Dynamic Bayesian Ratemaking: A Markov Chain Approximation Approach

Hong Li[1], Yang Lu[2], and Wenjun Zhu[3]

January 21, 2020

Abstract

We contribute to the non-life experience ratemaking literature by introducing a computationally efficient approximation algorithm for the Bayesian premium in models with dynamic random effect, where the risk of a policyholder is governed by an individual process of unobserved heterogeneity. Although intuitive and flexible, the biggest challenge of dynamic random effect models is that the resulting Bayesian premium typically lacks tractability. In this paper we propose to approximate the dynamics of the random effect process by a discrete (hidden) Markov chain, and replace the intractable Bayesian premium of the original model by that of the approximate Markov chain model, for which concise, closed form formula are derived. The methodology is general as it does not rely on any parametric distributional assumptions, and in particular allows for the inclusion of both the cost and the frequency components in pricing. Numerical examples show that the proposed approximation method is highly accurate. Finally, a real-data pricing example is used to illustrate the versatility of the approach.

[1]Warren Centre for Actuarial Studies and Research, Asper School of Business, University of Manitoba, Canada. Email: li.hong@umanitoba.ca.

[2]Corresponding author. University of Paris 13, CEPN and THEMA. Email: yang.lu@univ-paris13.fr

[3]Division of Banking & Finance, Nanyang Business School, Nanyang Technological University, Singapore. Email: wjzhu@ntu.edu.sg

Keywords: Markov chain discretization; Bayesian premium; dynamic random effect; risk heterogeneity.

MSC code: 62M10

JEL code: C0

Acknowledgement: We thank the the two referees for helpful comments. The authors gratefully acknowledge the Society of Actuaries CKER for an eponymous individual grant. In addition, Lu thanks the CNRS, the Labex MME-DII, University of Paris 13 faculty research grant BQR, as well as the ACPR chair Systemic Risk and Regulation; Part of the work is conducted while Lu is on sabbatical visit at THEMA, University of Cergy-Pontoise, whose hospitality is greatly acknowledged. Zhu also thanks the research funding support from the Nanyang Technological University Start Up Grant (M4082106.010).

1 Introduction

In experience ratemaking, it is standard to decompose the annual insurance premium into the product of a base (i.e., *a priori*) premium depending on individual characteristics (e.g. age, gender, and area of residency of policyholders), and a bonus-malus (i.e., *a posteriori*) adjustment coefficient. The latter also depends on the claim history $\underline{Z_T} = (Z_1, ..., Z_T)$ of this individual, where Z_t is the observed claim information at year t, such as the annual claim count (or the claim amounts, or both). In the standard Bonus-Malus System (BMS), the bonus-malus coefficient depends on the annual number of claims only, but not on individual characteristics [see e.g. Lemaire and Zi (1994)]. In addition, it is assumed to evolve as a Markov chain, whose transition matrix is typically imposed by the regulator. This system has the disadvantage of over-penalizing policyholders with a high base premium [see Taylor (1997); Tan et al. (2015)]. A popular alternative is the random effect model, in which an unobserved individual heterogeneity captures the serial dependence between the Z_t's. The random effect is assumed to be static in the early literature. For instance, Dionne and Vanasse (1989) propose a model with gamma distributed random effect, and derive closed form Bayesian premium using the Poisson-gamma

conjugacy property[1]. However, an important deficiency of the static assumption is that past claims have the same weight in the determination of the Bayesian premium, regardless of their seniority (i.e. age). Since then, this model has been extended to allow for a dynamic random effect, which effectively assigns more weights to recent claims [see e.g. Englund et al. (2009); Pinquet et al. (2001)].

One unsolved limitation of the dynamic random effect models is that the Bayesian forecast (i.e. the Bayesian premium) $\mathbb{E}[Z_{T+1}|\underline{Z_T}]$ is usually computationally intractable for mildly large T, except in some special cases. For instance, recently Lu (2018) shows that if the dynamic random effect X_t follows the autoregressive gamma (ARG) process, and the annual accident count is conditionally Poisson given X_t, then the Bayesian premium is available through the computation of the joint Laplace transform of the process, as well as its various partial derivatives. This result can be seen as the dynamic extension of the Poisson-gamma model of Dionne and Vanasse (1989).

However, this ARG-Poisson approach has several significant limitations. Firstly, when the policyholder's claim history is long, or when the numbers of claims are large, the approach loses its efficiency since the multivariate Laplace transform quickly becomes cumbersome. Secondly, it is not clear how different parametric assumptions, such as the ARG dynamics of the random effect or the conditional Poisson assumption, can be relaxed. Thirdly, the model only accounts for the annual number of claims for pricing. However, since the ultimate variable of interest is the total cost of claims, it is clearly advantageous to include claim costs in the pricing formula, so that policyholders with a large number of minor claims is not over-penalized compared to those with fewer claims but larger total amounts. While dependence between claim count and cost has been extensively studied [see e.g. Gschlößl and Czado (2007); Frees et al. (2011); Czado et al. (2012); Shi et al. (2015); Garrido et al. (2016)], most of these papers are focused on one-period models. Only a few works [see Frangos and Vrontos (2001); Gómez-Déniz (2016)] consider experience ratemaking using both counts and costs but their frameworks are limited to the static random effect.

The computational burden of the Bayesian premium has also favored the simpler linear cred-

[1]See also Boucher and Guillen (2011) for a semi-parametric extension of this approach.

ibility method, which approximates the Bayesian premium by a linear combination of $Z_1, ..., Z_T$ via linear regression. Despite its simplicity, it can lead to substantial pricing errors [Lu (2018)]. Moreover, the credibility formula does not necessarily ensure all the weight coefficients in the linear regression to be positive, leaving room for potentially negative premium, which should be avoided at all costs [see e.g. Pinquet (2018) for a discussion]. Finally, similar to the ARG-Poisson model, the linear credibility approach does not allow the user to account for *both* the cost and the frequency of claims.

In this paper, we propose a Markov chain approximation approach to reduce the computational complexity in dynamic random effect models. The underlying idea is that a continuously valued Markov process can be discretized into a finite state Markov chain. This leads to an approximate hidden Markov model, for which filtering techniques are readily available [see Kitagawa (1981); Hamilton (1989)]. Thus the intractable Bayesian premium in the original model can be replaced by the simple Bayesian premium of the approximate model. Such a discretization method, first proposed in the statistical literature, has become quite popular in finance following the binomial option pricing model proposed by Cox et al. (1979)[2], but is yet to be introduced in the insurance literature. Numerical examples show that the resulting approximate Bayesian premia are highly accurate across different ratemaking model specifications. More importantly, our approach is widely applicable beyond the setting of Lu (2018), in the sense that it relies neither on the conditional Poisson assumption, nor on any particular dynamics of the random effect. In particular, it is still applicable when Z_t contains information other than the annual claim count.

The paper is organized as follows. Section 2 describes the general methodology. Section 3 provides detailed approximation formulas under two popular dynamic specifications of the random effect: the ARG process and the log-Gaussian AR process. Section 4 uses simulation studies to compare the pricing accuracy between the approximate Bayesian premium and the credibility premium with the benchmark Bayesian premium. Section 5 presents the real-data analysis using the proposed approximation approach. Section 6 concludes. Technical details are gathered in Appendices.

[2]See also Duan and Simonato (2001) for American option pricing.

2 The approximation approach

In this section we describe the general discretization methodology for approximating the Bayesian premium under the dynamic random effect framework.

2.1 General setting

Throughout the paper we will make the following two assumptions.

Assumption 1. *There exists an unobservable policyholder-specific, univariate random effect process* (X_t) *which measures the time-varying risk of each policyholder*[3]*. Moreover,* (X_t) *is exogenous, that is, the joint conditional distribution of* (X_t, Z_t) *can be decomposed as:*

$$\ell(x_t, z_t|\underline{z_{t-1}}, \underline{x_{t-1}}) = \ell(x_t|\underline{x_{t-1}})\ell(z_t|x_t),$$

where the symbol ℓ *is the abbreviation of "law".*

While Assumption 1 assumes the random effect (X_t) to be univariate[4], it does not exclude cases where (Z_t) is multivariate. For instance, in the case where for each individual, both the annual number of claims and total claim amount are observed, (Z_t) is bivariate. Moreover, if individual claim amounts are available, then (Z_t) has a time-varying dimension depending on the annual number of claims. Note also that some ratemaking models in the literature do not satisfy Assumption 1 in the sense that (X_t) is endogenous, that is, the dynamics of (X_t) depends on lagged values of (Z_t) [see e.g. Abdallah et al. (2016)]. One shortcoming of these models is that the claim process is not automatically stationary. For this reason, they are not explored in this paper.

[3]For expository purpose, the individual index i is omitted in the paper when no ambiguity arises, but reader should keep in mind that the realized trajectory of the process (X_t) differs among policyholders, although its dynamics is assumed i.i.d. across the policyholders.

[4]The methodology of discretizing a Markov process into a discrete time Markov chain still works even if the Markov process is multivariate [see e.g. Galindev and Lkhagvasuren (2010) for the discretization of a bivariate VAR(1) model]. However, as a first attempt in the insurance literature, in this paper we focus on the univariate case.

