Insurance data can be asymmetric with heavy tails, causing inadequate adjustments of the usually applied models. To deal with this issue, a hierarchical model for collective risk with heavy tails of the claims distributions that also takes into account overdispersion of the number of claims is proposed. In particular, the distribution of the logarithm of the aggregate value of claims is assumed to follow a Student $t$ distribution. Additionally, to incorporate possible overdispersion, the number of claims is modeled as having a negative binomial distribution. Bayesian decision theory is invoked to calculate the fair premium based on the modified absolute deviation utility. An application to a health insurance data set is presented together with some diagnostic measures to identify excess variability. The variability measures are analyzed using the marginal posterior predictive distribution of the premiums according to some competitive models. Finally, a simulation study is carried out to assess the predictive capability of the model and the adequacy of the Bayesian estimation procedure.

1. INTRODUCTION

The presence of observations far from the center of the data often occurs in insurance data sets. This can be the result of the presence of outliers in the data set or a heavy-tailed distribution. In the actuary context, many authors have proposed classical approaches to deal with both issues. Under the frequentist approach, the proposal of using robust estimators to attenuate the impact of outliers was first introduced by Kunsch (1992). He examined cases of claims with extreme values and proposed the use of an M-robust estimator instead of the linear estimator of credibility.

In this context, Gisler and Reinhard (1993) proposed the use of the T-robust estimator. They used a particular influence function that reduced the influence of observations larger than a certain prespecified value.

Klugman and Hogg (1983) presented an empirical analysis of the size of loss in casualty insurance data and observed that these kinds of data have a very heavy-tailed and skewed distribution. On the other hand, there are several works describing how to model heavy-tailed distributions and detect outliers using the Bayesian approach. West (1984) considered a robust regression model to fit heavy-tailed data. The distribution assigned to the error term was structured as a mixture in the scale parameter of the normal distribution. In the actuarial context, Cipra (1996) included outliers and missing observations in a dynamic credibility model. Huisman et al. (1998) applied heavy-tailed distributions to FX return. Further details on how models for return response with outliers are fitted can be seen in Günay (2017). Bühmann and Gisler (2005, Chapter 2) presented the Bayes premium under the principles of credibility theory. These concepts are highly interpretable and understandable in actuary practice.

Based on the collective risk model and the principles of credibility theory, our challenge is to obtain robust estimates in scenarios where heavy-tailed distributions are fitted under the Bayesian paradigm. The most popular form in the literature is that the distribution assigned to the error term is structured as a mixture in the scale parameter of the normal distribution. This is obtained through a scale mixture of Gaussian and inverse-gamma distributions and is equivalent to considering all of the Gaussian errors as Student $t$ errors. For a comprehensive overview of other types of mixture models commonly used to detect and account for outliers, see Hyungsuk et al. (2019).

In the actuarial context, the log Student $t$ distribution with a known degree of freedom was first proposed by Klugman and Hogg (1983).
Our contribution here is focused on incorporating a heavy tail to the response error in the values of claims belonging to the collective risk process. Other possibilities to incorporate robustness in the estimation can be seen in Andrade and O’Hagan (2011) and Desgagné and Gagnon (2019), among others.

In this direction, our principal contribution is to propose a collective risk model in the Bayesian framework that accommodates the presence of observations far from the center of the data. We propose to model the logarithm of the claim values as Student $t$ distributed with a degree of freedom parameter assumed to be unknown. Another common aspect of the uncertainty inherent to insurance data is the overdispersion issue. To take into account possible overdispersion of the claim number distribution, we replace the Poisson distribution with the negative binomial one. We also analyze the parameter that quantifies the extra-Poisson variation (Zhou et al. 2012). To identify groups of risks, two measures are calculated: VaR (value at risk) and TVaR (tail value at risk). These measures began being used by insurance companies to assess possible portfolio losses in the late 1990s (Linsmeier and Pearson 1996; Larsen et al. 2002; Jorion 2007). VaR is a measure of possible portfolio losses; TVaR is the arithmetic mean of the VaR’s in the tail of the loss distribution; see Kaas et al. (2009, Chapter 5) for further details.

The article is organized as follows. In Section 2 we present a brief introduction to the collective risk model. Section 3 proposes our model and its particular cases, as well as the inference procedure. Section 4 presents an application to a health insurance data set, where the models described in Section 3 are also compared using different model criteria. Section 5 presents a simulation study to evaluate the efficiency of the estimation procedure. Finally, Section 6 provides the main conclusions and Section 7 presents a summary of the results with suggestions for further research.

2. HIERARCHICAL COLLECTIVE RISK MODEL: A BRIEF REVIEW

The basic collective risk model comes from important contributions of Harald Cramér and Filip Lundberg to insurance risk theory. Therefore, it is also known as the Cramér–Lundberg model; see Embrechts et al. (1997, Chapter 1) for details. Migon and Moura (2005) considered a hierarchical extension of the Cramér–Lundberg model and applied it to a health insurance plan. They predicted the pure premium based on the available past information concerning the number and total amount of claims, and also the population exposed to risk.

