and density functions are prespecified by $\boldsymbol{\Phi}(\cdot)$ and $\boldsymbol{\phi}(\cdot)$ with no extra parameters such that

$$\boldsymbol{\Phi}(\boldsymbol{w}) = \prod_{l=1}^{L}\prod_{s=1}^{S_l}\Phi_l\left(w_l^{(s)}\right) \quad \text{and} \quad \boldsymbol{\phi}(\boldsymbol{w}) = \prod_{l=1}^{L}\prod_{s=1}^{S_l}\phi_l\left(w_l^{(s)}\right), \tag{3}$$

where $\Phi_l(\cdot)$ and $\phi_l(\cdot)$ are, respectively, the distribution and density functions for the lth level of random effects $\{w_l^{(s)}\}_{s=1,2,\ ...,\ S_l}$ for $l = 1, 2, ..., L$. In general, one may specify a priori any distribution for $\boldsymbol{\Phi}(\cdot)$, but a common choice for random effects is the normal distribution. In this paper, we will set each $\Phi_l(\cdot)$ to be a standard normal distribution for $l = 1, 2, ..., L$, which results in a multivariate standard normal distribution for $\boldsymbol{\Phi}(\cdot)$, since all levels of the random effects $\{w_l^{(s)}\}_{s=1,2,\ ...,\ S_l}$ are marginally standard normal and are mutually independent. More discussions on the choice of $\boldsymbol{\Phi}(\cdot)$ are given in Section 3.3.

Similar to the covariates $\boldsymbol{x}_i$, we assume the random effects $\boldsymbol{w}_i$ influences only the gating function. In addition, we assume there are coefficients $\boldsymbol{\beta}_j$, $j = 1, 2, ..., g$, multiplied to the random effects, which serve as scaling factors that also affect the gating functions and add to the modeling flexibility by compensating the lack of parameters in $\boldsymbol{\Phi}(\cdot)$. Consequently, for the Mixed LRMoE model, the gating function, given covariates $\boldsymbol{x}_i$, realization of random effects $\boldsymbol{w}_i$, and parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is specified by

$$\pi_j(\boldsymbol{x}_i, \boldsymbol{w}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{\exp\left(\boldsymbol{\alpha}_j^T\boldsymbol{x}_i + \boldsymbol{\beta}_j^T\boldsymbol{w}_i\right)}{\sum_{j'=1}^{g}\exp\left(\boldsymbol{\alpha}_j^T\boldsymbol{x}_i + \boldsymbol{\beta}_j^T\boldsymbol{w}_i\right)}, \quad j = 1, 2, ..., g. \tag{4}$$

Unlike the gating functions, the expert functions are assumed to be independent of both the covariates $\boldsymbol{x}_i$ and the random effects $\boldsymbol{w}_i$, as illustrated in Figure 1. Note this is the same assumption used in the LRMoE model without random effects. Consequently, given the realization of random effects $\boldsymbol{w}$, the likelihood function of Mixed LRMoE is

$$\tilde{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Psi}; \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{w}) = \prod_{i=1}^{n}\left[\sum_{j=1}^{g}\pi_j(\boldsymbol{x}_i, \boldsymbol{w}_i; \boldsymbol{\alpha}, \boldsymbol{\beta})f_j(\boldsymbol{y}_i; \boldsymbol{\psi}_j)\right], \tag{5}$$

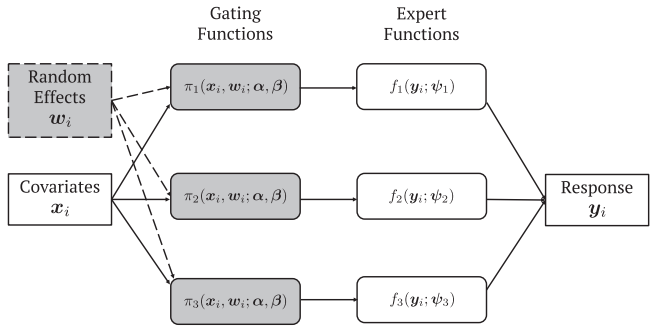

FIGURE 1 Model structure of a three-class Mixed LRMoE model. The shaded boxes indicate the addition of random effects to the original LRMoE model to account for policyholder-level individual risks and temporal dependence among different policy years for the same policyholder. LRMoE, Logit-weighted Reduced Mixture-of-Experts.

while the likelihood with random effects integrated out is given by

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Psi}; \boldsymbol{X}, \boldsymbol{Y}) = \int \tilde{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Psi}; \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{w}) \times \boldsymbol{\phi}(\boldsymbol{w})\, d\boldsymbol{w} = \mathbb{E}_{\boldsymbol{w}\sim\boldsymbol{\phi}(\cdot)}[\tilde{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Psi}; \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{w})], \tag{6}$$

where $d\boldsymbol{w} = \prod_{l=1}^{L} \prod_{s=1}^{S_l} dw_l^{(s)}$ and the subscript of the expectation operator $\mathbb{E}$ indicates the expectation is calculated by integrating out $\boldsymbol{w}$ with respect to $\boldsymbol{\phi}(\cdot)$.

We conclude the introduction of Mixed LRMoE with a remark on its formulation and a comparison with previous works which attempt to incorporate random effects in the general MoE framework. In the statistical literature, Yau et al. (2003) propose a two-component MoE with random effects in both the logit gating function and normal experts. Ng and McLachlan (2007) consider a similar framework but uses Bernoulli experts for a classification problem, while Ng and McLachlan (2014) add random effects only to the expert functions. For the application in insurance contexts, we focus on a special subclass of Mixed MoE model where random effects only influence the latent class probabilities through the gating function, while the expert functions are kept independent of covariates and random effects. Besides possessing the same level of modeling flexibility (see Section 3.3), this simplified model structure leads to an easier implementation of parameter estimation. As will be evident in Section 4, since the estimation procedures of gating and expert functions can be separated to some extent, the Mixed LRMoE model actually allows for more flexible choices and combinations of expert functions which are customized to different modeling problems (see also Section 6). By restricting the random effects to only the gating functions, we are able to develop a unified estimation algorithm which caters for different choices and combinations of expert functions.

3.3 | Denseness property of the Mixed LRMoE

The most important property of the Mixed LRMoE is the *denseness* property, which justifies the flexibility of the proposed model in capturing a broad range of complex multilevel data characteristics. While the theoretical result has been rigorously developed by Fung and Tseung

(2022), we hereby briefly describe and interpret the result without extensive mathematical treatments.

Let $F(\boldsymbol{Y}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Psi}|\boldsymbol{X})$ be the joint distribution function of $\boldsymbol{Y}$ given $\boldsymbol{X}$ under the proposed Mixed LRMoE model, which is given by

$$F(\boldsymbol{Y}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Psi}|\boldsymbol{X}) = \int \prod_{i=1}^{n} \left[\sum_{j=1}^{g} \pi_j(\boldsymbol{x}_i, \boldsymbol{w}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}) F_j(\boldsymbol{y}_i; \boldsymbol{\psi}_j) \right] \times \boldsymbol{\phi}(\boldsymbol{w})\, d\boldsymbol{w}, \tag{7}$$

where $F_j(\boldsymbol{y}_i; \boldsymbol{\psi}_j)$ is the distribution function of $f_j(\boldsymbol{y}_i; \boldsymbol{\psi}_j)$. Also, denote $H(\boldsymbol{Y}|\boldsymbol{X})$ as the joint distribution of $\boldsymbol{Y}$ given $\boldsymbol{X}$ under an arbitrary mixed effects model. Under some mild regularity conditions, Fung and Tseung (2022) prove that for any target mixed effects model $H(\boldsymbol{Y}|\boldsymbol{X})$, there exists a sequence of model parameters $\{(\boldsymbol{\alpha}^{[s]}, \boldsymbol{\beta}^{[s]}, \boldsymbol{\Psi}^{[s]})\}_{s=1,2,\ldots}$ (note that the number of latent risk classes g may increase as s increases) such that $F(\boldsymbol{Y}; \boldsymbol{\alpha}^{[s]}, \boldsymbol{\beta}^{[s]}, \boldsymbol{\Psi}^{[s]}|\boldsymbol{X})$ converges in distribution to $H(\boldsymbol{Y}|\boldsymbol{X})$ uniformly on $\boldsymbol{X}$ as $s \to \infty$. Note that the target mixed effects model $H(\boldsymbol{Y}|\boldsymbol{X})$ may carry very complicated model characteristics, including but not limited to the joint loss distribution (e.g., distributional multimodality and dependence across business lines), the regression link (e.g., nonlinear or interactive influence of policyholder attributes to the losses), the random intercept (e.g., latent impacts to each policyholder), and the random slope (e.g., random effects interact with policyholder attributes). As a result, the *denseness* theorem justifies the versatility of the proposed Mixed LRMoE in simultaneously capturing all these features to an arbitrary degree of accuracy. Moreover, the *denseness* theorem only requires that $\boldsymbol{\Phi}(\cdot)$ is continuous. Hence, one has the freedom to choose any continuous distributions for the random effects without impeding the flexibility of the Mixed LRMoE. Motivated by the computational convenience (see Section 4), we select $\Phi_l(\cdot)$ (Equation 3) to be a standard normal distribution, such that $\boldsymbol{\Phi}(\cdot)$ follows a multivariate standard normal distribution.

3.4 | A posteriori risk classification and ratemaking

In Section 3.2, the Mixed LRMoE modeling framework has been introduced in its most general form. The application of a posteriori risk classification and ratemaking is special case when $L = 1$, that is, there are policyholder-level random effects $\{w_1^{(s)}\}_{s=1,2,\ldots,N_0}$. Consequently, the sample size n is the total number of policy year observations out of N_0 unique policyholders, such that each factor in $\{w_1^{(s)}\}_{s=1,2,\ldots,N_0}$ represents the individual risk of one unique policyholder.