Assumption 2. *The dynamics of process* (X_t) *is Markov with continuous conditional density:*

$$\ell(x_t|\underline{x_{t-1}}) = \ell(x_t|x_{t-1}).$$

Moreover, the process is strictly stationary.

The above two assumptions are rather standard and have been adopted by, among others, Pinquet (1997), Pinquet et al. (2001), Englund et al. (2008), Englund et al. (2009), Bolancé et al. (2003), Brouhns et al. (2003), and Lu (2018). In this branch of literature, Z_t is typically the annual claim count and is conditionally Poisson $\mathcal{P}(X_t)$ distributed given X_t. Moreover, the dynamics of process (X_t) is either assumed to be log-Gaussian AR [see e.g. Brouhns et al. (2003)] or ARG [see e.g. Lu (2018)]. These two parametric restrictions will be relaxed in our framework.

2.2 The discretization scheme

Under Assumptions 1 and 2, the Bayesian premium satisfies, by the law of iterated expectation:

$$\begin{aligned}\mathbb{E}[Z_{T+1}|\underline{Z_T}] &= \mathbb{E}\Big\{\mathbb{E}[Z_{T+1}|X_{T+1}, \underline{Z_T}]|\underline{Z_T}\Big\} \\ &= \int \mathbb{E}[Z_{T+1}|X_{T+1}]\ell(X_{T+1}|\underline{Z_T})\mathrm{d}X_{T+1},\end{aligned}$$

where $\ell(X_{T+1}|\underline{Z_T})$ is the predictive distribution of random effect X_{T+1} given past observations $\underline{Z_T}$. Moreover, $\ell(X_{T+1}|\underline{Z_T})$ satisfies the recursive formula:

$$\ell(X_{T+1}|\underline{Z_T}) \propto \int \ell(Z_T|X_T)\ell(X_T|\underline{Z_{T-1}})\ell(X_{T+1}|X_T)\mathrm{d}X_T, \tag{2.1}$$

where $\propto$ means "proportional to" and the normalization constant c_T is the integral (over X_{T+1}) of the RHS of Equation (2.1):

$$\begin{aligned}c_T &= \iint \ell(Z_T|X_T)\ell(X_T|\underline{Z_{T-1}})\ell(X_{T+1}|X_T)\mathrm{d}X_T\mathrm{d}X_{T+1} \\ &= \int \ell(Z_T|X_T)\ell(X_T|\underline{Z_{T-1}})\mathrm{d}X_T.\end{aligned}$$

The initial value $\ell(X_1|\underline{X_0}) = \ell(X_1)$ is, by definition, the marginal density of process (X_t).

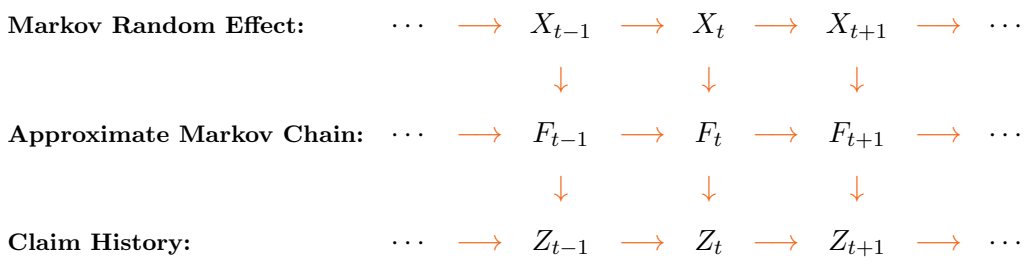
In general, the predictive distribution $\ell(X_{T+1}|\underline{Z_T})$ does *not* belong to any standard parametric family of distributions. Therefore, its value has to be approximated recursively, by computing numerically $\ell(X_1)$, $\ell(X_2|\underline{Z_1})$, $\cdots$, $\ell(X_T|\underline{Z_{T-1}})$ using Equation (2.1). This is the standard computational difficulty in non-linear/non-Gaussian state-space models[5]. Roughly speaking, a state-space model is a model with an unobserved, dynamic process (such as (X_t)), as well as an observable process (such as (Z_t)). Although advances in simulation techniques [see e.g. Pitt and Shephard (1999)] and computing power has partially mitigated this difficulty, it still remains highly intensive, especially in insurance applications, because such simulation needs to be conducted *separately* for each policyholder.

One special case under which simple, recursive updating formulas do exist, regardless of the functional form of $\ell(Z_t|X_t)$, is when (X_t) is discrete with a finite state space, i.e., when (X_t) is a Markov chain[6]. In this case the updating formula is given by the renowned Hamilton filter [see e.g. Kitagawa (1981); Hamilton (1989)]. Nonetheless, specifying a (hidden) Markov model *directly* for the dynamic random effect process is not without downsides. Indeed, the dynamics of a Markov chain is characterized by a (high-dimensional) transition matrix, and thus involves a large number of parameters. On the contrary, models for continuously valued processes are generically much more parsimonious, and are thus much easier and robust when it comes to model estimation. Therefore, in this paper we only consider the case where (X_t) is continuously valued.

One natural solution is therefore to discretize the process $(X_t) = (X_1, ..., X_t)$ into a finite state Markov chain, say (F_t), whose trajectory closely mimics that of (X_t). Specifically, let us denote the state space of (F_t) by $\mathbf{S} = \{s_1, ..., s_K\}$, where K is an integer, and $s_1 < s_2 < \cdots < s_K$, then we have the following causal scheme:

[5]We say that the state-space model is Gaussian if the joint process (X_t, Z_t) is Gaussian. In this case the conditional distribution $\ell(X_{T+1}|\underline{Z_T})$ is also Gaussian and its mean and variance can be estimated using the Kalman filter. However, given that Z_t is typically nonnegative in our application, this Gaussian framework is not appropriate.

[6]As is mentioned in Introduction, another exception is the framework of Lu (2018). However this latter is more restrictive and requires further distributional assumptions and is thus not discussed here.



Then we approximate the continuous conditional density $\ell(X_{T+1}|\underline{Z_T})$ by the discrete conditional probability $\ell(F_{T+1}|\underline{Z_T})$, and approximate the corresponding Bayesian premium by:

$$\mathbb{E}[Z_{T+1}|\underline{Z_T}] \approx \sum_{k=1}^{K} \mathbb{E}[Z_{T+1}|F_{T+1} = s_k]\mathbb{P}[F_{T+1} = s_k|\underline{Z_T}].$$

Moreover, the predictive distribution $\mathbb{P}[F_{T+1} = s_k|\underline{Z_T}]$ satisfies the following recursion, that is the Hamilton filter:

$$\mathbb{P}[F_{T+1} = s_k|\underline{Z_T}] \propto \sum_{i=1}^{K} \ell(Z_T|F_T = s_i)\mathbb{P}(F_T = s_i|\underline{Z_{T-1}})\mathbb{P}(F_{T+1} = s_k|F_T = s_i), \tag{2.2}$$

where the normalization constant is the sum (over k) of the RHS:

$$\begin{aligned} c_T &= \sum_{k=1}^{K}\sum_{i=1}^{K} \ell(Z_T|F_T = s_i)\mathbb{P}[F_T = s_i|\underline{Z_{T-1}}]\mathbb{P}[F_{T+1} = s_k|F_T = s_i] \\ &= \sum_{i=1}^{K} \ell(Z_T|F_T = s_i)\mathbb{P}[F_T = s_i|\underline{Z_{T-1}}]. \end{aligned}$$

The initial values are given by $\mathbb{P}[F_1 = s_k] = \mathbb{P}[F_1 = s_k|\underline{Z_0}]$, $k = 1, ..., N$, that is, the marginal distribution of the Markov chain.

In contrast to the integral in Equation (2.1), Equation (2.2) only involves a finite summation, rather than a complex integration. Therefore, calculating the predictive distribution, $\mathbb{P}[F_{T+1} = s_k|\underline{Z_T}]$, $k = 1, ..., K$, amounts to solving the transition probabilities of the Markov chain, $\mathbb{P}[F_{T+1} = s_k|F_T = s_i]$, for different $i, k = 1, ..., K$.

2.3 Constructing the Markov chain

Let us now specify the approximate Markov chain (F_t) given the continuously valued process (X_t). We assume that the state space $\mathcal{X}$ of (X_t) is a connected interval such as $]-\infty, \infty[$ and $[0, \infty[$, and denote by G the cumulative distribution function (CDF) of its marginal distribution.

There exists an extensive statistical literature on the discretization of a continuously valued Markov process into a finite state Markov chain [see e.g. Tauchen (1986); Tauchen and Hussey (1991); Galindev and Lkhagvasuren (2010); Gospodinov and Lkhagvasuren (2014); Farmer and Toda (2017)]. However, many of these methods either are model-specific[7], or involve rather complicated techniques such as moment matching or entropy minimization. In this subsection we propose a simple method that is applicable to all univariate Markov processes.