Let $N_{a,t}$ and $X_{a,t}$ respectively be defined as the number of claims and the total value of claims produced by a portfolio of policies in a given time $t = 1, \ldots, T$ for the age class $a = 1, \ldots, A$.

$$X_{a,t} = \begin{cases} \sum_{j=1}^{N_{a,t}} Z_{a,t,j}, & \text{if } N_{a,t} > 0, \\ 0, & \text{if otherwise}. \end{cases}$$

The $Z_{a,t,j}$ term is the value of the $j$th claim at time $t$ for the age class $a$.

The main assumptions in the Cramér–Lundberg process are:

1. The number of claims in the interval $(t-1,t]$ is a random variable denoted by $N_{a,t}$; conditional on $N_{a,t} = n_{a,t}$, the claim sizes $Z_{a,j}$, $j = 1, 2, \ldots, n_{a,t}$, are positive independent and identically distributed random variables with finite mean $\mu_{a} = E(Z_{a,j})$ and variance $\sigma_{a}^2 = var(Z_{a,j}) < \infty$.

2. The claim timing occurs at random instants $t_{1,a} \leq t_{2,a} \leq \cdots$ and the interarrival times $T_{k,a} = t_{k,a} - t_{k-1,a}$ are assumed to be independent and identically exponentially distributed random variables with finite mean $E(T_{k,a}) = \lambda_{a}^{-1}$.

Assuming that the sequences $T_k$ and $Z_k$ are independent of each other and the preceding conditions hold, it follows that $N_{a,t}$ is a homogeneous Poisson process with rate $\lambda_{a}$.

If a general distribution with finite mean $\lambda_{a}^{-1}$ is assumed for the interarrival time, a more general process is introduced, known as the renewal model, which allows for a slightly more general process $N_{a,t}$.

Supposing that $Z_{a,t,j} \sim G(\kappa_{a}, \theta_{a})$ and that the interarrival times are exponentially distributed, Migon and Moura (2005) concluded that the claims follow a Poisson process, and the claim sizes are independent and identically gamma distributed:

$$N_{a,t} \mid \lambda_{a}, \pi_{a,t} \sim P(\lambda_{a} \pi_{a,t}), \quad \lambda_{a} > 0$$

$$X_{a,t} \mid n_{a,t}, \theta_{a} \sim G(\kappa_{a}, \theta_{a}), \quad \theta_{a} > 0$$

(1)
where the symbol $Y \sim G(a, b)$ generically denotes that $Y$ is gamma distributed with a probability density function given by

$$p_Y(y | a, b) = \frac{a^b}{\Gamma(b)} y^{b-1} \exp(-by);$$

$n_{a,t}$ is the observed number of claims; $\kappa_{a,t} = n_{a,t} \kappa_a$; and $\pi_{a,t}$ denotes the insured population at time $t$ for the age class $a$.

Migon and Moura (2005) modeled the evolution of the insured population via a generalized hierarchical exponential growth model. In particular, we consider the case where the insured population at each time $t$ for age class $a$ is known. This implies that our “hierarchical collective risk model” (HCR) does not assume exchangeability by age of class like the baseline model. The hierarchical modeling comes from the nested probabilistic structure given by the conditional distribution of the total value of claims $X$ given the number of claims $N$.

To highlight this characteristic, the directed acyclic graph (DAG) is shown in Figure 1.

3. HEAVY-TAILED AND OVERDISPERSED HIERARCHICAL COLLECTIVE RISK MODEL

This section presents our model, which is suitable to accommodate high claim values. Its flexibility also enables taking into account possible overdispersion of the claim number.

Figure 1 presents the DAG related to our proposed model. Square nodes denote known constants while circular nodes describe stochastic quantities. The stochastic or functional conditional dependences are denoted by arrows. Two nested rounded rectangles define the hierarchy of the model. The inner one with random variable $N$ shows the count submodel and the outer one indicates the conditional dependence of $X$ with respect to $N$ ($X|n$).

3.1. Model Specification

It is assumed that the claim value, $X_{a,t}$, $\forall a$ and $\forall t$, is log Student $t$ ($\mathcal{L}\mathcal{T}$) distributed, $X_{a,t} \sim \mathcal{L}\mathcal{T}(\mu_{a,t}^X, \sigma_{a,t}^2, \nu_a)$; that is, the density function of $X_{a,t}$ is given by

$$p(x_{a,t} | \mu_{a,t}^X, \sigma_{a,t}^2, \nu_a) = \frac{\Gamma((\nu_a+1)/2)}{\Gamma(\nu_a/2) \sqrt{\pi \nu_a \sigma_{a,t}^2}} \left[ 1 + \frac{(\log x_{a,t} - \mu_{a,t}^X)^2}{\nu_a \sigma_{a,t}^2} \right]^{-(\nu_a+1)/2},$$