An illustration for one such policyholder with 3 years of claim history is shown in Figure 2. Suppose we would like to conduct a posteriori risk classification and ratemaking for year 3 based on the previous 2 years. The claim history is represented by two rows in the data set, that is, $(\boldsymbol{x}_1, \boldsymbol{w}_1, \boldsymbol{y}_1)$ and $(\boldsymbol{x}_2, \boldsymbol{w}_2, \boldsymbol{y}_2)$, while the future claim to be predicted is represented by yet another row of data $(\boldsymbol{x}_3, \boldsymbol{w}_3, \boldsymbol{y}_3)$. Since these three observations are generated by the same policyholder, $\boldsymbol{w}_1 = \boldsymbol{w}_2 = \boldsymbol{w}_3 = (w_1^{(1)})$, assuming this individual is encoded as the first unique policyholder in the portfolio (thus the superscript for $(w_1^{(1)})$).

Similar to many previous works such as those cited in Section 1, our paper also utilizes random effects for modeling temporal dependence among different policy years of the same

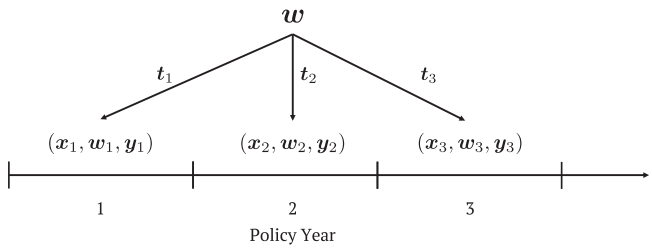

FIGURE 2 Example of a Mixed LRMoE model with $L = 1$ level of random effects on N_0 unique policyholders. A policyholder with 3 years of claim history is represented by three separate yet dependent observations in the data set, where the dependence is modeled by sharing the same factor in the random effect $\boldsymbol{w}$. Assuming this individual is encoded as the first factor $w_1^{(1)}$, the mapping vectors are $\boldsymbol{t}_{11} = \boldsymbol{t}_{21} = \boldsymbol{t}_{31} = (1, 0, 0, ..., 0)$ so that $\boldsymbol{w}_1 = \boldsymbol{w}_2 = \boldsymbol{w}_3 = (w_1^{(1)})$ are the same 1-length vector of random effect. This can be equivalently described by the mapping function $c_1(1) = c_1(2) = c_1(3) = 1$. Note that some elements in the covariates $\boldsymbol{x}_i$ may change over time, such as the policyholder's age. LRMoE, Logit-weighted Reduced Mixture-of-Experts.

policyholder, but we have done so in a slightly different fashion. Many previous papers have proposed mixed models whereby the certain model parameters are shared across different observations. For example, one may assume the claim frequency N_{it} of policyholder i in the tth year follows Poisson(θ_{it}), and then uses the observed data $\{N_{it} : t = 1, 2, ...\}$ to infer the posterior of the intensity parameter. In contrast, our formulation of the Mixed LRMoE treats the random effects $\boldsymbol{w}$ in a similar way as the fixed effects $\boldsymbol{x}_i$, which essentially serve as a regressor in the gating function. Rather than imposing certain changing dynamics on model parameters, the formulation of Mixed LRMoE actually resembles, to a large extent, classical approaches of longitudinal data modeling with random effects, see, for example, Diggle et al. (2002) and Fitzmaurice et al. (2012).

While the denseness property guarantees the modeling flexibility of Mixed LRMoE, some previous works in a posteriori risk classification and ratemaking have investigated other formulations and assumptions of random effects, for example, Boucher and Denuit (2006) considered the potential dependence of random effects on the covariates, and Bolancé et al. (2007) imposed temporal dynamics on the random effects. In contrast, our proposed framework makes more simplified assumptions, that is, independence between random effects and covariates, as well as the same realization of random effects over different policy years. Relaxing these assumptions will lead to different interpretations of the model at various degrees of complexity, which consequently may (or may not) result in significantly different risk classification and ratemaking decisions. In this paper, we will focus on the formulation of Mixed LRMoE presented in Section 3.4 and leave these model extensions for future research.

In practice, a framework for a posteriori risk classification and ratemaking should account for time-varying covariates, such as the policyholder's age. Consider the example in Figure 2 for a policyholder with 2 years of history whereby their age is increasing annually (say, 30 and 31 years old). In the Mixed LRMoE, this policyholder's experience is represented by two separate rows with covariates $\boldsymbol{x}_1, \boldsymbol{x}_2$ and responses $\boldsymbol{y}_1, \boldsymbol{y}_2$. The values for age in $\boldsymbol{x}_1, \boldsymbol{x}_2$ are correspondingly filled with 30 and 31. However, these two rows of data are not independent, because they are describing the same policyholder (thus the same factor in the policyholder-level random effect). In particular, the corresponding random effects are

$\boldsymbol{w}_1 = \boldsymbol{w}_2 = (w_1^{(1)})$, where $w_1^{(1)}$ indicates the unobserved risks of this policyholder. When making predictions for the upcoming policy year, we would represent the same policyholder by yet another row of data, say with covariates $\boldsymbol{x}_3$ and random effects $\boldsymbol{w}_3$. The age in $\boldsymbol{x}_3$ will be valued at 32, while $\boldsymbol{w}_3 = (w_1^{(1)})$ remains the same for the same policyholder, which creates dependency between the past and the future. In short, our framework treats time-varying covariates as regular ones, whose fixed effects are reflected by the corresponding entries in the regression coefficients $\boldsymbol{\alpha}$, while the temporal dependence among claim experiences is accounted for by the random effects $\boldsymbol{w}$.

Another practical issue in a posteriori risk classification and ratemaking is the varying length of available claim history per policyholder, in that an insurer's portfolio rarely remains unchanged as policyholders come in and out of the insured population. Similar to standard mixed effects models, our framework is able to handle imbalanced data, that is, policyholders with varying lengths of claim history. As detailed in the preceding example, policyholders with a longer claim history will have more rows of $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ to represent their claim history, all of which share the same policyholder-level random effects. Naturally, a longer claim history is desirable for obtaining more accurate posterior inference on the random effects, which may yield better results for a posteriori risk classification and ratemaking.

4 | PARAMETER ESTIMATION

In this section, we develop a stochastic variational ECM algorithm for estimating model parameters and for inferring the posterior distribution of random effects for Mixed LRMoE. We first present an overview of VI methods in general, and then provide details of the implementation for Mixed LRMoE with one single type of random effect. Discussion on model identifiability, model selection, and generalization of this algorithm is given at the end of this section.

4.1 | Overview of variational inference

In this subsection, we first provide an overview and motivation of VI methods. We start with the exact posterior distribution of random effects $\boldsymbol{w}$

$$\boldsymbol{p}(\boldsymbol{w}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Psi} | \boldsymbol{X}, \boldsymbol{Y}) \propto \prod_{i=1}^{n} \left[\sum_{j=1}^{g} \pi_j(\boldsymbol{x}_i, \boldsymbol{w}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}) f_j(\boldsymbol{y}_i; \boldsymbol{\psi}_j) \right] \times \boldsymbol{\phi}(\boldsymbol{w}), \tag{8}$$

which may be complicated due to the dependence on both the model parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Psi})$ and the observed data $(\boldsymbol{X}, \boldsymbol{Y})$. To circumvent this numerical challenge, we assume the exact posterior can be reasonably approximated by a *variational distribution* $\boldsymbol{q}(\boldsymbol{w}; \boldsymbol{\Theta})$ where $\boldsymbol{\Theta}$ is the *variational parameters*, which are assumed to be independent of the model parameters and observed data. This produces a numerically more tractable lower bound of the marginal likelihood in Equation (6), also known as the Evidence Lower Bound (ELBO) in the VI literature. More specifically, by taking the logarithm of Equation (6), utilizing the variational distribution, and applying Jensen's inequality, we obtain the following ELBO of the marginal loglikelihood.

$$\begin{aligned}\ell(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Psi}; \boldsymbol{X}, \boldsymbol{Y}) &= \log \int \tilde{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Psi}; \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{w}) \times \boldsymbol{\phi}(\boldsymbol{w})\, d\boldsymbol{w} \\ &= \log \int \tilde{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Psi}; \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{w}) \times \frac{\boldsymbol{\phi}(\boldsymbol{w})}{\boldsymbol{q}(\boldsymbol{w}; \boldsymbol{\Theta})} \times \boldsymbol{q}(\boldsymbol{w}; \boldsymbol{\Theta})\, d\boldsymbol{w} \\ &\geq \mathbb{E}_{\boldsymbol{w} \sim \boldsymbol{q}(\cdot; \boldsymbol{\Theta})}[\log \tilde{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Psi}; \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{w}) + \log \boldsymbol{\phi}(\boldsymbol{w}) - \log \boldsymbol{q}(\boldsymbol{w}; \boldsymbol{\Theta})] \\ &= \mathbb{E}_{\boldsymbol{w} \sim \boldsymbol{q}(\cdot; \boldsymbol{\Theta})}[\log \tilde{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Psi}; \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{w})] - \mathrm{KL}[\boldsymbol{q}(\boldsymbol{w}; \boldsymbol{\Theta}) \| \boldsymbol{\phi}(\boldsymbol{w})] \\ &:= \underline{\ell}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\Theta}; \boldsymbol{X}, \boldsymbol{Y}),\end{aligned} \tag{9}$$

where $\mathrm{KL}[\boldsymbol{q}(\boldsymbol{w}; \boldsymbol{\Theta}) \| \boldsymbol{\phi}(\boldsymbol{w})]$ is the Kullback–Leibler (KL) divergence between the variational posterior $\boldsymbol{q}(\boldsymbol{w}; \boldsymbol{\Theta})$ and the prior $\boldsymbol{\phi}(\boldsymbol{w})$ of random effects.