First, consider the following partition $I_1, I_2, ..., I_K$ of $\mathcal{X}$:

$$I_k = G^{-1}(]\frac{k-1}{K}, \frac{k}{K}]), \qquad k = 1, ..., K. \tag{2.3}$$

In this way, variable X_t has equal (marginal) probability of falling on each interval I_k. Then, we define the state space of the discrete Markov chain, $\{s_1, ..., s_K\}$, by:

$$s_k = G^{-1}(\frac{2k-1}{2K}), \qquad k = 1, ..., K. \tag{2.4}$$

Note that $\frac{2k-1}{2K}$ is the average of the two bounds of interval $]\frac{k-1}{K}, \frac{k}{K}]$, hence Equation (2.4) ensures that $s_k \in I_k$ for each k. The discrete Markov chain F_t is then defined by:

$$F_t = \ s_k, \text{ if and only if } X_t \in I_k. \tag{2.5}$$

In other words, whenever X_t belongs to interval I_k for a certain k, it is approximated by a constant s_k belonging to the same interval. Since the intervals I_k, $k = 1, ..., K$ form a partition of $\mathcal{X}$, F_t is uniquely defined by the above equation. Then we construct the transition matrix of

[7] For instance, the model of Tauchen (1986) is designed for linear AR(1) processes.

the Markov chain (F_t) as follows:

$$\begin{aligned}\Pi_{ij} &:= \mathbb{P}[F_{t+1} = s_j | F_t = s_i] = \mathbb{P}[X_{t+1} \in I_j | X_t \in I_i] \\ &\approx \mathbb{P}[X_{t+1} \in I_j | X_t = s_i] \\ &= \int_{G^{-1}(\frac{j-1}{K})}^{G^{-1}(\frac{j}{K})} \ell(x_{t+1}|s_i) \mathrm{d}x_{t+1}, \end{aligned} \tag{2.6}$$

that is the integral of the conditional density $\ell(x_{t+1}|s_i)$ over interval I_j. This integral is one-dimensional and can be easily computed numerically. We can remark that when j varies between 1 and K, the RHS of (2.6) sums up to 1, thus we have defined a valid transition matrix of a Markov chain.

Finally, by Perron-Frobenius theorem [see Seneta (2006)], the vector of marginal distribution of the finite state Markov chain (F_t) is the unique[8], normalized left eigen-vector v such that $v' = v'\Pi$. Thus by conducting an eigen-decomposition, we get the vector of initial values used in recursion (2.2):

$$(\mathbb{P}[F_1 = s_1], .., \mathbb{P}[F_1 = s_K])' = v.$$

The quality of the above approximation scheme depends on the fineness of the discretization grid: the larger K, the smaller the approximation error. In the limiting case where K goes to infinity, the approximation error would disappear, and (F_t) would converge to (X_t). However, increasing K also increases the dimension of the transition matrix Π, and hence the computational cost. In Section 4 we will see that a reasonably large K can lead to rather good approximation quality.

It should be noted that the Markov chain obtained by discretizing the dynamic random effect has nothing to do with the Markov chain used in the basic bonus-malus system. Indeed, the premium in the latter system does not result from any Bayesian principle[9], hence the issue of unfairness towards individuals of high basis premium. Instead, in our discretized model, the

[8]Under the assumption that the Markov chain is irreducible, a-periodic, which is implicitly assumed in our paper.

[9]That is, the premium cannot be written as the conditional expectation of future claims given past claims, for a suitable dynamic specification of the claim process.

Bayesian premium is available, making it a fair ratemaking system[10].

Thus under our framework, the pricing procedure of an actuary is the following. Suppose, for instance, that she assumes that for each individual i and time t, $Z_{i,t}$ is conditionally Poisson with parameter $\lambda_i X_{i,t}$, where λ_i is a function (with unknown parameters $\theta_1 \in \mathbb{R}^{m_1}$) of the observed individual characteristics of individual i, and the distribution of the random effect process $(X_{i,t})_t$ is i.i.d. across different individuals, satisfying some simple continuous-valued dynamics. Let us denote by $\theta_2 \in \mathbb{R}^{m_2}$ the unknown parameters characterizing this dynamics. Then the actuary estimates parameters $\theta = (\theta_1', \theta_2')' \in \mathbb{R}^{m_1+m_2}$ with any standard method [see e.g. Bolancé et al. (2003), Lu (2018)]. Finally, she discretizes the dynamics of the random effect and computes the approximate Bayesian premium for each policyholder.

3 Examples on discretization

In this section we illustrate the discretization approach with two popular dynamic specifications of (X_t): the ARG process and the log-Gaussian AR process, both of which have been applied in existing ratemaking literature. For each model, we derive explicit expression of the transition matrix Π of the Markov chain.

3.1 ARG process

The ARG process is the exact time-discretization of the continuous time Cox-Ingersoll-Ross process (also known as the square root diffusion) [see Gouriéroux and Jasiak (2006)]. In the insurance literature, it has been used for ratemaking [see Lu (2018)], for mortality forecasting [see Li and Lu (2018)], as well as for claims reserving [see de Alba and Nieto-Barajas (2008)]. The ARG process is positively valued and is defined along with an intermediate count[11] variable N_t through:

- conditional on $\underline{X_t}$, variable N_t follows Poisson distribution with parameter βX_t.

[10] See e.g. Dionne and Vanasse (1989) for a discussion of why a Bayesian ratemaking rule is necessarily "fair".

[11] Note that this count variable N_t is unobserved. It is introduced to define the exogenous dynamics of ARG process, and is *not* equal to the observed annual number of claims Z_t.

- conditional on $\underline{X_t}$ and $\underline{N_t}$, variable X_{t+1} follows gamma distribution with shape parameter $\delta + N_t$ and scale parameter c.

In the above definition, δ, β, and c are all positive parameters. To summarize, we have the following causal chain:

$$X_t \to N_t \to X_{t+1} \to N_{t+1} \to X_{t+2} \cdots$$

This process is stationary if $\rho := \beta c < 1$, and the terminology of ARG comes from the fact that first, $\mathbb{E}[X_{t+1}|X_t] = \rho X_t + \delta c$, which resembles a linear AR process, and second, its marginal distribution is gamma $\gamma(\delta, c^*)$, i.e., with shape parameter δ and scale parameter $c^* := \frac{c}{1-\rho}$. Let us denote by G_{δ,c^*} the CDF of this gamma distribution.

The conditional density of the process is, by definition:

$$\ell(x_{t+1}|x_t) = \sum_{n=0}^{\infty} \frac{e^{-\beta x_t}\beta^n x_t^n}{n!} \frac{x_{t+1}^{\delta+n-1} e^{-x_{t+1}/c}}{\Gamma(\delta+n)c^{\delta+n}}.$$

Thus the transition matrix Π is given by:

$$\begin{aligned}\Pi_{i,j} &= \int_{G^{-1}_{\delta,c^*}(\frac{j-1}{K})}^{G^{-1}_{\delta,c^*}(\frac{j}{K})} \ell(x_{t+1}|s_i)\mathrm{d}x_{t+1} \\ &= \sum_{n=0}^{\infty} \frac{e^{-\beta s_i}\beta^n s_i^n}{n!}\Big[G_{\delta+n,c}(G^{-1}_{\delta,c^*}(\frac{j}{K})) - G_{\delta+n,c}(G^{-1}_{\delta,c^*}(\frac{j-1}{K}))\Big],\end{aligned}$$

where $s_i = G^{-1}_{\delta,c^*}(\frac{2i-1}{2K})$, $i = 1, ..., K$.

While the above formula involves an infinite summation, it can be easily approximated by truncating the summation at a high, but finite order. In practice, since the transition matrix is common for all the policyholders, it only needs to be computed once.

3.2 Log-Gaussian AR(1) process

It has been assumed in Brouhns et al. (2003) that the annual number of claims is conditionally Poisson distributed with parameter $\lambda \exp(X_t)$, where λ is a deterministic function of the observed individual characteristics of the policyholder, and X_t is an unobserved Gaussian AR(1) process

defined by:

$$X_{t+1} = c + \rho X_t + \epsilon_t, \qquad \forall t,$$

where the autocorrelation coefficient ρ lies between 0 and 1, and ϵ_t is i.i.d. $\mathcal{N}(0, \sigma^2)$ distributed. Under this specification, (X_t) is stationary with Gaussian marginal distribution $\mathcal{N}(\frac{c}{1-\rho}, \frac{\sigma^2}{1-\rho^2})$. Due to the identity $\lambda \exp(X_t) = s\lambda \ \exp(X_t - \log s)$ for any positive constant s, the dynamics of the process (X_t) is identified only up to a additive constant. Thus without loss of generality, we impose the following normalization constraint:

$$\mathbb{E}[\exp(X_t)] = \exp\Big[\frac{c}{1-\rho} + \frac{\sigma^2}{2(1-\rho^2)}\Big] = 1,$$

which is equivalent to $c = -\dfrac{\sigma^2}{2(1+\rho)}$. Then the CDF of X_t becomes:

$$G(x) = \mathbb{P}[X_t < x] = \Phi\Big\{\frac{\sqrt{1-\rho^2}}{\sigma}\big[x + \frac{\sigma^2}{2(1-\rho^2)}\big]\Big\},$$

where Φ is the CDF of the standard normal distribution. On the other hand, the conditional distribution of X_{t+1} given $X_t = x_t$ is:

$$\ell(x_{t+1}|x_t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big[-\frac{(x_{t+1} - c - \rho x_t)^2}{2\sigma^2}\Big].$$

Hence, applying Equation (2.6) leads to the following transition matrix:

$$\begin{aligned}\Pi_{i,j} &= \int_{G^{-1}(\frac{j-1}{K})}^{G^{-1}(\frac{j}{K})} \ell(x_{t+1}|s_i)\mathrm{d}x_{t+1} \\ &= \Phi\Big[\frac{G^{-1}(\frac{j}{K}) - \rho s_i - c}{\sigma}\Big] - \Phi\Big[\frac{G^{-1}(\frac{j-1}{K}) - \rho s_i - c}{\sigma}\Big].\end{aligned}$$

4 Numerical illustration of the pricing accuracy

Let us now apply the proposed Markov chain approximation method to hypothetical policyholders. The approximate Bayesian premium is compared to the Bayesian premium (calculated

with closed form formula whenever available, or otherwise by simulation), as well as the linear credibility premium. Three models are considered, including an ARG-based claim frequency model (Section 4.1), a log-Gaussian-AR-based claim frequency model (Section 4.2), as well as a log-Gaussian frequency-cost model (Section 4.3).