$$0 \leq x_{a,t} < \infty, \quad \nu_a > 0, \quad (2)$$

where $\mu_{a,t}^X$, $\sigma_{a,t}^2$, and $\nu_a$ denote the parameters of location, scale, and degrees of freedom of the $\mathcal{L}\mathcal{T}$ distribution, respectively. We also assume conditional independence for all time periods $t$ and all age risk classes $a$, and we set $\kappa_a = 1$, $\forall a = 1, ..., A$. To facilitate the comparison with the model stated in Equation (1), we set

$$\mu_{a,t}^X = \log \left( \frac{n_{a,t}}{\theta_a} \right) - \frac{1}{2} \sigma_{a,t}^2, \quad \text{and} \quad \sigma_{a,t}^2 = \log \left( \frac{1}{n_{a,t}} + 1 \right). \quad (3)$$
The inference procedure is carried out by sampling from the posterior distribution. The posterior distribution is proportional to the likelihood \( L(M_{a,t}) \) presented in Section 2 with the negative binomial distribution \( (\text{NB}) \), that is, \( N_{a,t} | \mu_{a,t}, \delta_a \sim \text{NB} \left( \left(1 + \frac{\mu_{a,t}}{\delta_a}\right)^{-1}, \delta_a \right) \) with mean \( \mathcal{M} = \mu_{a,t} \delta_a / \delta_a \) and variance \( \mathcal{M}(1 + \mathcal{M}) \) where \( \mu_{a,t} = \lambda_a \pi_{a,t} \). This model is parameterized to enable comparisons with the model presented in Section 3. Note that \( N_{a,t} | \mu_{a,t}, \delta_a \) can be obtained as a mixture of \( N_{a,t} | \mu_{a,t}, \gamma_a \sim \mathcal{P}(\mu_{a,t}, \gamma_a) \) and \( \gamma_a \sim \mathcal{G}(\delta_a, \delta_a) \).

To complete the model specification, we assign a proper prior distribution to the model parameters. Independent priors of the model parameters across age class are employed. The choice of the prior distributions was driven by the need to make inferences with minimum prior subject information. For the degrees of freedom parameter, \( \nu_a \), a gamma distribution, \( \mathcal{G}(a_0, b_0) \), is used (Fernandez and Steel 1999) to specify priors with exponential tails. We assume \( \delta_a \sim \mathcal{G}(a_0, b_0) \) for the dispersion parameter. The conditional distributions of the \( \theta_a \) and \( \lambda_a \) parameters are assigned as \( \theta_a \sim \mathcal{G}(a_0, b_0) \) and \( \lambda_a \sim \mathcal{G}(a_2, b_2) \), respectively. To allow subjectivity in the choice of the hyperparameter, a third hierarchical level assigns to all of them \( \mathcal{G}(0.1, 0.1) \) and Half-Cauchy \( \mathcal{H}(0.1) \) independent prior distributions (Polson and Scott 2012; Klein et al. 2014; Klein and Kneib 2016).

A sensitivity analysis verified that the assumption of a gamma prior for the hyperparameters does not show anomalous behavior (see Appendix) when used to calculate the 95th percentile of the marginal posterior distribution of the premium.

Table 1 presents a summary of the six models analyzed in this work.

### 3.2. Inference procedure

Let \( \Theta = (\theta, \kappa, \nu, \lambda, \delta) \) be the vector of parameters for all risk classes \( a, a = 1, \ldots, A \), and denote by \( D \) all information available for all time periods \( t, t = 1, \ldots, T \). Therefore, the likelihood function of the \( \mathcal{L} \mathcal{T} - \text{NB} \) model is given by

\[
\log L(\Theta | D) = \sum_{a,t} \log \left( \frac{\nu_a^2 w(\nu_a)}{2^{\nu_a} \Gamma(\nu_a)} \right) - \frac{\nu_a + 1}{2} \log \left( \nu_a + P^{a,t} \right) + \psi_a \tag{4}
\]

with \( w(\nu_a) = \frac{\Gamma(\nu_a + 1)}{\Gamma(\nu_a)} \), \( P^{a,t} = \sigma_{a,t}^{-1} \log \left( \frac{\nu_a^2 \nu_{a,t} (1 + \nu_{a,t})^2}{\nu_a^2 \nu_{a,t}} \right) \), and \( \psi_a = \Gamma(\nu_{a,t} + \delta_a) / \Gamma(\nu_{a,t}) \). The inference procedure is carried out by sampling from the posterior distribution. The posterior distribution \( \pi(\Theta | D) \) is proportional to the likelihood \( L(\Theta | D) \) times the prior information \( \pi(\Theta) \). Sampling from \( \pi(\Theta | D) \) can be done by the Markov-chain Monte Carlo (MCMC) technique.
3.3. The Premium Value

Migon and Moura (2005) obtained the premium value as a solution to a decision problem. Using the modified absolute deviation utility with some appropriate constants, it can be proved under some regularity conditions that the value \( d \) that maximizes the utility function is the 95th percentile of the predictive posterior of the unknown total claim value per insured.