Instead of directly maximizing the marginal likelihood in Equation (6), we aim to maximize the ELBO $\underline{\ell}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\Theta}; \boldsymbol{X}, \boldsymbol{Y})$ in Equation (9), hoping that the optimal parameters which maximize this lower bound are close to the true optimal parameters which maximize the actual loglikelihood. The main advantage is the tractability of the approximate posterior of random effects $\boldsymbol{w}$, which is essentially specified by parameters $\boldsymbol{\Theta}$ independent of all the other model parameters and observed data. As will be evident in Section 4.2, sampling from the approximated posterior is easier and faster than MCMC methods, since the latter works with a more complex exact posterior and typically requires a burn-in period. This may offer significant numerical efficiency, especially in high-dimensional cases where there are many types of random effects and each type of random effect has many levels. Meanwhile, the obvious trade-off is obtaining only the approximated solutions to the estimated model parameters and the approximated posterior distributions of the random effects. While the goodness of approximation and convergence properties for VI remain an open problem (see, e.g., Blei et al., 2017), our numerical simulations in Supporting Information Appendix B and real data analysis in Section 5 show promising results. This may serve as an empirical evidence for applying VI methods to insurance problems where an approximated solution may be acceptable in the presence of large data sets.

For VI, one needs to specify a family of parametric distributions for the approximated posterior $\boldsymbol{q}(\boldsymbol{w}; \boldsymbol{\Theta})$. In this paper, we follow standard practices and use the *mean-field variational family*, whereby the posterior of latent variables, that is, random effects $\boldsymbol{w}$, is a factorized multivariate normal distribution. More specifically, we assume the posterior of $w_l^{(s)}$ is a normal distribution with mean $\mu_l^{(s)}$ and standard deviation $\sigma_l^{(s)}$ for $s = 1, 2, ..., S_l$ and $l = 1, 2, ..., L$, which are independent across all levels l and all factors s. Mathematically,

$$\boldsymbol{q}(\boldsymbol{w}; \boldsymbol{\Theta}) = \prod_{l=1}^{L} \prod_{s=1}^{S_l} \frac{1}{\sqrt{2\pi \left(\sigma_l^{(s)}\right)^2}} \exp\left[-\frac{1}{2\left(\sigma_l^{(s)}\right)^2} \left(w_l^{(s)} - \mu_l^{(s)} \right)^2 \right]. \tag{10}$$

For notational convenience, we write $\boldsymbol{\Theta} = \{(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)\}_{l=1,2,...,L}$, where $\boldsymbol{\mu}_l = (\mu_l^{(1)}, \mu_l^{(2)}, ..., \mu_l^{(S_l)})^T$ is the posterior mean vector and $\boldsymbol{\Sigma}_l = \mathrm{diag}((\sigma_l^{(1)})^2, (\sigma_l^{(2)})^2, ..., (\sigma_l^{(S_l)})^2)$ the diagonal covariance matrix for the lth level of random effect.

When $L = 1$, given the factorization of likelihood across $s = 1, 2, ..., S_1$, different factors of the same level of random effect are in fact independent, both in the prior and the posterior distribution. Hence, in our application of the Mixed LRMoE with only policyholder-level random effects, the only source of error of VI is the approximation of the exact posterior by a normal distribution. However, when there are multiple types of random effects, especially in the case of certain

dependence structures (e.g., multiple crossed random effects), the independence assumption in the mean-field variational family may create an additional source of error of approximation.

4.2 | A stochastic variational ECM algorithm

With the approach of VI and the choice of the mean-field variational family $\boldsymbol{q}(\boldsymbol{w}; \boldsymbol{\Theta})$, we now develop a stochastic variational ECM algorithm for estimating the model parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Psi})$, as well as inferring the posterior of random effects $\boldsymbol{w}$ represented by the variational parameters $\boldsymbol{\Theta} = \{(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)\}_{l=1,2,\ldots,L}$.

On a high level, our estimation algorithm proceeds in an iterative manner which seeks to conditionally maximize the ELBO in Equation (9) with respect to one set of parameters while keeping others fixed. Consequently, the algorithm will ultimately arrive at a local optimum for the ELBO of the marginal loglikelihood. First, we initialize the model parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Psi})$ using the clusterized method of moments, similar to, for example, Gui et al. (2018). Meanwhile, the variational parameters $\boldsymbol{\Theta}$ can be initialized such that $\boldsymbol{\mu}_l = \boldsymbol{0}$ and $\boldsymbol{\Sigma}_l = \boldsymbol{I}$ for $l = 1, 2, \ldots, L$ (i.e., assuming a multivariate standard normal distribution), which is consistent with standard practices in the VI literature. Then, our algorithm iterates through the following steps until convergence. A detailed description of these steps can be found in Supporting Information Appendix A.1.

E-Step: At iteration $t + 1$, given the current model parameters $(\boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta}^{(t)}, \boldsymbol{\Psi}^{(t)})$ and variational parameters $\boldsymbol{\Theta}^{(t)} = \{(\boldsymbol{\mu}_l^{(t)}, \boldsymbol{\Sigma}_l^{(t)})\}_{l=1,2,\ldots,L}$, we calculate the expectation of the complete-data ELBO, which results in the objective function $Q^{(t+1)}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\Theta}; \boldsymbol{X}, \boldsymbol{Y})$.

CM-Steps:

(i) Given the current values of the variational parameters $\boldsymbol{\Theta}^{(t)}$, we conditionally maximize the objective function in $(\boldsymbol{\alpha}^{(t+1)}, \boldsymbol{\beta}^{(t+1)}, \boldsymbol{\Psi}^{(t+1)})$.
(ii) Given the updated $(\boldsymbol{\alpha}^{(t+1)}, \boldsymbol{\beta}^{(t+1)}, \boldsymbol{\Psi}^{(t+1)})$, find the updated variational parameters $\boldsymbol{\Theta}^{(t+1)} = \{(\boldsymbol{\mu}_l^{(t+1)}, \boldsymbol{\Sigma}_l^{(t+1)})\}_{l=1,2,\ldots,L}$ by optimizing the complete-data ELBO.

In addition to the estimated model parameters $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Psi}})$, our algorithm also yields the variational parameters $\hat{\boldsymbol{\Theta}} = \{(\hat{\boldsymbol{\mu}}_l, \hat{\boldsymbol{\Sigma}}_l)\}_{l=1,2,\ldots,L}$ which completely specify the approximated posterior distribution of random effects $\boldsymbol{w}$. For illustration purposes, Supporting Information Appendix B contains two simulation studies which show our proposed algorithm can recover both model parameters and the realizations of random effects to a reasonable degree. For applications such as a posteriori risk classification and ratemaking, despite no closed-form formulas for various quantities of interests such as the posterior mean of response $\boldsymbol{y}_i$ (see also Section 5), their approximated values can be efficiently calculated by sampling from the variational posterior distribution which is assumed to be multivariate normal. Note the approximated posterior distribution of $\boldsymbol{y}_i$ still retains a similar mixture structure as Equation (7), whereby the integration is now with respect to the approximated posterior of $\boldsymbol{q}(\boldsymbol{w})$ rather than the prior $\boldsymbol{\phi}(\boldsymbol{w})$.

4.3 | Model identifiability and selection

As with many mixture models, certain restrictions are imposed for the Mixed LRMoE to be identifiable when conducting parameter estimation. To avoid label-switching between latent

components (see, e.g., Fung et al., 2019a; Jiang & Tanner, 1999), we fix $\boldsymbol{\alpha}_g = \mathbf{0}$ and $\boldsymbol{\beta}_g = \mathbf{0}$ as vectors of zeros, so the last latent class serves as a reference class. In addition, we fix $\boldsymbol{\beta}_1 = \mathbf{1}$ as a vector of ones to avoid arbitrary scaling of magnitude and switching of positive and negative signs of the random effects $\boldsymbol{w}$. Consequently, we need to estimate the coefficients $\boldsymbol{\beta}$ multiplied to the random effects only when there are at least three latent classes (see the examples in Supporting Information Appendix B).

Model selection when parameters are estimated using VI remains an open problem in general. One may accept the ELBO as a good approximation of the marginal likelihood and use it as the basis of model selection, but this has not been justified in theory (Blei et al., 2017). Other approaches include sequential selection (Sato, 2001), cross validation (Nott et al., 2012), and Generalized Evidence Bounds (Chen et al., 2018). For the purpose of this paper, we take a more practical approach by using the standard train-test split and examining the approximated loglikelihood and ELBO on the test set to obtain a conservative gauge of goodness-of-fit. Examples are given in the real data analysis in Section 5.

5 | REAL DATA ANALYSIS

In this section, we apply the Mixed LRMoE model to a real automobile insurance data set for a posteriori risk classification and ratemaking, and then compare its performance with a number of benchmark models. More specifically, we will investigate whether the Mixed LRMoE model can outperform benchmark models like Logistic and Lognormal GL(M)M and LRMoE without random effects in terms of goodness-of-fit. We will also investigate whether the Mixed LRMoE produces reasonable results for a posteriori risk classification and ratemaking, that is, policyholders who made claims in the past should generally be considered riskier and should be assigned a higher a posteriori premium.

5.1 | Description of data

The data set contains the Bodily Injury (BI) claim history of 76,049 unique policyholders from policy years 2014 to 2019 (330,781 records in total) of a major North American automobile insurer.

Since we are only working with a one-dimensional response, it will be represented by y_i in this section. The description of available covariates $\boldsymbol{x}_i$ and the summary statistics of the response y_i are given in Table 1. We observe the loss distribution has significant zero inflation and a heavy tail in certain policy years. There also seems to be an increasing trend of claim severity over the years, in addition to varying degrees of skewness and kurtosis. The empirical distribution of positive losses is also plotted in Figure 3, which shows slightly different shapes across policy years.

Within the period of 2014–2019, the lengths of policyholders' available claim history vary from 1 year (new contracts) to 6 years (multiple renewals). For the purpose of this section, we will limit ourselves to a subset of policyholders with at least 3 years of claim history, whereby the last available year will be used as a holdout testing set and all preceding years will be used as a training set for model fitting. This filtering step will ensure all policyholders have at least some history (2 years at a minimum) for inferring the distribution of random effects. In a preliminary analysis whereby we use a train-test split based on calendar years (i.e., 2014–2017/

TABLE 1 Overview of real data set.