4.1 ARG-based frequency model

We first consider the case where process (X_t) follows the ARG process as defined in Subsection 3.1, and assume that claim frequency Z_t is conditionally Poisson distributed with parameter λX_t, where λ is a known constant. Thus the p.m.f. of Z_t conditional on X_t is:

$$\ell(z|\lambda, X_t) = \frac{e^{-\lambda X_t}(\lambda X_t)^z}{z!}, \qquad \forall z \in \mathbb{N}. \tag{4.1}$$

The advantage of the ARG assumption is that the exact Bayesian premium is available in closed form using the Laplace transform formula derived in Lu (2018). Thus it allows us to rigorously assess the accuracy of our approximation. For each hypothetical claim history, we compare the approximate Bayesian premium, the exact Bayesian premium, as well as the linear credibility premium[12]. Moreover, in order to examine the impact of parameter values on pricing errors, different parameter sets are specified and compared.

The numerical study is designed as follows. First, we compute the premia in three ARG-Poisson models which differ in terms of the value of $\delta \in \{0.75, 1, 2\}$. For each value of δ, we follow the specification in Lu (2018) by letting[13] $\rho = 0.73$ and setting the value of c such that the marginal mean $\mathbb{E}[X_t] = \frac{c}{1-\rho}\delta = 1$. Finally, we set $K = 100$ for the discretization scheme. Although it is theoretically possible to further diminish the approximation error by increasing K, we limit K to 100 for two reasons. First, in some statistical packages such as R, the gamma function, which appears in the conditional and unconditional density formulas, is not defined for arguments larger than a certain threshold. Secondly, further increasing K would swell the computational cost, without introducing substantial improvement of the approximation.

Tables 1 and 2 display the exact Bayesian premium (Bayes.), the discretization-based ap-

[12]The linear credibility premium is computed using standard procedures [see e.g. Bolancé et al. (2003)].
[13]This value of ρ has been previously estimated by Bolancé et al. (2003) using a real insurance database.

proximate Bayesian premium (Disc.), and the linear credibility premium (Cred.) for a set of hypothetical claim histories with $T = 2$ and $T = 3$, respectively. We have expressed the premia in terms of a multiple of the *a priori* premium λ. In other words, in these tables, a premium that is smaller (resp. larger) than 1 corresponds to a policyholder with a bonus (resp. malus) coefficient. Thus the more past claims there are, the higher the premium for period $T + 1$. We can see that the premium obtained from the discretization is closer to the exact Bayesian premium in most cases. Moreover, the approximation is better when Z_1, Z_2 and Z_3 are smaller.

	$\delta = 0.75$			$\delta = 1$			$\delta = 2$		
	Bayes.	Disc.	Cred.	Bayes.	Disc.	Cred.	Bayes.	Disc.	Cred.
$p_3(0,0)$	0.89	0.89	0.90	0.92	0.92	0.92	0.96	0.96	0.96
$p_3(0,1)$	1.75	1.73	1.76	1.57	1.57	1.58	1.30	1.30	1.30
$p_3(1,0)$	1.49	1.48	1.50	1.38	1.37	1.38	1.20	1.20	1.20
$p_3(1,1)$	2.47	2.38	2.37	2.11	2.06	2.04	1.57	1.55	1.55

Table 1: Premium for the $T + 1 = 3$rd period under the ARG-based frequency model, given hypothetical claim histories of the first $T = 2$ periods.

	$\delta = 0.75$			$\delta = 1$			$\delta = 2$		
	Bayes.	Disc.	Cred.	Bayes.	Disc.	Cred.	Bayes.	Disc.	Cred.
$p_4(0,0,0)$	0.87	0.876	0.87	0.95	0.95	0.95	0.90	0.90	0.90
$p_4(0,0,1)$	1.69	1.68	1.72	1.53	1.53	1.55	1.29	1.28	1.29
$p_4(0,1,0)$	1.43	1.43	1.45	1.34	1.34	1.35	1.19	1.18	1.19
$p_4(0,1,1)$	2.38	2.31	2.30	2.05	2.02	2.00	1.55	1.53	1.53
$p_4(1,0,0)$	1.25	1.25	1.26	1.21	1.20	1.22	1.12	1.12	1.12
$p_4(1,0,1)$	2.25	2.19	2.11	1.95	1.91	1.87	1.49	1.47	1.47
$p_4(1,1,0)$	1.90	1.86	1.85	1.70	1.67	1.67	1.38	1.36	1.37
$p_4(1,1,1)$	2.91	2.75	2.69	2.46	2.36	2.32	1.76	1.71	1.71

Table 2: Premium for the $T + 1 = 4$th period under the ARG-based frequency model, given hypothetical claim histories of the first $T = 3$ periods.

4.2 Log-Gaussian-AR frequency model

Let us now assume that the dynamic random effect (X_t) follows a log-Gaussian AR(1) process and that given X_t, the observed count Z_t is Poisson distributed with parameter $\lambda \exp(X_t)$. As Brouhns et al. (2003) point out, in this model, the exact Bayesian premium no longer has

closed form and has to be computed through intensive simulation methods (see Appendix A for details).

We compute the simulation-based Bayesian premium, the Markov-chain-based approximate Bayesian premium, and the linear credibility premium, and express them as multiples of λ, for a set of hypothetical claim histories with $T = 2$ and $T = 3$. Furthermore, we evaluate the pricing errors in two models which differ only in terms of the value of ρ, with $\rho = 0.73$ and 0.9, respectively. For each ρ, we let $\sigma = 1$ and set the value of c such that $\mathbb{E}[\exp(X_t)] = 1$. By the law of large numbers, when the number of simulated trajectories goes to infinity, the simulation-based Bayesian premium converges to the theoretical, intractable Bayesian premium. We have checked that with 50,000 draws, this convergence is approximately attained. Note that this large number of draws spells a significant computational cost for the simulation-based approach. For the Markov chain approximation method, the number of states is chosen to be $K = 100$.

Tables 3 and 4 display the simulated Bayesian premium (Bayes.), the approximate Bayesian premium (Disc.), and the linear credibility premium (Cred.) for a set of hypothetical claim histories with $T = 2$ and $T = 3$, respectively.

	$\rho = 0.73$			$\rho = 0.9$		
	Bayes	Disc.	Cred.	Bayes	Disc.	Cred.
$p_3(0,0)$	0.80	0.80	0.78	0.45	0.46	0.43
$p_3(0,1)$	2.43	2.42	3.07	3.84	4.03	7.81
$p_3(1,0)$	1.67	1.68	1.77	3.25	3.41	1.22
$p_3(1,1)$	3.95	3.84	4.07	8.31	8.44	8.62

Table 3: Premium for the $T + 1 = 3$rd period under the log-Gaussian AR frequency model, for hypothetical claim histories of the first $T = 2$ periods.

	$\rho = 0.73$			$\rho = 0.9$		
	Bayes	Disc.	Cred.	Bayes	Disc.	Cred.
$p_4(0,0,0)$	0.77	0.77	0.74	0.40	0.40	0.42
$p_4(0,0,1)$	2.27	2.32	3.01	3.44	3.57	2.79
$p_4(0,1,0)$	1.59	1.59	1.67	2.21	2.31	1.08
$p_4(0,1,1)$	3.71	3.71	3.93	7.25	7.58	8.45
$p_4(1,0,0)$	1.27	1.27	1.21	1.77	1.78	0.70
$p_4(1,0,1)$	3.15	3.18	3.48	6.67	6.46	8.07
$p_4(1,1,0)$	2.15	2.17	2.14	4.01	4.12	4.35
$p_4(1,1,1)$	4.24	4.45	4.40	9.83	9.78	8.72

Table 4: Premium for the $T+1 = 4$th period under the log-Gaussian AR frequency model, for hypothetical claim histories of the first $T = 3$ periods.

Similar to the ARG example, we can see that the approximate Bayesian premium is in general closer to the simulation-based premium than the linear credibility premium. In particular, the linear credibility premium tends to underestimate the risk when there is no claim in the most recent period, and overestimate the risk otherwise. The estimation error is more substantial with larger ρ. This observation echos the previous analysis by Lu (2018). On the other hand, the approximate Bayesian premium derived from discretization is in general much closer to the simulation-based premium.