Let the total claim value in the time horizon and the insured population for each age class \( a \), \( a = 1, \ldots, A \), at time \( T + H \), respectively, be \( X_{a,T+H} \) and \( \Pi_{a,T+H} \), where \( H \) is the planning horizon. Based on the decision theory fundamentals, if the modified absolute deviation is used as a utility function, then the decision value (premium value) that maximizes the utility function in the marginal predictive posterior distribution \( (R_{a,T+H} = \frac{X_{a,T+H}}{\Pi_{a,T+H}}) \) is given by its 95th percentile. Therefore, the premium value is obtained for each model as the 95th percentile of \( R_{a,T+H} \), which is calculated by

\[
p(R_{a,T+H} \mid D_T) = \int_\Theta p(R_{a,T+H} \mid \theta)p(\theta \mid D_T)d\Theta,
\]

where \( p(\theta \mid D_T) \) denotes the posterior distribution of all model parameters.

3.4. Measure of Risk in the Marginal Predictive Posterior Distribution

The marginal predictive distribution of \( R \) (see Section 3) is used to obtain the value at risk (VaR) and the tail value at risk (TVaR).

Let \( R \) be a random variable that represents the risk. The value at risk of \( R \) at a level \( \tau \in (0, 1) \), \( \text{VaR}_\tau(R) \), is defined as \( \text{VaR}_\tau(R) = F_R^{-1}(\tau) = \inf \{ r : F_R(r) \geq \tau \} \). The VaR measure is the inverse cumulative distribution function (cdf) of \( R \) computed at \( \tau \). The tail value at risk of \( R \) at a level \( \tau \) is defined as \( TVaR_\tau(R) = \frac{1}{1-\tau} \int_\tau^1 \text{VaR}_\tau(R) dt \). Therefore, \( TVaR \) can be interpreted as the arithmetic mean of the VaR's of \( R \) from \( \tau \) onward. An alternative expression is given by \( TVaR_\tau(R) = \text{VaR}_\tau(R) + \frac{1}{1-\tau} \text{ES}_\tau(R) \), where \( \text{ES}_\tau(R) \) describes the expected shortfall (ES) at level \( \tau \) of \( R \). The ES is defined as \( \text{ES}_\tau(R) = E[(R - \text{VaR}_\tau(R))_+] \), and ES can be interpreted as the net stop-loss premium in the hypothetical situation of excess over \( \text{VaR} \).

The second form of \( TVaR \) is used in Section 4.
3.5. Bayesian Model Selection

To assess model fitness, the deviance information criterion (DIC) (Spiegelhalter et al. 2002) is used. The DIC is defined by

$$\text{DIC} = D(\overline{\Theta}) + 2p_D,$$

where $D(\Theta) = -2\log(L(\Theta \mid y))$, where $y$ denotes all information provided by the data, $\Theta$ is the vector of parameters, and $L(\Theta \mid y)$ is the likelihood function, $p_D = D(\overline{\Theta}) - D(\overline{\Theta})$, where $D(\overline{\Theta})$ and $\overline{\Theta}$ denote the posterior mean of $D(\Theta)$ and $\Theta$, respectively. To verify the accuracy of prediction values, the continuous ranked probability score (CRPS) criterion (Lopes et al. 2012) was also applied. The CRPS can be expressed as

$$\text{CRPS} = \sum_{a,t} \left( E[y_{\text{rep,at}} - y_{at}] - 0.5E[y_{\text{rep,at}} - \tilde{y}_{\text{rep,at}}] \right)/AT,$$

where the expectations in Equation (5) are calculated under the posterior marginal distribution, and $y_{\text{rep,at}}$ and $\tilde{y}_{\text{rep,at}}$ are independent replicates from it. Assuming there is a sample of size $L$ of the marginal posterior distribution, these expectations can be approximated by $\sum_{i=1}^{L} y_{\text{rep,at}}^{(i)} - y_{at}/L$ and $\sum_{i=1}^{L} y_{\text{rep,at}}^{(i)} - \tilde{y}_{\text{rep,at}}^{(i)}/L$, respectively. Models with lower DIC and CRPS values are preferred.

4. ILLUSTRATION WITH A REAL DATA SET

The data set used to illustrate our alternative models was extracted from a health care plan implemented in a university in northeast Brazil. This data set was also used by Migon and Moura (2005). Its main characteristics are being self-administered and having a small number of insureds. The data set consists of only 12 monthly observations of the number of claims, the total value of the observed claims, and the number of insured persons per type of service (doctor’s office visit, diagnostic test, and hospital care), by age class. There are seven age classes, $a = 1, \ldots, 7$, and they correspond to the following intervals: [0,17), [17,30), [30,40), [40,50), [50,60), [60,70), [70 +). For the sake of simplicity, the notation employed in the next sections does not include an additional index to describe the type of service. All calculations were performed separately for each type of service, so the respective predicted values of the size and the number of claims per age class were obtained by aggregation.