Covariate	Range	Description
x_{i0}	1	Intercept. Baseline for Female drivers and Rural region.
x_{i1}	{0, 1}	Indicator for Male drivers. Mean is 0.47.
x_{i2}	[16, 99]	Driver's age. Mean is 64 and median is 66.
x_{i3}	[0, 27]	Vehicle age. Mean is 6.5 and median is 6.
x_{i4}	[9.35, 12.82]	Natural logarithm of vehicle price. Mean is 10.28 and median is 10.28.
x_{i5}	[1, 99]	Vehicle's collision rating (an indicator of risk). Mean is 26 and median is 27.
x_{i6}	{0, 1}	Indicator for policies issued in the Capital. Mean is 0.09.
x_{i7}	{0, 1}	Indicator for policies issued in Urban region. Mean is 0.75.

Response y_i		Claim severity						
Year	Claim rate	Mean	SD	Lower Quart.	Median	Upper Quart.	Skewness	Kurtosis
2014	0.0260	10,440	44,715	1355	2688	5667	11	146
2015	0.0247	13,395	57,044	1409	2992	6345	10	120
2016	0.0264	12,363	78,032	1648	3334	6769	19	420
2017	0.0269	11,469	65,311	1719	3471	7175	24	684
2018	0.0248	9,910	38,639	1993	3819	7975	18	397
2019	0.0161	13,194	80,346	2361	4977	8750	22	547
Overall	0.0244	11,723	62,364	1699	3425	7019	21	570

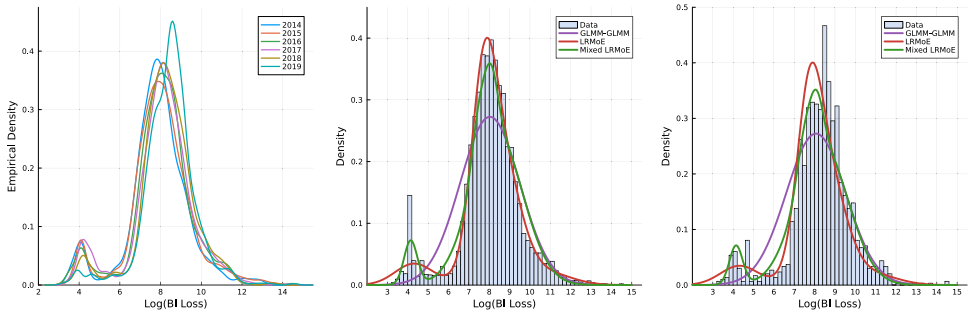

FIGURE 3 Histogram and fitted density of positive claim distribution. (Left) Empirical density of positive losses by policy year. (Middle/right) Training/testing set. Only GLMM–GLMM is shown because all the benchmark models yield very similar results. BI, Bodily Injury; GLMM, Generalized Linear Mixed Model; LRMoE, Logit-weighted Reduced Mixture-of-Experts. [Color figure can be viewed at wileyonlinelibrary.com]

2018 for training and 2018/2019 for testing), the differences in loss distributions (both frequency and severity) across the years make it very difficult to gauge and compare model performances. Our train-test split based on policyholder-level history will also smooth out some of the distributional differences across calendar years. Furthermore, 20% of the policyholders from the training period are randomly selected as the validation set for selecting the number of components in the (Mixed) LRMoE model. In the end, there are 203,579/50,831/75,742

two-year contracts, or 60,594/15,148/75,742 unique policyholders, in the training/validation/testing set, respectively. Overall, 34%/22%/18%/26% of the policyholders have 3/4/5/6 years of claim history before the train-test split, while 11%/12%/13%/64% of the testing set are contracts from year 2016/2017/2018/2019, respectively.

For illustration purposes, we will model the total amount of loss per year. As a benchmark, we will consider various combinations of GLM and GLMM against which we compare the proposed Mixed LRMoE model. For these benchmark models, we assume independence between claim frequency and severity. We use a probability mass $\delta_i(\boldsymbol{x}_i)$ at zero for no occurrence of claims and a continuous distribution $g_i(y_i|\boldsymbol{x}_i)$ for the total loss amount given there is at least one claim. Consequently, using $I_{\{y_i=0\}}$ and $I_{\{y_i>0\}}$ as indicators for the occurrence of claims, the distribution of total loss of policyholder i is given by

$$f(y_i|\boldsymbol{x}_i) = \delta_i(\boldsymbol{x}_i) \times I_{\{y_i=0\}} + (1-\delta_i(\boldsymbol{x}_i))g_{id}(y_i|\boldsymbol{x}_i) \times I_{\{y_i>0\}}, \quad i = 1, 2, ..., n. \tag{11}$$

where both $\delta_i(\boldsymbol{x}_i)$ and $g_i(y_i|\boldsymbol{x}_i)$ may be modeled by either GLM or GLMM. In the case of GLMM, we will add policyholder-level random effects with 60,594 levels which corresponds to the number of unique policyholders in the training data set. For the claim probability, we use the standard Logistic GL(M)M with the logit link function. For the claim severity, we choose the Lognormal GL(M)M with the log link function which shows the best fit to data after some initial experimentation.

For the models to investigate, we will consider (mixed) LRMoE with zero-inflated (ZI) Lognormal expert functions. With the expert functions fixed, we only need to select the number of latent components for both LRMoE and Mixed LRMoE. We have selected a five-component LRMoE and a five-component Mixed LRMoE based on the Akaike Information Criterion (AIC) calculated on the validation data set.

5.2 | Goodness-of-fit

The fitted loglikelihood values of all benchmark models are summarized in Table 2. As expected, on the training set, the GLMM–GLMM model produces the highest loglikelihood since the policyholder-level random effects are used twice. The combinations of GLMM–GLM and GLM–GLMM offer worse fit to data, followed by the GLM–GLM model without any random effects. Meanwhile, all benchmark models perform very similarly on the testing set. Table 3 summarizes the fitting results of the (Mixed) LRMoE models. We see that the Mixed

TABLE 2 Benchmark models for real data analysis.

Benchmark	Claim probability	Log of claim severity	Number of parameters	Training		Testing	
				loglik	AIC	loglik	AIC
GLM–GLM	Logistic GLM	Normal GLM	16	−75,084	150,199	−22,314	44,660
GLM–GLMM	Logistic GLM	Normal GLMM	17	−74,700	149,434	−22,317	44,668
GLMM–GLM	Logistic GLMM	Normal GLM	17	−73,995	148,023	−22,304	44,641
GLMM–GLMM	Logistic GLMM	Normal GLMM	18	−73,611	147,258	−22,306	44,648

Abbreviations: AIC, Akaike Information Criterion; GLM, Generalized Linear Model; GLMM, Generalized Linear Mixed Model.

TABLE 3 (Mixed) LRMoE models for real data analysis.

g	Number of parameters	Training loglik	Training AIC	Validation loglik	Validation AIC	Testing loglik	Testing AIC
LRMoE							
2	14	−74,905	149,839	−19,449	38,925	−22,226	44,481
3	25	−74,774	149,599	−19,415	38,880	−22,183	44,417
4	36	−74,712	149,495	−19,403	38,877	−22,173	44,419
5	47	−74,614	149,323	−19,389	38,873	−22,169	44,432
6	58	−74,578	149,272	−19,386	38,888	−22,169	44,453
7	69	−74,555	149,248	−19,381	38,899	−22,173	44,483
Mixed LRMoE							
2	14	−73,479	146,985	−19,458	38,944	−22,210	44,448
3	26	−72,918	145,889	−19,408	38,868	−22,161	44,374
4	38	−73,844	147,763	−19,397	38,870	−22,153	44,382
5	50	−73,285	146,670	−19,380	38,860	−22,153	44,406
6	62	−72,718	145,559	−19,374	38,873	−22,160	44,444
7	74	−72,561	145,270	−19,379	38,906	−22,153	44,455

Abbreviations: AIC, Akaike Information Criterion; LRMoE, Logit-weighted Reduced Mixture-of-Experts.

LRMoE model offer much better fit to data in terms of loglikelihood on training and testing data sets, and outperforms the LRMoE model without random effects. This demonstrates the flexibility of Mixed LRMoE as well as the advantage of incorporating policyholder-level random effects for more accurate modeling of the loss distribution. As for penalization on model complexity, the AIC values are included for all model candidates in the tables, which also demonstrates the outperformance of the Mixed LRMoE model. Even though the Mixed LRMoE model has a more complex structure in terms of the number of parameters, as will be evident in Section 5.3, this added model complexity greatly improves a posteriori risk classification and ratemaking, which is the ultimate goal in this context.

Besides loglikelihood values, we also look at how each model candidate fits the probability of claim and the distribution of positive losses. For the probability of claim, all model candidates offer very similar fitting performance. On the training period, all models are able to fit the observed claim probability 0.0255 to the fourth decimal place. However, on the testing period where the observed claim probability is 0.0196, all models candidates have produced a slightly higher prediction, ranging from 0.0248 to 0.0253 (or +26% to +29% of relative error), which can be attributed particularly to the lower claim frequency in year 2019 as observed in Table 1. Meanwhile, the (Mixed) LRMoE models have provided a better fit to the distribution of positive losses, as indicated by Figure 3 which compares the fitted densities against the empirical distribution. Most notably, the (Mixed) LRMoE models have successfully captured the multimodality in the distribution of positive losses, while GLM and GLMM only fit a unimodal density to the entire distribution of positive losses, and the LRMoE model without random effects fits slightly worse on smaller claims. Even though our data processing procedures have mixed up different calendar years for the testing period, note there still seems

to be some distributional shift from the training to the testing period (most notably due to year 2019), so the estimated density curves from all model candidates appear to be slightly off to the left.

5.3 | Risk classification and ratemaking

For insurance pricing purposes, it is crucial that policyholders' claim history is adequately incorporated in the calculation of premium at policy renewal. In short, higher risks, as reflected by the occurrence of claim and/or higher claim amounts, should lead to a higher a posteriori premium. In this subsection, we compare the model performance in terms of a posteriori risk classification and ratemaking.