4.3 Frequency-cost model

We now take into account both claim frequency and cost in the premium calculation. We denote by $Z_t^{(1)}$ the claim count at time t, $Z_{i,t}^{(2)}$ the amount of the i-th claim in period t, and $Z_t^{(2)} = \sum_{i=1}^{Z_t^{(1)}} Z_{i,t}^{(2)}$ the total claim amount at time t. The distribution of these random variables are summarized in the following assumption.

Assumption 3. *For each t, given the dynamic random effect X_t,*

- *$Z_t^{(1)}$ is conditionally Poisson distributed with mean λX_t;*
- *$Z_{i,t}^{(2)}$ is conditionally i.i.d., exponential distributed with scale parameter $\alpha + \beta X_t$ for each i, where α and β are nonnegative constants.*

As a consequence, given the claim count $Z_t^{(1)}$, the total cost $Z_t^{(2)}$ is conditionally gamma, with scale parameter $\alpha + \beta X_t$ and shape parameter $Z_t^{(1)}$. This result will be useful, if instead of observing individual claim amounts, we observe the annual total claim amounts.

Under the above assumption, the claim frequency and cost random variables are conditionally independent given X_t. Moreover, if $\beta = 0$, then the distribution of the individual cost does not depend on the dynamic random effect and the model reduces to a standard frequency model.

In the simulation study, we further assume that $\alpha = 0$. The joint probability of $Z_t^{(1)}$ and each $Z_{i,t}^{(2)}$ can then be written as:

$$\mathbb{P}[Z_t^{(1)} = n \text{ and } Z_{i,t}^{(2)} = z_i, \forall i = 1, ..., n | \lambda, \beta, X_t] = \mathbb{P}[Z_t^{(1)} = n | \lambda, X_t] \prod_{i=1}^{n} \mathbb{P}[Z_{i,t}^{(2)} = z_i | \beta, X_t]. \quad (4.2)$$

Denote by $\underline{Z_T^{(2)}}$ the history of claim amounts up to time T. The Bayesian premium at time $T+1$, given past information of both claim counts $\underline{Z_T^{(1)}}$ and amounts $\underline{Z_T^{(2)}}$, is equal to:

$$\begin{aligned}
& \mathbb{E}[Z_{T+1}^{(2)} | \underline{Z_T^{(1)}}, \underline{Z_T^{(2)}}, \beta, \lambda] \\
= \; & \mathbb{E}[\sum_{i=1}^{Z_{T+1}^{(1)}} Z_{i,T+1}^{(2)} | \underline{Z_T^{(1)}}, \underline{Z_T^{(2)}}, \beta, \lambda] \\
= \; & \int_0^\infty \mathbb{E}[\sum_{i=1}^{Z_{T+1}^{(1)}} Z_{i,T+1}^{(2)} | \underline{Z_T^{(1)}}, \underline{Z_T^{(2)}}, X_{T+1}, \beta, \lambda,] \ell(X_{T+1} | \underline{Z_T^{(1)}}, \underline{Z_T^{(2)}}) dX_{T+1} \\
= \; & \int_0^\infty \mathbb{E}\Big\{\mathbb{E}[\sum_{i=1}^{Z_{T+1}^{(1)}} Z_{i,T+1}^{(2)} | Z_{T+1}^{(1)}, \underline{Z_T^{(1)}}, \underline{Z_T^{(2)}}, X_{T+1}, \beta, \lambda]\Big\} \ell(X_{T+1} | \underline{Z_T^{(1)}}, \underline{Z_T^{(2)}}) dX_{T+1} \\
= \; & \int_0^\infty \mathbb{E}[Z_{T+1}^{(1)} | \underline{Z_T^{(1)}}, \underline{Z_T^{(2)}}, X_{T+1}, \lambda] \mathbb{E}[Z_{1,T+1}^{(2)} | \underline{Z_T^{(1)}}, \underline{Z_T^{(2)}}, X_{T+1}, \beta] \ell(X_{T+1} | \underline{Z_T^{(1)}}, \underline{Z_T^{(2)}}) dX_{T+1} \\
= \; & \int_0^\infty \lambda X_{T+1} \beta X_{T+1} \ell(X_{T+1} | \underline{Z_T^{(1)}}, \underline{Z_T^{(2)}}) dX_{T+1} \\
= \; & \int_0^\infty \lambda \beta X_{T+1}^2 \ell(X_{T+1} | \underline{Z_T^{(1)}}, \underline{Z_T^{(2)}}) dX_{T+1}. \qquad (4.3)
\end{aligned}$$

Let us now evaluate the premium using the log-Gaussian AR model. The parameter values are set as follows: $\rho = 0.73$, $\sigma = 1$, and $\beta = 100$. The value of c is set such that $\mathbb{E}[\exp(X_t)] = 1$. The Bayesian premium is computed by simulation with 50,000 draws (see Appendix A.2 for more details). The Markov chain approximated premium is computed with $K = 100$.

We illustrate the frequency-cost model with $T = 2$. Table 5 reports the simulated Bayesian premium and the approximate premium derived from discretization under both the frequency-cost model and the frequency model. Again, all premia are expressed as multiples of their respective *a priori* premium. In each period, when the number of claim is positive, we assume that the claim amount is either 100 (Column 2 to 5), or 200 (Column 6 to 9). We can see that, in this model, the approximate premium is still rather close to the simulated premium. Moreover, we can remark that the insurance premium depends heavily on the amounts of previous claims. This demonstrates the necessity of considering both claims and amounts in the pricing exercise.

	Claim amount = 100				Claim amount = 200			
	Total amount		Claim number		Total amount		Claim number	
	Bayes.	Disc.	Bayes.	Disc.	Bayes.	Disc.	Bayes.	Disc.
$p_3(0,1)$	2.02	2.11	2.05	2.05	3.70	3.68	2.76	2.80
$p_3(1,0)$	1.13	1.19	1.55	1.56	2.04	2.17	2.17	2.17
$p_3(1,1)$	3.02	3.07	2.48	2.47	4.88	4.80	3.19	3.21

Table 5: Premium for the $T+1 = 3$rd period under the log-Gaussian AR frequency-cost model with $\rho = 0.73$, for hypothetical claim histories of the first $T = 2$ periods.

Two remarks are made regarding the proposed frequency-cost model. First, we make the simplifying assumption that both the claim frequency and cost per claim are governed by the same random effect, X_t. While this assumption reconciles many existing claim frequency-cost models[14], we note that a more general, non-degenerate joint distribution for multiple random

[14]For instance, Shi et al. (2015) use a copula to characterize the random effects governing the claim frequency and severity, respectively. They have considered, among others, Archimedean copulas such as Gumbel, Clayton and Frank copulas, which have an interpretation of a (single) factor representation [see e.g. Oakes (1989)]. Therefore, the model discussed in thus Subsection can be viewed as a dynamic extension of Shi et al. (2015), in which the single factor is time-varying. Moreover, among models considering different types of claims, Shi and Valdez (2014) propose to model the dependence between different frequency variables through Archimedean copulas [see also Wen et al. (2009); Zhang et al. (2018) for an analysis of the credibility premium in such a model]. More remotely, models in which a single random effect is introduced to capture the dependence between two risk variables is widely studied in actuarial science under the terminology of background risk models, both theoretically [see e.g. Côté and Genest (2018)] and empirically [see e.g. Frees et al. (1996); Luciano et al. (2008)

effects could lead to more realistic ratemaking results in many insurance applications. For instance, Pinquet (1997) uses a portfolio of automobile policyholders of a French insurance company, and finds that correlation between the random effects governing the claim frequency and cost is very low. Second, costs of claim are represented by exponential distributions, which means that the mode for costs per claim is zero. While this is true in some cases (examples of such datasets include the motor insurance portfolios used in Yip and Yau (2005), Yang et al. (2018), De Jong and Heller (2008), and Lu (2019)), it does not apply to all insurance applications. In the latter cases, other distributions leading to a positive mode should be considered.

5 Pricing real-life insurance policies

In this section, we illustrate the usefulness of the proposed approximation scheme by analyzing a hypothetical non-life insurance portfolio constructed from real data. In particular, a unique dataset recording historical storm events in different locations across the U.S. is utilized. The insurance policies we consider have long loss experience (longer than 30 years) and occasionally large claim numbers. This renders existing methods unfeasible, and highlights the appeal of the proposed Markov chain approximation framework.

5.1 Data and model assumptions

We consider a hypothetical catastrophe insurance policy, which protects against losses from severe storm events above a certain deductible level, D. In other words, for each policyholder j, the annual loss frequency of the insurance portfolio is $Z_{j,t} = \sum_{i=1}^{M_{j,t}} \mathbb{I}_{\{L_{i,j,t}>D\}}$, $t = 1, \ldots, T$, where $L_{i,j,t}$ is the economic damage of the ith storm event of policyholder j in year t, and $M_{j,t}$ is the total number of storms in year t of policyholder j. Here, we take each geographic location as a hypothetical policyholder. A natural source of the random effect is the geographic heterogeneity effect. Due to the low-frequency-high-severity nature, as well as the geographic heterogeneity, forecasting future loss frequencies is usually a challenging task, which requires

for joint annuity pricing].

typically a long observation history. The Markov chain approximation provides a convenient way to forecast into the future based on historical records.