The six models described in Table 1 of Section 3 were fitted to the data. Sampling from $\pi^D(\Theta \mid D)$ is done by the Markov-chain Monte Carlo (MCMC) method, using the No-U-Turn-Sampler implemented in the Stan software (Stan Development Team 2018). Three chains of length 10,000, starting from different locations with burn-in of 5,000 iterations, were generated. Convergence was monitored via Markov-chain Monte Carlo autocorrelation, cross-correlation, and density plots using the approach developed by Cowles and Carlin (1996) and implemented in R (R Development Core Team 2018).

The marginal predictive posterior distribution $R_{a,T+H}$ for the considered models is presented in Figure 1(a), where (+) corresponds to the 95th percentile of the predictive distribution, accumulated over the planning horizon.
These results indicate that the premiums obtained from the $\mathcal{L}T$ models are considerably smaller for age class $[40,50)$ than the ones obtained from the other models. Thus, the $\mathcal{L}T$ models provide greater shrinkage of the premium. For the age classes $[60,70)$ and $[70,\infty)$, the marginal posteriors of $R_{a,T+H}$ are approximately the same for the $\mathcal{L}N-NB$ and $\mathcal{L}T-NB$ models, but different for the $G-NB$ model.

We additionally considered an analysis by age class of the posterior distributions of $d$ and $C_{23}$, shown in Figure 2. An analysis of the posterior distribution of $d$ shows there is a high probability of the true values of $d$ being greater than 1 for all three considered models and all age classes. The small values of the $m$ estimates indicate heavy tails, most evident for the $[40,50)$ age class. These results show evidence of overdispersion and the need to introduce a heavy-tailed distribution to model the claim values.

Figure 3 shows box plots of the marginal posterior of the (a) $k$ and (b) $h$ effects based on the six models by age class. As expected, an analysis by age class of the posterior distributions of $k$ for all six models shows their posterior means increase as the insured gets older. Nevertheless, the three models that take into account overdispersion have more uncertainty in estimating $k$ than their respective counterparts. The parameter $\theta_a$ increases as the respective mean claim value decreases (see Eq. [3]). Since the mean claim value tends to increase with age, the pattern shown in Figure 3(b) is expected. It is worth noting that the $\mathcal{L}T-NB$ and $\mathcal{L}T-P$ models perform more in line with this pattern than the other models.

Some measurements to study the impact of overdispersion on the age classes were also developed. These measurements are based on the posterior predictive variance of both size and number of claims. To quantify the over- or underdispersion, the variance of the posterior marginal predictive distribution of $N_{a,t}$ is calculated. The term $1 + \frac{R_{a,T+H}}{N_{a,t}}$ measures the excess of variability in each age class $a$.

Figure 4 shows the coefficient of variation by age class (a) and for the age class and the three types of service (circle, triangle, square) (b) for model M6. The hospital care service (square) presents the largest premium values. The age classes $[40,50)$ and $[70,\infty)$ show the largest coefficients of variation when all the services are averaged.

To identify possible high-risk age groups assuming the presence of overdispersion, that is, assuming a negative binomial distribution for the number of claims, the measures $VaR$ and $TVaR$ were calculated for the models whose number of claims were assumed to be negative binomial distributed, that is, models M4, M5, and M6.

Moreover, the $VaR$ estimates from the $G$ distribution are very close to those obtained from the Student $\mathcal{L}T$ distribution, except for the age class $[40,50)$. Figure 5(b) describes the tail of the value at risk in $\tau \in (0,0.1)$. According to Figure 6, $TVaR$ statistics present approximately the same values for age classes $[0,17],[17,30],[30,50],[50,60)$. Higher values of $TVaR$ are observed for model M4 for all age classes.
FIGURE 5. Marginal Posterior Distributions of the Variation Coefficients Based on M6: \( \mathcal{L} - \mathcal{N} \beta \) by (a) Age Class and Service, Each Curve Corresponds to \( \mathcal{T} = 12 \) Values and (b) Age Class, Each Curve Corresponds to \( \mathcal{T} \times \mathcal{S} = 36 \) Values. Note: In (a) the solid circle symbol and gray curve indicate information with respect to service 1. The solid triangle symbol and light gray curve indicate information with respect to service 2. The solid square symbol and white curve indicate information about service 3. The distance between solid symbols (circle, triangle, square) denote the 2.5%, 97.5% empirical quantiles with respect to services 1, 2 and 3, respectively, for each age class. In (b), the symbol (+) denotes the 2.5%, 97.5% empirical quantiles and \( \times \) the posterior mean by age class without services.

FIGURE 6. Values-at-risk Calculated for the Seven Age Classes and for Models M4, M5 and M6: (a) VaR Calculated Over a Complete Range of Confidence Levels \( \mathcal{t} \in (0, 1) \). (b) VaR Calculated Over a Range of Confidence Levels \( \mathcal{t} \in (0, 0.1) \).
Table 2 presents the values of model comparison measures. According to these criteria, the $L_T/C_0 NB$ distributions seem to be the most appropriate to model the number of claims and their total value.