For risk classification, the latent classes in (Mixed) LRMoE models can be naturally interpreted as different clusters of policyholders based on their risk profile. To compare how risk classification is affected by claim history, we categorize all policyholders into two groups: those with at least one claim and those without any claim during the training period, and summarize their latent class probabilities in Table 4. Most notably, with the addition of random effects, the Mixed LRMoE models are able to strongly distinguish risky policyholders who have at least one claim in the past, by assigning almost a much higher probability to the riskiest latent class. Meanwhile, the LRMoE model without random effects only suggests a slight increase in the risky class probability based solely on covariate information, given the independence assumption for observations across different policy years.

Different decisions in a posteriori risk classification will also lead to differences in ratemaking. For a posteriori ratemaking, we calculate the premium for policy renewals in the testing period based on the posterior distribution given the claim history in the training period.

TABLE 4 Comparison of latent classes and the predicted probabilities by claim history.

	LRMoE			Mixed LRMoE		
Risk level	**Class**	**Mean**	**SD**	**Class**	**Mean**	**SD**
Low	1	0.87	15.59	1	0.00	0.00
	2	5.52	189.64	2	1.68	11.03
Medium	3	128.56	686.58	3	11.56	157.36
High	4	830.20	18,512.30	4	732.83	1532.26
	5	1168.66	4689.53	5	1958.62	9756.57

	LRMoE		Mixed LRMoE	
Risk level	**No**	**Yes**	**No**	**Yes**
Low	0.5580	0.5321	0.5216	0.4546
Medium	0.1779	0.1867	0.3213	0.3063
High	0.2641	0.2812	0.1571	0.2391

Note: The first table summarizes the mean and standard deviation of the response by latent class, and we have manually categorized them into three risk levels. The second table compares the predicted latent class probabilities for different groups of policyholders by their claim history (No: no claim during the training period; Yes: at least one claim during the training period), calculated from different models.

Abbreviation: LRMoE, Logit-weighted Reduced Mixture-of-Experts.

For illustration purposes, we only consider the pure premium which is equal to the probability of claim multiplied by the expected positive mean loss amount.

On a higher level, we investigate all policyholders based on the same grouping (with and without claims in the training period). The distributions of the predicted posterior premium are shown in Figure 4 for all model candidates. For models without random effects, that is, GLM–GLM and LRMoE, the predicted distributions of posterior premium for the two groups appear to be highly overlapping, which indicates that fixed effects alone cannot distinguish policyholders based on claim history. For benchmark models with random effects, namely, GLM–GLMM, GLMM–GLM, and GLMM–GLMM, there appear to be some differences between the two groups, whereby some policyholders with claim history will have a higher predicted premium. Most notably, the Mixed LRMoE model shows much larger differences between the distributions of predicted premium, which better captures the riskiness of policyholders reflected by their claim history.

On a more detailed level, Table 5 summarizes the predicted posterior premium, based on the two groups above in addition to the relative size of incurred total losses. We observe that the Mixed LRMoE model, as well as benchmark GLMM–GLM and GLMM–GLMM, heavily penalizes policyholders who have at least one claim, as shown by the additional premium loadings.

For both a posteriori risk classification and ratemaking discussed above, we have primarily focused on differentiating policyholders based on the occurrence of claims and the claim sizes when applicable, whereby the Mixed LRMoE model is shown to have effectively incorporated such information. However, we can still observe the effects of a priori information, that is, policyholder covariates, when determining the a posteriori premium. Most notably, in Figure 4, there is a good level of overlap between the histograms of the predicted premium for people with and without claim history, even for all model candidates with random effects. For example, certain policyholders with claim history (lower end of the orange histogram) would still be charged a lower premium than some policyholders without claim history (upper end of

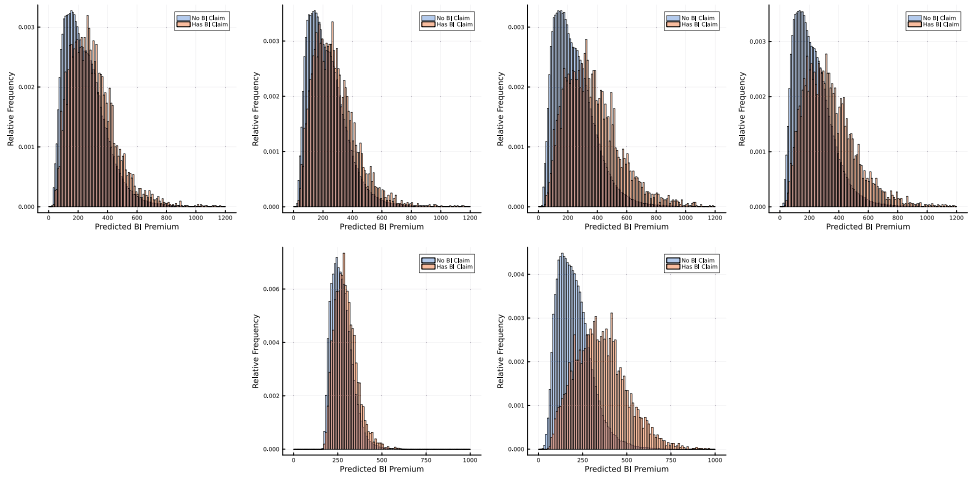

FIGURE 4 Histogram of predicted posterior premium based on different models. Top row, from left to right: GLM–GLM, GLM–GLMM, GLMM–GLM, and GLMM–GLMM. Bottom row, from left to right: LRMoE and Mixed LRMoE. AIC, Akaike Information Criterion; BI, Bodily Injury; GLM, Generalized Linear Model; GLMM, Generalized Linear Mixed Model. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 5 Average of predicted posterior premium based on policyholders' claim history.

	Claim indicator		Claim size		
Model	**No**	**Yes**	**Small**	**Medium**	**Large**
GLM–GLM	249	300 (21%)	283 (14%)	293 (18%)	324 (30%)
GLM–GLMM	229	277 (21%)	234 (2%)	269 (18%)	329 (44%)
GLMM–GLM	247	367 (49%)	347 (40%)	359 (45%)	396 (61%)
GLMM–GLMM	227	340 (50%)	287 (26%)	330 (45%)	403 (77%)
LRMoE	273	295 (8%)	288 (6%)	292 (7%)	304 (12%)
Mixed LRMoE	204	353 (73%)	321 (58%)	337 (65%)	401 (97%)

Note: The cutoff points for positive claim sizes are the 33% and 67% percentiles of its distribution. Percentages in brackets indicate the additional premium loadings compared with policyholders without any claim history, that is, Claim Indicator = No.

Abbreviations: GLM, Generalized Linear Model; GLMM, Generalized Linear Mixed Model; LRMoE, Logit-weighted Reduced Mixture-of-Experts.

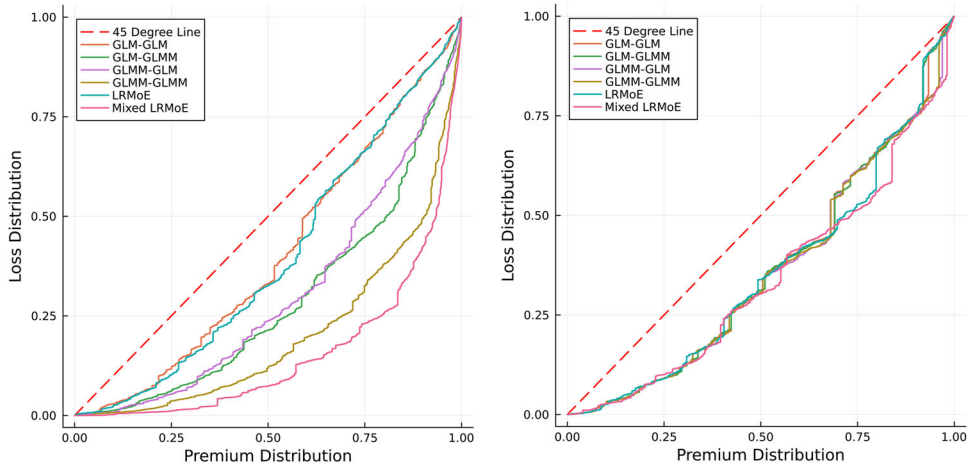

FIGURE 5 Comparison of the Ordered Lorenz Curves generated from various model candidates. (Left/right) Training/testing set. GLM, Generalized Linear Model; GLMM, Generalized Linear Mixed Model; LRMoE, Logit-weighted Reduced Mixture-of-Experts. [Color figure can be viewed at wileyonlinelibrary.com]

the blue histogram), which should be attributed to covariates such as the inherent risk level of certain age groups or the collision rating of a particular group of vehicles.

5.4 | Gini Index

Next, we examine the model performance using the Gini Index as a measurement of adequacy for insurance risk scoring (see, e.g., Frees et al., 2011). We first plot the Ordered Lorenz Curve in Figure 5 for both the training and testing sets, where the x-axis represents the cumulative percentage of premium and y-axis represents the cumulative percentage of the incurred losses

during the training or testing period, In the training set, the Mixed LRMoE model produces an Ordered Lorenz Curve farthest away from the 45° Line of Equality which represents a null model where all policyholders are assigned the same premium. This indicates the Mixed LRMoE yields the highest degree of differentiation of policyholders based on their relative riskiness. Meanwhile, in the testing set, it may be difficult to visually compare the model candidates. Hence, we rely on the Gini Index, calculated as twice the area between the Ordered Lorenz Curve and the Line of Equality, as a measurement of model performance for comparison.

We implement a bootstrapping procedure to obtain the exact distributions and comparisons of Gini Index values (see also Corollary 3 of Frees et al., 2011, for an asymptotic version). More specifically, we obtain 10,000 bootstrapped samples of the training and testing sets, from which the Ordered Lorenz Curves are produced and the corresponding Gini Index values are calculated. We note the potential correlation of Gini Index values produced by all model candidates, since they are all regression models based on the same set of covariates. Consequently, broadly similar policyholders (e.g., low-risk vs. high-risk) will be assigned similar premium values. The difference in performance will stem from a finer differentiation within broadly similar policyholders (e.g., high-risk vs. higher-risk), by better capturing the nonlinear regression relationship and/or the latent, unobserved risks. Table 6 summarizes the bootstrapped Gini Index values of all model candidates, as well as the pairwise comparisons of their differences based on both two-sided and one-sided tests against zero.