The storm loss frequency dataset is constructed from the Storm Events Database of the National Oceanic and Atmospheric Administration (NOAA; `https://www.ncdc.noaa.gov/stormevents/`), which documents the occurrence of storms and other economically significant weather-linked events across the U.S. The dataset covers the period 1986-2016. In this study, we consider a deductible level of $D = 25$ millions (USD), that is, a payment is triggered whenever the economic damage of a storm event exceeds D. We analyze all of the 50 States that are indemnified at least once over the entire sample period[15]. Figure 1 displays $Z_{j,t}$, the paths of loss process for each state, where where $j = 1, \ldots, J$, and $J = 50$ is the total number of states included in the dataset. In particular, we can remark the large magnitude of loss frequencies. For example, the peak in 1995 corresponds to the Atlantic hurricane season, which caused 92 severe economic damages in excess of 25 millions in the state of Georgia. The loss processes in Figure 1 also highlights some other catastrophic storm events including the Hurricane Katrina in 2005, the 2016 Southern and Midwestern floods, etc. Summary statistics of the storm loss frequency are displayed in Table 6. For each state, the mean, standard deviation (Std.Dev.), coefficient of variation (CV, defined as the ratio of the standard deviation over the mean), skewness, and kurtosis are reported.

5.2 Pricing using the Markov chain approximation

We assume that the loss frequency due to storm events in year t for state j follows a conditional Poisson distribution with intensity $\lambda_j \exp(X_{j,t})$, i.e., $Z_{j,t}|X_{j,t} \sim \text{Poi}(\lambda_j \exp(X_{j,t}))$, where λ_j is the fixed effect parameter. Further, for a given j, $(X_{j,t})$ is specified as an unobservable AR(1) process with autocorrelation coefficient ρ and i.i.d. Gaussian $\mathcal{N}(0, \sigma^2)$ residuals. We use the Generalized Method of Moments (GMM) to estimate the parameters $\lambda_1, ..., \lambda_J, \sigma, \rho$ and the estimation procedure is described in Appendix B.

Note that this specification is quite general and does not say anything about the inter-

[15]This includes 49 States and the District of Columbia. Only the state of New Hampshire is excluded since it does not have a loss larger than 25 million USD over the whole sample period.

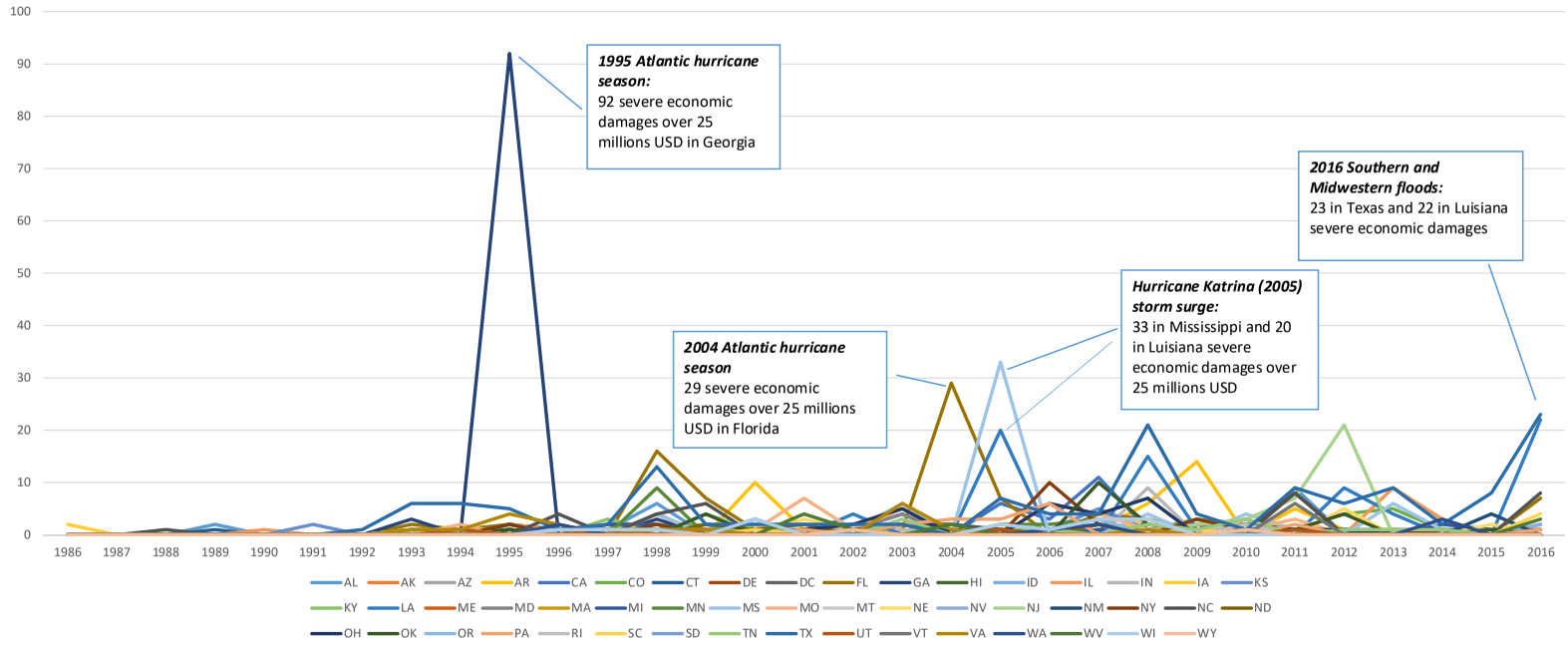


Figure 1: Time series of the loss frequency above a deductible level of 2.5 Million USD for each State.

Table 6: Summary Statistics of Storm Frequency Data for Each State. The table summarizes, for each State, the mean, standard deviation (Std.Dev.), coefficient of variation (CV, defined as the ratio of the standard deviation over the mean), skewness, and kurtosis.

STATE	Mean	Std.Dev.	CV	Skewness	Kurtosis	STATE	Mean	Std.Dev.	CV	Skewness	Kurtosis
AL	0.677	1.922	2.837	3.627	13.426	MO	1.097	1.904	1.736	2.022	3.297
AK	0.032	0.180	5.568	5.568	31.000	MT	0.097	0.539	5.568	5.568	31.000
AZ	0.290	0.588	2.027	1.958	2.968	NE	0.516	1.092	2.115	2.915	9.557
AR	1.387	3.201	2.308	2.974	8.899	NV	0.129	0.428	3.313	3.561	13.021
CA	1.258	2.250	1.789	3.220	11.939	NJ	1.452	3.940	2.714	4.428	21.539
CO	0.871	1.258	1.444	1.869	3.521	NM	0.226	0.762	3.375	4.417	21.322
CT	0.032	0.180	5.568	5.568	31.000	NY	0.935	2.308	2.468	3.160	9.984
DE	0.032	0.180	5.568	5.568	31.000	NC	1.194	2.330	1.952	2.117	3.577
DC	0.032	0.180	5.568	5.568	31.000	ND	0.387	0.667	1.724	1.523	1.122
FL	2.677	5.946	2.221	3.476	13.400	OH	0.968	1.906	1.969	2.180	3.882
GA	3.387	16.470	4.862	5.542	30.804	OK	0.903	2.006	2.221	3.532	14.403
HI	0.065	0.250	3.871	3.728	12.717	OR	0.194	0.601	3.106	3.853	16.436
ID	0.097	0.301	3.106	2.868	6.654	PA	0.710	1.346	1.897	2.498	7.232
IL	0.774	1.726	2.230	3.905	17.837	RI	0.065	0.359	5.568	5.568	31.000
IN	0.452	1.630	3.609	5.130	27.511	SC	0.226	0.762	3.375	4.417	21.322
IA	0.323	0.748	2.318	2.448	5.549	SD	0.032	0.180	5.568	5.568	31.000
KS	0.710	1.039	1.464	2.540	8.948	TN	0.548	0.850	1.550	1.753	2.819
KY	0.548	0.925	1.687	1.604	1.549	TX	4.581	5.702	1.245	2.032	4.286
LA	2.645	5.834	2.206	2.555	5.708	UT	0.161	0.374	2.318	1.937	1.868
ME	0.129	0.718	5.568	5.568	31.000	VT	0.194	1.078	5.568	5.568	31.000
MD	0.194	0.749	3.871	4.745	23.925	VA	0.581	1.311	2.258	3.134	10.550
MA	0.097	0.301	3.106	2.868	6.654	WA	0.258	0.729	2.824	2.865	7.551
MI	0.516	0.851	1.649	1.506	1.266	WV	0.323	0.945	2.929	3.087	9.083
MN	0.645	1.723	2.671	4.144	19.363	WI	0.484	0.962	1.987	1.972	2.726
MS	1.968	6.036	3.067	4.841	25.155	WY	0.065	0.250	3.871	3.728	12.717

dependence between the dynamic random effects processes $(X_{j,t})$ for different j. In fact, such dependence could arise, especially between (geographically) neighbouring states. However, while such dependence is not taken into account, the GMM estimation will still lead to consistent estimators. The reason is that GMM estimation approach only focuses on (joint) moments between different $X_{j,t}$'s within each same state j, and is *robust* to the presence of such dependence. In the pricing stage, such a dependence will also be neglected in our Bayesian premium formula. We acknowledge that there might be efficiency losses by ignoring this dependence and leave this issue for future research.