5. A SIMULATION STUDY

In this section we present a simulation study to evaluate the efficiency of the estimation procedure in recovering the true values of the model hyperparameters ($\alpha_1, \beta_1, \alpha_2, \beta_2, \alpha_0, \beta_0, \alpha_r, \beta_r$). Three hundred data sets based on $M_6$ were generated and the population size was kept fixed and taken from the health insurance data set described in Section 4.

To make the simulated data as similar as possible to the real data, the true hyperparameters in Table 3 were calculated by equating the corresponding means of the gamma distribution to the posterior means obtained by fitting the real data and setting their respective precision values approximately equal to 11. The system of two equations was solved and the respective true values were obtained.

The prior distributions used to fit the $S = 300$ simulated data sets were $G(0.1, 0.1)$ for the eight hyperparameters.
To obtain samples of the posterior distribution, the MCMC algorithm was implemented using the Stan software (Stan Development Team 2018). To monitor the convergence of chains, the procedures described in Section 3.2 were implemented.

The mean relative bias (MRB), as a percentage, is defined as

\[
MRB = \frac{100}{S} \sum_{s=1}^{S} \frac{\hat{h}_s - h}{C_0}\frac{1}{C_1},
\]

where \(\hat{h}_s\) is the posterior mean of the respective true hyperparameter \(h\) for the \(s\) simulated data sets, \(s = 1, \ldots, S\).

Another criterion to evaluate the performance of the MCMC results is given by the mean squared error, expressed in Table 3 as a percentage, that is,

\[
MSE = 100 \sum_{s=1}^{S} \frac{(\hat{h}_s - h)^2}{C_0C_1}.
\]

Table 3 shows small MRB and MSE values.

Since the premium value \(P\) is a conservative measurement (see Migon and Moura, 2005), we developed a statistic \(U_a\) that compares the simulated value of \(R_a\), with the 95% quantile obtained from the marginal posterior predictive distribution of \(R_a\), that is, the premium value \(P_a\); see Section 3.3 for details. The quantity \(U_a\) indicates the level of protection that the proposed model gives to the insurance company for each age class \(a = 1, \ldots, A\).

Table 4 shows the values of \(U_a = 100 \sum_{s=1}^{S} \frac{\#Q_s(R_a, 0.95) > P_a^{(s)}}{S}\), and \(Q_s(R_a, 0.95)\) denotes the 95% quantile of \(R_a\) and \(P_a^{(s)}\) defines the true premium value, calculated for each simulated sample \(s = 1, \ldots, S\) and age class \(a = 1, \ldots, A\). Although the age classes \(30 \leq a \leq 40\) and \(\geq 70\) yielded slightly higher percentages than the others, on average all values of \(U_a\) were approximately 2%, indicating that only in approximately 2% of the samples were the premium values smaller than the total claim value per insured.

### 6. MAIN CONCLUSIONS

Collective risk models that incorporate heavy-tailed distributions and overdispersion were proposed to provide realistic estimates and forecasts, allowing practitioners to set fair premium prices.

Insurance premium values not greatly affected by observations far from the center and allowing for the number of claims to have a larger variance than that prescribed by the Poisson distribution were calculated. The inference procedure was applied in the Bayesian approach and estimation was carried out through MCMC methods via the Stan software (Stan Development Team 2018). Table 2 shows comparison criterion values in favor of our proposal.

### 7. CONCLUDING REMARKS

An application was presented and six hierarchically structured models (Gelman and Hill 2007) were fitted. The results showed shrinkage of the premium for some age classes. Excess of variability and variation coefficients were used to identify

---

**TABLE 3**

| Hyperparameters | \(\alpha_k\) | \(\beta_k\) | \(\alpha_l\) | \(\beta_l\) | \(\alpha_0\) | \(\beta_0\) | \(\alpha_r\) | \(\beta_r\) |
|-----------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| True            | 2.85        | 5.62        | 12.37       | 3.71        | 1.11        | 11.11       | 10.12       | 3.35        |
| MRB (in%)       | 1.75        | 0.41        | 0.58        | 0.02        | -0.30       | 0.13        | 0.23        | -1.69       |
| MSE (in%)       | 1.63        | 0.36        | 0.65        | 1.38        | 2.73        | 0.06        | 0.17        | 1.05        |

*Note:* Mean relative bias (MRB) and mean squared error (MSE).