In the training set, all model candidates are significantly better than the null model, which is not surprising. The GLMM–GLMM model is the best among the benchmark models, and it also outperforms the LRMoE model by capturing unobserved policyholder-level risks with random effects. The proposed Mixed LRMoE model performs the best, showing a significant margin of outperformance against all other model candidates.

In the testing set, all model candidates are also better than the null. However, when comparing against each other, they perform quite similarly whereby most of the two-sided tests yield insignificant results at $p = 0.20$. If one is interested in determining the outperformance of one model against another, a one-sided test of the positivity of the difference in Gini Index values may also be appropriate. Under this test, the GLMM–GLM model is the best among all benchmark models, while the Mixed LRMoE model may still be considered better than all others by producing a higher Gini Index at least 86% of the time.

A data drift in year 2019 has been previously noted in Table 1 and Figure 3, which comprises 64% of the testing set. To this end, we further compare the model performance by bootstrapping the testing set by policy years 2016–2018 and year 2019. In the former experiment, the proposed Mixed LRMoE offers superior outperformance, assuming the testing set is generated from a distribution similar to years 2016–2018. In the latter, the margin of outperformance by Mixed LRMoE becomes less significant due to the sudden change of loss distribution. However, such unprecedented data drift is outside the scope of what statistical and predictive models can address based on historical data only.

5.5 | Comparison of individual policyholders

Apart from the analysis on a portfolio level, we also investigate the model performance on a policyholder level. In particular, we examine pairs of policyholders with similar covariates but different claim experiences, to investigate how the model candidates determine the a posteriori

TABLE 6 Summary of Gini Index of all model candidates and their differences.

		Difference in Gini Index (row–column)				
Model	**Gini Index**	**GLM–GLM**	**GLM–GLMM**	**GLMM–GLM**	**GLMM–GLMM**	**LRMoE**
Training set						
GLM–GLM	0.1900***	–				
	(0.1112, 0.2660)	–				
		–				
GLM–GLMM	0.3890***	0.1990***	–			
	(0.3315, 0.4476)	(0.1643, 0.2340)	–			
		100%	–			
GLMM–GLM	0.3832***	0.1932***	−0.0057	–		
	(0.3081, 0.4564)	(0.1860, 0.1998)	(−0.0376, 0.0257)	–		
		100%	36%	–		
GLMM–GLMM	0.5643***	0.3743***	0.1753***	0.1811***	–	
	(0.5174, 0.6112)	(0.3310, 0.4186)	(0.1616, 0.1891)	(0.1419, 0.2217)	–	
		100%	100%	100%	–	
LRMoE	0.1938***	0.0038	−0.1952***	−0.1894***	−0.3705***	–
	(0.1232, 0.2634)	(−0.0005, 0.0130)	(−0.2253, −0.1650)	(−0.1997, −0.1793)	(−0.4828, −0.4291)	–
		80%	0%	0%	0%	–
Mixed LRMoE	0.6496***	0.4596***	0.2606***	0.2664***	0.0853***	0.4558***
	(0.5743, 0.7198)	(0.4350, 0.4850)	(0.2192, 0.3031)	(0.2442, 0.2895)	(0.0387, 0.1312)	(0.4291, 0.4829)
		100%	100%	100%	100%	100%

(Continues)

TABLE 6 (Continued)

		Difference in Gini Index (row–column)				
Model	**Gini Index**	**GLM–GLM**	**GLM–GLMM**	**GLMM–GLM**	**GLMM–GLMM**	**LRMoE**
Testing set						
GLM–GLM	0.2766***	–				
	(0.1564, 0.4162)	–				
		–				
GLM–GLMM	0.2745***	−0.0021	–			
	(0.1566, 0.4078)	(−0.0086, 0.0036)	–			
		72%	–			
GLMM–GLM	0.2865***	0.0099*	0.0120	–		
	(0.1602, 0.4389)	(−0.0024, 0.0251)	(−0.0034, 0.0329)	–		
		92%	90%	–		
GLMM–GLMM	0.2848***	0.0082*	0.0103	−0.0017	–	
	(0.1598, 0.4338)	(−0.0029, 0.0207)	(−0.0031, 0.0277)	(−0.0062, 0.0023)	–	
		90%	90%	23%	–	
LRMoE	0.2854***	0.0088	0.0109	−0.0011	0.0006	–
	(0.1539, 0.4227)	(−0.0165, 0.0507)	(−0.0149, 0.0513)	(−0.0376, 0.0499)	(−0.0334, 0.0502)	–
		63%	74%	46%	2%	–
Mixed LRMoE	0.3174***	0.0408*	0.0429*	0.0309	0.0326	0.0320***
	(0.1663, 0.4751)	(−0.0087, 0.0948)	(−0.0084, 0.0971)	(−0.0156, 0.0901)	(−0.0138, 0.0907)	(0.0005, 0.0594)
		94%	94%	86%	88%	98%
Mixed LRMoE	0.3126***	0.0469**	0.0449*	0.0476*	0.0460*	0.0139*

TABLE 6 (Continued)

		Difference in Gini Index (row–column)				
Model	**Gini Index**	**GLM–GLM**	**GLM–GLMM**	**GLMM–GLM**	**GLMM–GLMM**	**LRMoE**
Year 2016–2018	(0.2034, 0.4150)	(−0.0066, 0.0998)	(−0.0088, 0.0976)	(−0.0101, 0.1042)	(−0.0117, 0.1021)	(−0.0055, 0.0321)
		95%	95%	95%	94%	92%
Mixed LRMoE	0.3285***	0.0245*	0.0305*	0.0053	0.0099	0.0385**
Year 2019	(0.1480, 0.5212)	(−0.0177, 0.0577)	(−0.0157, 0.0682)	(−0.0258, 0.0310)	(−0.0223, 0.0352)	(−0.0042, 0.0738)
		89%	91%	67%	77%	95%

Note: Numbers in brackets indicate the 95%-level credible intervals estimated from 10,000 bootstrapped samples. Superscripts ***/**/* indicate the difference is significant at 0.05/0.10/0.20 levels, respectively, under a two-sided test against zero. Numbers below the brackets indicate the proportion of bootstrapped samples where the model in the row outperforms the model in the column under each pair of comparison, which is equivalent to a one-sided test of the difference against zero.

Abbreviations: GLM, Generalized Linear Model; GLMM, Generalized Linear Mixed Model; LRMoE, Logit-weighted Reduced Mixture-of-Experts.

premium for individual policyholders. For brevity, we only retain GLMM–GLMM as the benchmark and compare it with the (Mixed) LRMoE models.

We consider three pairs of policyholders (A_1, A_2), (B_1, B_2), and (C_1, C_2), whereby each pair of policyholders share the exact same covariates but different claim experiences and all of them have 6 years of full history from 2014 to 2019. A_1 and A_2 are both 65-year-old male, drive a 7-year-old vehicle worth of \$40,100 with a collision rating of 33, and purchased their policies in the Urban region. B_1 and B_2 are both 35-year-old female, drive a 6-year-old vehicle worth of \$29,400 with a collision rating of 32, and purchased their policies in the Urban region. C_1 and C_2 are both 40-year-old male, drive an 8-year-old vehicle worth of \$24,800 with a collision rating of 29, and purchased their policies in the Urban region. As for the claim history during 2014–2018, A_1, B_1, and C_1 have no claims, while A_2, B_2, and C_2 have a total claim amount of \$850, \$1950, and \$5704, respectively. Given that 97.9% of policyholders have no claims at all (see Table 1), these positive claims lie at the very tail of the overall loss distribution, with C_2 being close to the 99% percentile. From an a posteriori perspective, A_2, B_2, and C_2 should be considered increasingly riskier than their counterparts.

For these selected policyholders, Table 7 summarizes their a posteriori pure premium values. Since the LRMoE model does not incorporate claim history, each pair of (A_1, A_2), (B_1, B_2), and (C_1, C_2) is given the same premium value which is not reasonable. In contrast, all other models with random effects have produced higher a posteriori premium for A_2, B_2, and C_2, since their claim experiences during the training period are indicative of latent heterogeneous risks unobservable from covariates alone. Most notably, the Mixed LRMoE model has posed very large penalties for policyholders B_2 and C_2, whose a posteriori premium is more than double that of a comparable policyholder without any claim history.

In addition, Figure 6 illustrates their a posteriori predictive distribution for the positive losses of these selected policyholders. For A_2, B_2, and C_2 who have made claims in the past and should be considered riskier, the Mixed LRMoE model has assigned more probability masses on the positive losses, as indicated by the elevated density functions compared with their safer counterparts. Most notably, the Mixed LRMoE model has produced much heavier tails for B_2 and C_2 than those produced by other model candidates, which contributes to the drastic increase in the corresponding a posteriori pure premium compared with B_1 and C_1.

5.6 | Economic and business implications

In the preceding subsections, we have illustrated how the proposed Mixed LRMoE outperforms the benchmark models by providing a superior fit to data, producing reasonable results for a

TABLE 7 A posteriori pure premium values for sample policyholders.

Policyholder	A_1	A_2	B_1	B_2	C_1	C_2
GLMM–GLMM	228	269	359	420	286	408
LRMoE	282	282	337	337	309	309
Mixed LRMoE	225	247	277	566	249	535

Note: The pairs (A_1, A_2), (B_1, B_2), and (C_1, C_2) have the same covariates. During the training period, A_1, B_1, and C_1 have no claims, while $A_1/B_2/C_1$ has a total claim amount of \$850, \$1950, and \$5704, respectively. From an a posteriori perspective, A_2, B_2, and C_2 should be considered increasingly riskier than their counterparts.