Following the approximation scheme in Subsection 3.2, the Bayesian premia for year $T+1$ are calculated in Table 7, with the number of states of the Markov chain approximation set as $K = 100$. For each panel, the first column stands for the abbreviation of state names; the second column displays the Bayesian premia for existing policyholders (Bayes Premium); and the third column displays the Bayesian premium for a new policyholder with no claim history (Bayes Premium New). We can see that the two Bayesian premia are close to each other, with some states having slightly higher Bayesian premia for existing policyholders and the others having higher premia for new policyholders.

The pricing results are also illustrated in Figure 2, where the orange line shows the Bayesian premia for the existing policyholders; the grey line shows the Bayesian Premium for new policyholders with no claim histories; and the historical claim frequency average is displayed with the blue line. We can see that for some states (e.g., GA and TX), the historical average is quite far away from the the other two Bayesian premia. This result highlights the importance of reweighting historical records according to their seniority (i.e., assigning more weights to recent claims). Indeed, models assigning the same weights to historical records may fail to capture the time-varying systemic weather risk and thus induce mispricing [see Borman et al. (2013); Woodard et al. (2011)]. There have been a growing literature that tries to develop more comprehensive and scientific approaches to reweight historical losses to adjust for trending risk factors [see e.g. Coble et al. (2011); Borman et al. (2013); Woodard (2014); Porth et al. (2014); Zhu et al. (2018)]. In this illustrative pricing example, we show that using a dynamic random effect, the Bayesian ratemaking effectively assigns higher weights to more recent records, and

hence achieves more scientific pricing results.

Table 7: Bayesian premium for Year $T + 1 = 32$. The table displays the Bayesian premia for existing policyholders (Bayes. Premium) and the Bayesian premium for a new policyholder with no claim history (Bayes. Premium New).

STATE	Bayes Prem	Bayes Prem New	STATE	Bayes Prem	Bayes Prem New
AL	0.32	0.35	MO	0.78	0.88
AK	0.04	0.05	MT	0.03	0.03
AZ	0.29	0.32	NE	0.38	0.39
AR	0.72	0.82	NV	0.12	0.09
CA	0.97	0.96	NJ	0.64	0.74
CO	0.76	0.88	NM	0.13	0.11
CT	0.04	0.05	NY	0.48	0.54
DE	0.04	0.05	NC	1.49	0.84
DC	0.04	0.05	ND	0.49	0.46
FL	2.34	1.60	OH	0.61	0.69
GA	0.82	0.96	OK	0.54	0.58
HI	0.13	0.13	OR	0.11	0.12
ID	0.23	0.24	PA	0.58	0.56
IL	0.47	0.50	RI	0.02	0.02
IN	0.18	0.18	SC	0.19	0.11
IA	0.24	0.25	SD	0.04	0.05
KS	0.87	0.76	TN	0.55	0.61
KY	0.48	0.54	TX	7.35	4.21
LA	2.98	1.59	UT	0.49	0.54
ME	0.04	0.04	VT	0.05	0.05
MD	0.08	0.08	VA	0.36	0.39
MA	0.23	0.24	WA	0.16	0.16
MI	0.51	0.54	WV	0.27	0.18
MN	0.33	0.35	WI	0.37	0.40
MS	0.76	0.88	WY	0.13	0.13

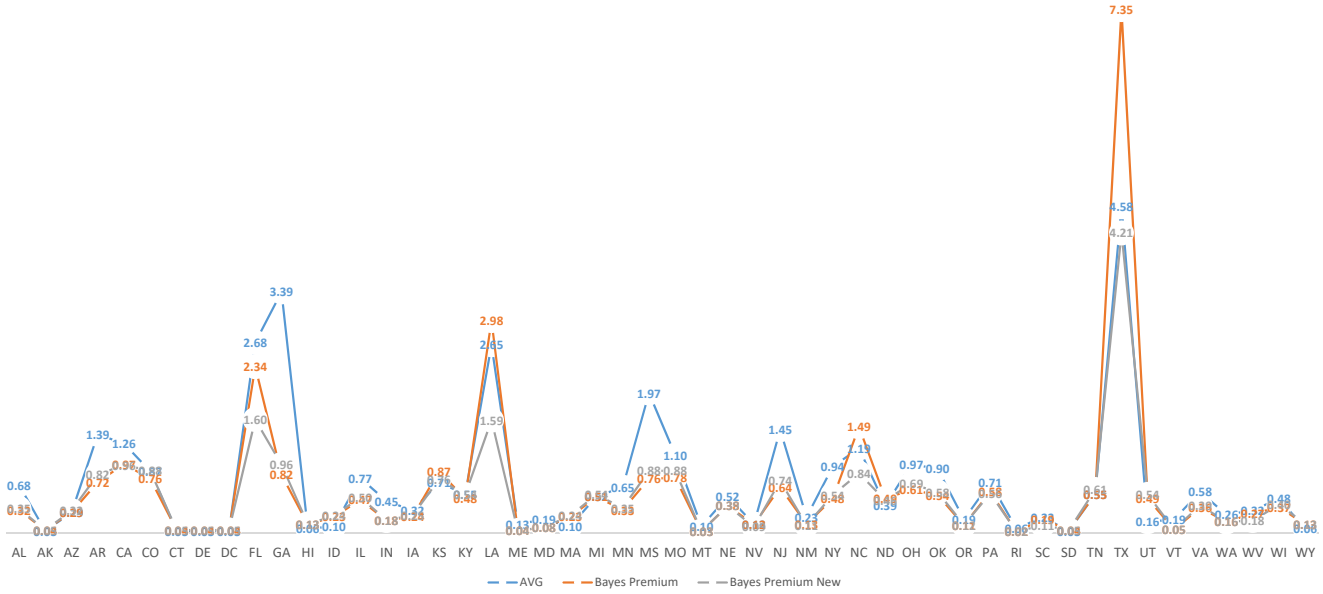


Figure 2: Bayesian premiums for Year $T+1=32$. Orange line shows the Bayesian premia for the existing policyholders. Grey line shows the Bayesian Premium for new policyholders with no claim histories. Blue line shows the historical claim frequency average.

6 Conclusion

We have proposed a simple and accurate Markov chain approximation approach for dynamic Bayesian ratemaking. Dynamic random effect model is popular due to its flexibility to capture the time-varying risk of policyholders. However, its appeal is currently limited since the Bayesian premium is typically intractable. Linear credibility model, although simpler, suffers from substantial pricing errors and is not able to include both the severity and the frequency pricing components. In this paper we have introduced the Markov chain approximation method to the actuarial literature, which strikes a balance between pricing accuracy and computational efficiency. Through hypothetical numerical examples, we have demonstrated the accuracy of our method with respect to the Bayesian approach and the credibility approach, as well as the wide applicability of our approximation method to various existing pricing models. Finally, we have demonstrated the versatility of our framework on a real-life example with an unusually long claim history.

There are several interesting directions for future research, among which we can think about situations when both claim frequency and cost are present. In the paper we have focused on the degenerate case where the two variables share the same random effect. In future research it would be interesting to develop approximation algorithm for models assigning different random effects to these two variables. Such an extension involves the approximation of bivariate Markov processes and is thus more computationally demanding, but the econometric literature has already made some attempts in this direction [see e.g. Galindev and Lkhagvasuren (2010)] that might deserve our attention.

Appendices

A Computing the Bayesian premium via simulation

In this appendix we introduce, for both the frequency model and the frequency-cost model, the simulation algorithm when the Bayesian premium does not have a closed form expression. This is for instance the case in the frequency model in which the dynamic random effect follows a log-Gaussian AR process.

A.1 Frequency model

Under the frequency model, the goal is to determine

$$\begin{aligned}
&\mathbb{E}[Z_{T+1}|\lambda, \underline{z_T}] \\
&= \int_0^\infty \mathbb{E}[Z_{T+1}|\lambda, X_{T+1}, \underline{z_T}]\ell(X_{T+1}|\underline{z_T})dX_{T+1} \\
&= \int_0^\infty \lambda \exp(X_{T+1})\ell(X_{T+1}|\underline{z_T})dX_{T+1}, \qquad \text{(A.1)}
\end{aligned}$$

where $\ell(X_{T+1}|\underline{z_T})$ is the predictive distribution of X_{T+1} given $\underline{z_T}$. The simulation procedure is as follows.

First, simulate J trajectories of process $\underline{X_{T+1}} = (X_0, X_1, X_2, ..., X_{T+1})$. For each trajectory, X_0 is simulated using the stationary normal distribution with mean $\frac{c}{1-\rho}$ and variance $\frac{\sigma^2}{1-\rho^2}$, then X_1, X_2, ... X_{T+1} are simulated iteratively using the Gaussian conditional distribution of the AR(1) dynamics. We denote by $\underline{x_t}^{(j)}$, $t = 0, 1, ..., T+1$, the j-th simulated trajectory of (X_t) up to time $T+1$.