**TABLE 4**

| Statistic | \(0 \overset{\sim}{=}\) 17 | \(17 \overset{\sim}{=}\) 30 | \(30 \overset{\sim}{=}\) 40 | \(40 \overset{\sim}{=}\) 50 | \(50 \overset{\sim}{=}\) 60 | \(60 \overset{\sim}{=}\) 70 | \(\geq 70\) |
|-----------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-------------|
| \(U_a\)   | 0.68            | 1.03            | 3.77            | 1.71            | 2.05            | 1.37            | 3.08        |

*Note:* \(U_a\) indicates the level of protection.
A HEAVY-TAILED AND OVERDISPERSED COLLECTIVE RISK MODEL

Pamela M. Chiroque-Solano

ORCID

ACKNOWLEDGMENTS

For future work, a model that relates the costs and the age of the beneficiaries could be considered. The Birnbaum Saunders $t$ and inverse Gaussian distributions should also be evaluated and used to model the claim values (Gilberto et al. 2012). Reference priors for the overdispersion parameter based on Liseo et al. (2010) could also be obtained and applied. Covariates such as gender, education level, and national economic indices could be included in the model to help to obtain the fairest premium values.

REFERENCES

A HEAVY-TAILED AND OVERDISPERSED COLLECTIVE RISK MODEL

age classes with non-standard premiums. The analysis of the insurance data revealed that service 3 (hospitalization) had more observations far from the center than the other services. The estimates of the parameters $\delta$ and $\nu$ presented in Figure 2 suggest overdispersion and heavy tails.

The $VaR$ and $TVaR$ measures were considerably different for the three models analyzed. This was due to the difference in the fat tails of models M4, M5, and M6. This result is consistent with the findings of Stoyanov et al. (2013) and Günay (2017). Furthermore, in the alternative formula for $TVaR$, the expectation shortfall term was taken into account for the premium calculation. This provides greater flexibility in the interpretation. The method shows evidence that the age classes [40,50),[60,70), and [70,+) have higher risk than the others (see Figures 1, 5, and 6).

To validate the algorithm, artificial data were generated from the proposed model (M6). The convergence of the chains of the parameters was verified by Geweke and Raftery’s method. The model correctly predicted the quantities of interest. The inference procedure showed the desired behavior, since the premium calculated via our proposed model did not cause a financial loss to the insurance company compared to the results of the standard model (see Figure 1). Furthermore, when we compared the most important quantity to be considered, the premium value, it was suitable, as indicated by the $U_{a}$ statistic.

The approach proposed in this article is advised to be used in scenarios where the claim counts heterogeneous and the claim amounts are often far from the center, as the health insurance data analyzed in Section 4. Many insurance loss data also have these features. However, careful model assessment, as presented in Section 4, should be done before considering the use of the approach to any particular insurance data.

For future work, a model that relates the costs and the age of the beneficiaries could be considered. The Birnbaum Saunders $t$ and inverse Gaussian distributions should also be evaluated and used to model the claim values (Gilberto et al. 2012). Reference priors for the overdispersion parameter based on Liseo et al. (2010) could also be obtained and applied. Covariates such as gender, education level, and national economic indices could be included in the model to help to obtain the fairest premium values.

acknowledgments

This work was part of the master’s dissertation of P. M. C. Solano under the supervision of F. A. S. Moura. P. M. C. Solano benefited from a scholarship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ). The authors thank the editor, associate editor, and two referees for very thoughtful and constructive comments.

ORCID

Pamela M. Chiroque-Solano http://orcid.org/0000-0002-2667-8755

Fernando Antônio da S. Moura http://orcid.org/0000-0002-3880-4675

REFERENCES

Andrade, J., and A. O’Hagan. 2011. Bayesian robustness modelling of location and scale parameters. Scandinavian Journal of Statistics 38(4): 691–711. http://www.jstor.org/stable/41411145

Bühlmann, H., and A. Gisler. 2005. A course in credibility theory and its applications. Berlin, Germany: Springer. https://www.springer.com/gp/book/9783540257530

Cipra, T. 1996. Dynamic credibility with outliers and missing observations. Applications of Mathematics 2: 149–59. https://am.math.cas.cz/am41-2-4.html doi:10.21136/AM.1996.134318

Cowles, M., and B. Carlin. 1996. Markov chain Monte Carlo convergence diagnostics: A comparative review. Journal of the American Statistical Association 91: 883–904. https://www.jstor.org/stable/2291683 doi:10.1080/01621459.1996.10476956

Desgagné, A., and P. Gagnon. 2019. Bayesian robustness to outliers in linear regression and ratio estimation. Brazilian Journal of Probability and Statistics 33: 205–21. doi:10.1214/17-BJPS385

Embrechts, P., C. Kluppelberg, and T. Mikosch. 1997. Loss models. Modelling extreme events for insurance and finance. https://www.springer.com/gp/book/9783540609315

Fernandez, C., and M. Steel. 1999. Multivariate Student-$t$ regression models: Pitfalls and inference. Biometrika 86(1): 153–67. http://www.jstor.org/stable/2673544. doi:10.1093/biomet/86.1.153

Gelman, A., and J. Hill. 2007. Data analysis using regression and multilevel/hierarchical models. New York, NY: Cambridge University Press. https://www.cambridge.org/core/books/data-analysis-using-regression-and-multilevelhierarchical-models/52A29531C7FD730C3A68951A17C9D983