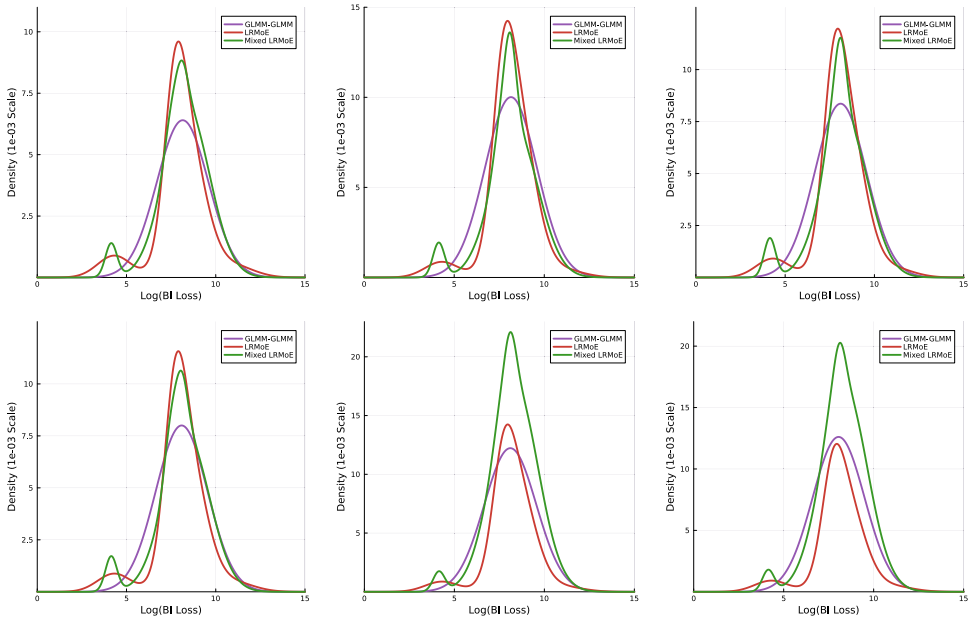

FIGURE 6 A posteriori predictive distributions for positive losses for sample policyholders. Top row, from left to right: A_1, B_1, and C_1. Bottom row, from left to right: A_2, B_2, and C_2. Since the LRMoE model does not consider claim history, the corresponding density functions are the same for each pair of (A_1, A_2), (B_1, B_2), and (C_1, C_2), and may be viewed as a reference to compare the top and bottom rows. GLM, Generalized Linear Model; GLMM, Generalized Linear Mixed Model; LRMoE, Logit-weighted Reduced Mixture-of-Experts. [Color figure can be viewed at wileyonlinelibrary.com]

posteriori risk classification and ratemaking, and adequately differentiating riskier policyholders from safer ones based on their claim history. Now we briefly discuss the economic and business implications of the potential application of the Mixed LRMoE model in practice.

As mentioned in Section 1, a well-designed framework for a posteriori risk classification and ratemaking is crucial for the insurer's profitability and risk management. With a better fit to empirically observed data shown in Section 5.2, the Mixed LRMoE can provide a more accurate description of the overall loss distribution, which lays the foundation for risk classification and ratemaking. Compared with benchmark models such as GLM and GLMM, our proposed model is flexible enough to capture complex data structures such as multimodality and heavy tails, which is particularly helpful for modeling extreme losses generated by risky policyholders. This is also illustrated by the a posteriori risk classification and ratemaking results in Sections 5.3 and 5.5, whereby riskier policyholders with large claims in the past are subject to a much higher posterior premium at policy renewal, while some safer policyholders are rewarded by a lower premium. Consequently, by capturing latent risks manifested in the claim history, policyholders are more appropriately priced (rather than mispriced) according to the Mixed LRMoE model, which results in better risk segmentation as indicated by the improved Gini Index values in Section 5.4. All these advantages of the Mixed LRMoE model will help increase the insurer's profitability and ensure better risk management.

However, in Section 5.3, certain risky policyholders with large claims in the past are very aggressively penalized by the Mixed LRMoE model, as illustrated by the drastic increase in the a posteriori pure premium. From a practical perspective, the insurer can undoubtedly expect

nonrenewal of insurance policies from some of these riskier policyholders. While such nonrenewals will lead to a decrease in premium income (all else held constant), it also comes with the advantage of reduced risk exposures especially in the tail. In the meantime, safer policyholders are rewarded with potential decreases renewal premium, which increases the likelihood of customer retention and could contribute further to the insurer's profitability, since these policyholders are less likely to incur losses after all. Consequently, this may lead to long-term changes in the composition of the insurer's portfolio, as the proportions of safe and risky policyholders are likely to change after a few years, assuming risky policyholders with claim history gradually drop out. While the insurer should constantly monitor their portfolio structure, especially after implementing a new model (whether the Mixed LRMoE or any model in general), we leave the detailed investigation and discussion on such long-term impacts for future research. We also recognize that ours is only an illustrative application of the Mixed LRMoE model for research purposes. In practice, another potential challenge is to properly communicate the a posteriori premium values to policyholders and other stakeholders, especially when policyholders have similar covariates but different claim experiences, as shown by the examples in Section 5.5. This may also have legal and regulatory implications, as well as interesting academic discussions on the fairness of insurance pricing, but we will leave these issues for future research.

6 | CONCLUSION

In this paper, we have proposed to incorporate policyholder-level random effects in a flexible regression framework, called the Mixed LRMoE, which is then applied to the problem of a posteriori risk classification and ratemaking. Although the addition of random effects has resulted in an intractable marginal likelihood function of the model, we have developed a stochastic variational ECM algorithm for efficient estimation of model parameters and inference of the posterior of random effects, which are crucial for updating policyholders' risk profile based on their claim history. Our numerical simulation and real data analysis have demonstrated the potentials of Mixed LRMoE as a powerful tool for more accurate insurance loss modeling and better a posteriori insurance risk classification and ratemaking. While our current work has already shown promising results in an illustrative example, some practical issues remain to be addressed in future research (see Section 5.6). From a technical and modeling perspective, one may consider the following extensions and directions for future work.

- In the current formulation of Mixed LRMoE, all past policy years are equally weighted by sharing the same realization of random effects. A more realistic and general approach is to apply a weighting scheme whereby recent claims are more influential in determining the posterior premium.
- We have taken the approach of modeling the total incurred loss as a mixture of ZI distributions, whereby the dependence between claim frequency and severity is not explicitly specified. An interesting extension is to incorporate such dependence in the (Mixed) LRMoE modeling framework.
- As observed in our numerical study, the shift of loss distributions over different policy years presents another challenge to a posteriori risk classification and ratemaking. This opens up potential research opportunities for modeling frameworks which account for, for example,

temporal trends of claim probability, inflation of claim severity, and more generally, a change of the overall loss distribution.

- While our estimation algorithm enjoys numerical efficiency and has been shown to yield reasonable results both in simulation and real data analysis, it could be worthwhile to investigate the theoretical properties, such as approximation errors and rate of convergence, of VI methods in the class of MoE models as well as the Mixed LRMoE.

ACKNOWLEDGMENTS

The authors thank two anonymous referees, a senior editor, and the editor-in-chief (Joan T. Schmit) for their valuable feedback and suggestions which have greatly improved this paper. They also thank Sebastián Calcetero for his insightful comments. The authors acknowledge the financial support provided by the Natural Sciences and Engineering Research Council of Canada (Andrei L. Badescu: Grant No. RGPIN 284246; X. Sheldon Lin: Grant No. RGPIN-2017-06684).

ORCID

Spark C. Tseung http://orcid.org/0000-0002-1790-0668

REFERENCES

Antonio, K., & Beirlant, J. (2007). Actuarial statistics with generalized linear mixed models. *Insurance: Mathematics and Economics*, *40*(1), 58–76.

Antonio, K., & Valdez, E. A. (2012). Statistical concepts of a priori and a posteriori risk classification in insurance. *AStA Advances in Statistical Analysis*, *96*(2), 187–224.

Bailey, R. A., & Simon, L. J. (1960). Two studies in automobile insurance ratemaking. *ASTIN Bulletin: The Journal of the IAA*, *1*(4), 192–217.

Bermúdez, L., & Karlis, D. (2012). A finite mixture of bivariate Poisson regression models with an application to insurance ratemaking. *Computational Statistics & Data Analysis*, *56*(12), 3988–3999.

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, *112*(518), 859–877.

Bolancé, C., Denuit, M., Guillén, M., & Lambert, P. (2007). Greatest accuracy credibility with dynamic heterogeneity: The Harvey–Fernandes model. *Belgian Actuarial Bulletin*, *7*(1), 14–18.

Booth, J. G., & Hobert, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *61*(1), 265–285.

Boucher, J.-P., & Denuit, M. (2006). Fixed versus random effects in Poisson regression models for claim counts: A case study with motor insurance. *ASTIN Bulletin: The Journal of the IAA*, *36*(1), 285–301.

Boucher, J.-P., Denuit, M., & Guillen, M. (2009). Number of accidents or number of claims? An approach with zero-inflated Poisson models for panel data. *Journal of Risk and Insurance*, *76*(4), 821–846.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, *88*(421), 9–25.

Brooks, S., Gelman, A., Jones, G., & Meng, X.-L. (2011). *Handbook of Markov chain Monte Carlo*. CRC Press.

Browne, W. J., & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, *1*(3), 473–514.

Bühlmann, H. (1967). Experience rating and credibility. *ASTIN Bulletin: The Journal of the IAA*, *4*(3), 199–207.

Bühlmann, H., & Gisler, A. (2005). *A course in credibility theory and its applications* (Vol. 317). Springer.

Chapados, N., Dugas, C., Vincent, P., & Ducharme, R. (2008). Scoring models for insurance risk sharing pool optimization. In *2008 IEEE International Conference on Data Mining Workshops* (pp. 97–105). IEEE.

Chen, L., Tao, C., Zhang, R., Henao, R., & Duke, L. C. (2018). Variational inference and model selection with generalized evidence bounds. In *International Conference on Machine Learning* (pp. 893–902). PMLR.

De Jong, P., & Heller, G. Z. (2008). *Generalized linear models for insurance data*. Cambridge Books.

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

Denuit, M., & Lu, Y. (2021). Wishart-gamma random effects models with applications to nonlife insurance. *Journal of Risk and Insurance*, *88*(2), 443–481.