Gilberto, P., V. Leiva, M. Barros, and S. Liu. 2012. Robust statistical modeling using the Birnbaum Saunders-$t$ distribution applied to insured. Applied Stochastic Models in Business and Industry 28: 16–34. https://onlinelibrary.wiley.com/doi/abs/10.1002/asmb.887 doi:10.1002/asmb.887

Gisler, A., and P. Reinhard. 1993. Robust credibility. Astin Bulletin: The Journal of the International Actuarial Association 23: 117–43. http://www.actuaries.org/LIBRARY/ASTIN/vol23no1/117.pdf
### APPENDIX

#### TABLE A.1

**Premium Value for Each Model by Age Class Using Gamma (G) and Half-Cauchy (HC) Distributions for the Hyperparameters**

| Age Class | M1 | M2 | M3 | M4 | M5 | M6 |
|-----------|----|----|----|----|----|----|
| [0,17)   | \(G\) | \(HC\) | \(G\) | \(HC\) | \(G\) | \(HC\) |
| \(l_i\)  | 14.520 | 14.530 | 21.954 | 22.014 | 25.570 | 25.576 |
| \(l_u\)  | 15.781 | 15.769 | 24.160 | 24.148 | 28.018 | 28.011 |
| \(P\)    | 15.672 | 15.661 | 23.993 | 23.954 | 27.827 | 27.814 |
| [17,30)  | \(G\) | \(HC\) | \(G\) | \(HC\) | \(G\) | \(HC\) |
| \(l_i\)  | 15.642 | 15.582 | 23.373 | 23.347 | 27.185 | 27.127 |
| \(l_u\)  | 15.526 | 15.496 | 23.201 | 23.167 | 27.001 | 26.922 |
| \(P\)    | 15.692 | 15.646 | 23.278 | 23.245 | 27.241 | 27.214 |
| [30,40)  | \(G\) | \(HC\) | \(G\) | \(HC\) | \(G\) | \(HC\) |
| \(l_i\)  | 14.423 | 14.426 | 21.086 | 21.111 | 24.783 | 24.807 |
| \(l_u\)  | 15.692 | 15.646 | 23.278 | 23.245 | 27.241 | 27.214 |
| \(P\)    | 15.794 | 15.724 | 24.169 | 24.130 | 28.026 | 27.974 |
| [40,50)  | \(G\) | \(HC\) | \(G\) | \(HC\) | \(G\) | \(HC\) |
| \(l_i\)  | 14.394 | 14.371 | 21.287 | 21.298 | 24.797 | 24.840 |
| \(l_u\)  | 15.592 | 15.591 | 23.315 | 23.325 | 27.192 | 27.153 |
| \(P\)    | 15.494 | 15.472 | 23.169 | 23.154 | 26.999 | 26.932 |
| [50,60)  | \(G\) | \(HC\) | \(G\) | \(HC\) | \(G\) | \(HC\) |
| \(l_i\)  | 14.430 | 14.420 | 21.132 | 21.119 | 24.806 | 24.813 |
| \(l_u\)  | 15.674 | 15.612 | 23.259 | 23.268 | 27.227 | 27.215 |
| \(P\)    | 15.563 | 15.524 | 23.085 | 23.066 | 26.998 | 27.010 |
| [60,70)  | \(G\) | \(HC\) | \(G\) | \(HC\) | \(G\) | \(HC\) |
| \(l_i\)  | 14.520 | 14.530 | 21.954 | 22.014 | 25.570 | 25.576 |
| \(l_u\)  | 15.781 | 15.769 | 24.160 | 24.148 | 28.018 | 28.011 |
| \(P\)    | 15.672 | 15.661 | 23.993 | 23.954 | 27.827 | 27.814 |
| [70,+]   | \(G\) | \(HC\) | \(G\) | \(HC\) | \(G\) | \(HC\) |
| \(l_i\)  | 14.520 | 14.530 | 21.954 | 22.014 | 25.570 | 25.576 |
| \(l_u\)  | 15.781 | 15.769 | 24.160 | 24.148 | 28.018 | 28.011 |
| \(P\)    | 15.672 | 15.661 | 23.993 | 23.954 | 27.827 | 27.814 |

**Note:** Age classes: [0,17),[17,30),[30,40),[40,50),[50,60),[60,70),[70,+.). Models M1 (\(G\) + \(\mathcal{P}\)), M2 (\(LN\) + \(\mathcal{P}\)), M3 (\(LT\) + \(\mathcal{P}\)), M4 (\(G\) \(\mathcal{N}\)), M5 (\(CN\) \(\mathcal{N}\)), M6 (\(LT\) \(\mathcal{N}\)). \(l_i\) indicates the lower limit’s 2.5th percentile, \(l_u\) indicates the upper limit’s 97.5th percentile, and \(P\) is the estimated premium (95th percentile).