Denuit, M., Maréchal, X., Pitrebois, S., & Walhin, J.-F. (2007). *Actuarial modelling of claim counts: Risk classification, credibility and bonus-malus systems*. John Wiley & Sons.

Diggle, P., Diggle, P. J., Heagerty, P., Liang, K.-Y., Zeger, S. (2002). *Analysis of longitudinal data*. Oxford University Press.

Dionne, G., & Vanasse, C. (1989). A generalization of automobile insurance rating models: The negative binomial distribution with a regression component. *ASTIN Bulletin: The Journal of the IAA*, *19*(2), 199–212.

Dionne, G., & Vanasse, C. (1992). Automobile insurance ratemaking in the presence of asymmetrical information. *Journal of Applied Econometrics*, *7*(2), 149–165.

Edlin, A. (1999). Per-mile premiums for auto insurance. National Bureau of Economic Research.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). *Applied longitudinal analysis* (Vol. 998). John Wiley & Sons.

Frangos, N. E., & Vrontos, S. D. (2001). Design of optimal bonus-malus systems with a frequency and a severity component on an individual basis in automobile insurance. *ASTIN Bulletin: The Journal of the IAA*, *31*(1), 1–22.

Frees, E. W., Meyers, G., & Cummings, A. D. (2011). Summarizing insurance scores using a Gini Index. *Journal of the American Statistical Association*, *106*(495), 1085–1098.

Fung, T. C., Badescu, A., & Lin, X. S. (2022). Fitting censored and truncated regression data using the mixture of experts models. *North American Actuarial Journal*, *26*(4), 496–520.

Fung, T. C., Badescu, A. L., & Lin, X. S. (2019a). A class of mixture of experts models for general insurance: Application to correlated claim frequencies. *ASTIN Bulletin: The Journal of the IAA*, *49*(3), 647–688.

Fung, T. C., Badescu, A. L., & Lin, X. S. (2019b). A class of mixture of experts models for general insurance: Theoretical developments. *Insurance: Mathematics and Economics*, *89*, 111–127.

Fung, T. C., & Tseung, S. C. (2022). Mixture of experts models for multilevel data: Modelling framework and approximation theory. arXiv preprint arXiv:2209.15207.

Fung, T. C., Tzougas, G., & Wüthrich, M. V. (2022). Mixture composite regression models with multi-type feature selection. *North American Actuarial Journal*, *27*(2), 396–428.

Gomes, C., Jin, Z., & Yang, H. (2021). Insurance fraud detection with unsupervised deep learning. *Journal of Risk and Insurance*, *88*(3), 591–624.

Gómez-Déniz, E. (2016). Bivariate credibility bonus-malus premiums distinguishing between two types of claims. *Insurance: Mathematics and Economics*, *70*, 117–124.

Gómez-Déniz, E., Sarabia, J. M., & Calderín-Ojeda, E. (2008). Univariate and multivariate versions of the negative binomial-inverse Gaussian distributions with applications. *Insurance: Mathematics and Economics*, *42*(1), 39–49.

Gui, W., Huang, R., & Lin, X. S. (2018). Fitting the Erlang mixture model to data via a GEM-CMM algorithm. *Journal of Computational and Applied Mathematics*, *343*, 189–205.

Henderson, C. R. (1973). Sire evaluation and genetic trends. *Journal of Animal Science*, *1973*(Symposium), 10–41.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, *31*(2), 423–447.

Jiang, W., & Tanner, M. A. (1999). On the identifiability of mixtures-of-experts. *Neural Networks*, *12*(9), 1253–1258.

Jordan, M. I., & Jacobs, R. A. (1994). Hierarchical mixtures of experts and the EM algorithm. *Neural Computation*, *6*(2), 181–214.

Kelly, M., & Nielson, N. (2006). Age as a variable in insurance pricing and risk classification. *The Geneva Papers on Risk and Insurance—Issues and Practice*, *31*(2), 212–232.

Kim, M., Jeong, H., & Dey, D. (2022). Approximation of zero-inflated Poisson credibility premium via variational Bayes approach. *Risks*, *10*(3), 54.

Klugman, S. A., Panjer, H. H., & Willmot, G. E. (2012). *Loss models: From data to decisions* (Vol. 715). John Wiley & Sons.

Kuo, K. (2020). Individual claims forecasting with Bayesian mixture density networks. arXiv preprint arXiv:2003.02453.

Lemaire, J. (1995). *Bonus-malus systems in automobile insurance* (Vol. 19). Springer Science & Business Media.
Lemaire, J., Park, S. C., & Wang, K. C. (2016). The use of annual mileage as a rating variable. *ASTIN Bulletin: The Journal of the IAA*, *46*(1), 39–69.
McCullagh, P., & Nelder, J. (1989). *Generalized linear models*.
McGilchrist, C. (1994). Estimation in generalized mixed models. *Journal of the Royal Statistical Society: Series B (Methodological)*, *56*(1), 61–69.
McGilchrist, C., & Yau, K. (1995). The derivation of BLUP, ML, REML estimation methods for generalised linear mixed models. *Communications in Statistics—Theory and Methods*, *24*(12), 2963–2980.
McLachlan, G., & Peel, D. (2004). *Finite mixture models*. John Wiley & Sons.
McLean, R. A., Sanders, W. L., & Stroup, W. W. (1991). A unified approach to mixed linear models. *The American Statistician*, *45*(1), 54–64.
Ng, S.-K., & McLachlan, G. J. (2007). Extension of mixture-of-experts networks for binary classification of hierarchical data. *Artificial Intelligence in Medicine*, *41*(1), 57–67.
Ng, S.-K., & McLachlan, G. J. (2014). Mixture models for clustering multilevel growth trajectories. *Computational Statistics & Data Analysis*, *71*, 43–51.
Ni, W., Li, B., Constantinescu, C., & Pantelous, A. A. (2014). Bonus-malus systems with hybrid claim severity distributions. In M. Beer, S-K. Au, & J. W. Hall (Eds.), *Vulnerability, uncertainty, and risk: Quantification, mitigation, and management* (pp. 1234–1244). American Society of Civil Engineers.
Norberg, R. (1979). The credibility approach to experience rating. *Scandinavian Actuarial Journal*, *1979*(4), 181–221.
Nott, D. J., Tan, S. L., Villani, M., & Kohn, R. (2012). Regression density estimation with variational methods and stochastic approximation. *Journal of Computational and Graphical Statistics*, *21*(3), 797–820.
Oh, R., Kim, J. H., & Ahn, J. Y. (2022). Designing a bonus-malus system reflecting the claim size under the dependent frequency-severity model. *Probability in the Engineering and Informational Sciences*, *36*(4), 963–987.
Oh, R., Shi, P., & Ahn, J. Y. (2020). Bonus-malus premiums under the dependent frequency-severity modeling. *Scandinavian Actuarial Journal*, *2020*(3), 172–195.
Ohlsson, E., & Johansson, B. (2010). *Non-life insurance pricing with generalized linear models* (Vol. 2). Springer.
Park, S. C., Kim, J. H., & Ahn, J. Y. (2018). Does hunger for bonuses drive the dependence between claim frequency and severity? *Insurance: Mathematics and Economics*, *83*, 32–46.
Pechon, F., Denuit, M., & Trufin, J. (2019). Multivariate modelling of multiple guarantees in motor insurance of a household. *European Actuarial Journal*, *9*, 575–602.
Pinheiro, J. C., & Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, *4*(1), 12–35.
Pinquet, J. (1998). Designing optimal bonus-malus systems from different types of claims. *ASTIN Bulletin: The Journal of the IAA*, *28*(2), 205–220.
Raudenbush, S. W., Yang, M.-L., & Yosef, M. (2000). Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate laplace approximation. *Journal of Computational and Graphical Statistics*, *9*(1), 141–157.
Sato, M.-A. (2001). Online model selection based on the variational Bayes. *Neural Computation*, *13*(7), 1649–1681.
Tseung, S. C., Badescu, A., Fung, T. C., & Lin, X. S. (2020). LRMoE: An R package for flexible actuarial loss modelling using mixture of experts regression model. Available at SSRN 3740215.
Tseung, S. C., Badescu, A. L., Fung, T. C., & Lin, X. S. (2021). LRMoE. jl: A software package for insurance loss modelling using mixture of experts regression model. *Annals of Actuarial Science*, *15*(2), 419–440.
Tzougas, G., & di Cerchiara, A. P. (2021). The multivariate mixed negative binomial regression model with an application to insurance a posteriori ratemaking. *Insurance: Mathematics and Economics*, *101*, 602–625.
Tzougas, G., Vrontos, S., & Frangos, N. (2014). Optimal bonus-malus systems using finite mixture models. *ASTIN Bulletin: The Journal of the IAA*, *44*(2), 417–444.
Tzougas, G., Vrontos, S., & Frangos, N. (2018). Bonus-malus systems with two-component mixture models arising from different parametric families. *North American Actuarial Journal*, *22*(1), 55–91.
Vickrey, W. (1968). Automobile accidents, tort law, externalities, and insurance: An economist's critique. *Law and Contemporary Problems*, *33*(3), 464–487.

Yau, K. K., Lee, A. H., & Ng, A. S. (2003). Finite mixture regression model with random effects: Application to neonatal hospital length of stay. *Computational Statistics & Data Analysis*, *41*(3–4), 359–366.

Zeger, S. L., & Karim, M. R. (1991). Generalized linear models with random effects; a Gibbs sampling approach. *Journal of the American Statistical Association*, *86*(413), 79–86.

Zhang, J., Fraser, S., Lindsay, J., Clarke, K., & Mao, Y. (1998). Age-specific patterns of factors related to fatal motor vehicle traffic crashes: Focus on young and elderly drivers. *Public Health*, *112*(5), 289–295.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Tseung, S. C., Chan, I. W., Fung, T. C., Badescu, A. L., & Lin, X. S. (2023). Improving risk classification and ratemaking using mixture-of-experts models with random effects. *Journal of Risk and Insurance Review*, 1–32. https://doi.org/10.1111/jori.12436