Then the RHS of (A.1) is approximated by:

$$\mathbb{E}[Z_{T+1}|\lambda, \underline{z_T}] \approx \sum_{j=1}^{J} \lambda \exp(x_{T+1}^{(j)}) p(x_{T+1}^{(j)}|\underline{z_T}),$$

where $p(x_{T+1}^{(j)}|\underline{z_T}), j = 1, ..., J$, is a vector of weights that sum up to 1:

$$\sum_{j=1}^{J} p(x_{T+1}^{(j)}|\underline{z_T}) = 1,$$

and is defined, up to a normalization constant, by:

$$p(x_{T+1}^{(j)}|\underline{z_T}) \propto \prod_{t=1}^{T} \ell(z_t|x_t^{(j)}).$$

In other words, the different simulated trajectories are weighted according to the conditional likelihood of the observed counts $\underline{Z_T}$ given $\underline{X_T}$.

A.2 Frequency-cost model

The goal is to determine:

$$\mathbb{E}[Z_{T+1}^{(2)}|\underline{z_T^{(1)}}, \underline{z_T^{(2)}}, \beta, \lambda] = \int_0^{\infty} \lambda\beta \exp(2X_{T+1}) \ell(X_{T+1}|\underline{z_T^{(1)}}, \underline{z_T^{(2)}}) \mathrm{d}X_{T+1} \tag{A.2}$$

After simulating different trajectories of (X_t^j) up to time $T+1$ for $j = 1, ..., J$, we approximate the above integral by:

$$\mathbb{E}[Z_{T+1}^{(2)}|\underline{z_T^{(1)}}, \underline{z_T^{(2)}}, \beta, \lambda] \approx \sum_{j=1}^{J} \lambda\beta \exp(2x_{T+1}^{(j)}) p(x_{T+1}^{(j)}|\underline{z_T^{(1)}}, \underline{z_T^{(2)}})$$

where the vector of weights $p(x_{T+1}^{(j)}|\underline{z_T^{(1)}}, \underline{z_T^{(2)}})$ sums up to unity, and is defined, up to a proportionality constant, by:

$$p(x_{T+1}^{(j)}|\underline{z_T^{(1)}}, \underline{z_T^{(2)}}) \propto \prod_{t=1}^{T} \Big[\ell(z_t^{(1)}|x_t^{(j)}) \prod_{i=1}^{z_t^{(1)}} \ell(Z_{i,t}^{(2)}|X_t^{(j)})\Big],$$

that is the joint conditional distribution of the number of claims $z_t^{(1)}, t = 1, ..., T$, as well as all the individual claim amounts.

B Generalized method of moments estimation

This section discusses the estimation of model parameters using the Generalized Method of Moments (GMM). We assume that the observable storm events loss frequency random variables in year t for the state j, $Z_{j,t}$, are conditionally independent given $X_{i,t}$, and follow a conditional Poisson distribution with intensity $\lambda_j \exp(X_{j,t})$, where $t = 1, \ldots, T$ and $j = 1, \ldots, J = 50$, i.e.,

$$Z_{j,t}|X_{j,t} \sim \text{Poi}(\lambda_j \exp(X_{j,t})), \quad t = 1, \ldots, T, j = 1, \ldots, J, \tag{B.1}$$

where λ_j is the fixed effect parameter. Further, for a given j, $(X_{j,t})$ is specified as an AR(1) process with autocorrelation coefficient ρ and i.i.d. Gaussian residuals:[16]:

$$X_{j,t} = \rho X_{j,t-1} + \epsilon_{j,t}, \tag{B.2}$$

where $\epsilon_{j,t} \sim \mathcal{N}(0, \sigma^2)$. Let $\theta^2 = \dfrac{\sigma^2}{1-\rho^2}$.

Note that for each fixed j and any t we have:

$$\mathbb{E}[Z_{j,t}] = \mathbb{E}[\mathbb{E}(Z_{j,t}|X_{j,t})] = \lambda_j \mathbb{E}[\exp(X_{j,t})] = \lambda_j \cdot \exp\left(\frac{1}{2}\theta^2\right). \tag{B.3}$$

[16] In this specification we have assumed that $\mathbb{E}[X_{j,t}] = 0$. This is due to the fact that $\exp(X_{j,t})$ (resp. $X_{j,t}$) is only identified up to a multiplicative (resp. additive) constants. Thus for identification reasons there is no loss of generality to impose $\mathbb{E}[X_{j,t}] = 0$.

A natural estimate of $\mathbb{E}[Z_{j,t}]$ is the sample mean, $\frac{1}{T}\sum_{t=1}^{T} Z_{j,t}$, hence,

$$\lambda_j \cdot \exp\left(\frac{1}{2}\theta^2\right) \approx \frac{1}{T}\sum_{t=1}^{T} Z_{j,t}. \tag{B.4}$$

Thus the fixed effect λ_j can be expressed as a function of the sample moments and the parameters of the AR(1) process, according to Equation (B.4). Other moment conditions are chosen to be:

$$\begin{aligned}
\mathbb{E}[Z_{j,t}^2] &= \mathbb{E}[\mathbb{E}(Z_{j,t}^2|X_{j,t})] = \lambda_j \mathbb{E}[\exp(X_{j,t})] + \lambda_j^2 \mathbb{E}[\exp(2X_{j,t})] \\
&= \lambda_j \cdot \exp\left(\frac{1}{2}\theta^2\right) + \lambda_j^2 \cdot \exp\left(2\theta^2\right), \qquad \text{(B.5)} \\
\mathbb{E}[Z_{j,t}Z_{j,t-1}] &= \mathbb{E}[\mathbb{E}(Z_{j,t}Z_{j,t-1}|X_{j,t}, X_{j,t-1})] = \lambda_j^2 \mathbb{E}[\exp(X_{j,t}X_{j,t-1})] \\
&= \lambda_j^2 \exp\left((1+\rho)\theta^2\right). \qquad \text{(B.6)}
\end{aligned}$$

Likewise, the theoretical variance and lag-1 covariance of the process $(Z_{j,t})_t$ can be replaced by its empirical counterparts $\tilde{\mathbb{E}}[Z_{j,t}^2]$ and $\tilde{\mathbb{E}}[Z_{j,t}Z_{j,t-1}]$. Thus we obtain a total of $2I$ moment conditions, where I is the number of individuals, and 2 parameters (ρ, θ) to calibrate. We choose to be the minimiser of the term:

$$\sum_{j=1}^{I} \left(\tilde{\mathbb{E}}[Z_{j,t}^2] - \mathbb{E}[Z_{j,t}^2]\right)^2 + \left(\tilde{\mathbb{E}}[Z_{j,t}Z_{j,t-1}] - \mathbb{E}[Z_{j,t}Z_{j,t-1}]\right)^2.$$

The estimation results are summarized in Table 8 below.

Table 8: GMM Estimation Results. The first panel displays the induced estimates of the fixed effect parameter, λ_j, for each state. The bottom panel displays the estimated parameters in the AR(1) process. "*" and "***" indicate statistical significance at 10% and 1% level, respectively.

STATE	λ_j	STATE	λ_j	STATE	λ_j	STATE	λ_j
AL	0.246	IL	0.358	MT	0.021	SC	0.080
AK	0.032	IN	0.131	NE	0.275	SD	0.032
AZ	0.225	IA	0.178	NV	0.063	TN	0.436
AR	0.586	KS	0.539	NJ	0.524	TX	3.000
CA	0.682	KY	0.386	NM	0.080	UT	0.387
CO	0.626	LA	1.129	NY	0.381	VT	0.037
CT	0.032	ME	0.026	NC	0.599	VA	0.278
DE	0.032	MD	0.059	ND	0.329	WA	0.114
DC	0.032	MA	0.174	OH	0.494	WV	0.127
FL	1.136	MI	0.387	OK	0.411	WI	0.285
GA	0.686	MN	0.251	OR	0.083	WY	0.093
HI	0.093	MS	0.625	PA	0.397		
ID	0.174	MO	0.623	RI	0.016		
	θ		0.825^{***}				
	ρ		0.308^{*}				

References

Abdallah, A., Boucher, J.-P., and Cossette, H. (2016). Sarmanov family of multivariate distributions for bivariate dynamic claim counts model. *Insurance: Mathematics and Economics*, 68:120–133.

Bolancé, C., Guillén, M., and Pinquet, J. (2003). Time-varying credibility for frequency risk models: estimation and tests for autoregressive specifications on the random effects. *Insurance: Mathematics and Economics*, 33(2):273–282.

Borman, J. I., Goodwin, B. K., Coble, K. H., Knight, T. O., and Rejesus, R. (2013). Accounting for short samples and heterogeneous experience in rating crop insurance. *Agricultural Finance Review*, 73(1):88–101.

Boucher, J.-P. and Guillen, M. (2011). A semi-nonparametric approach to model panel count data. *Communications in Statistics-Theory and Methods*, 40(4):622–634.

Brouhns, N., Guillén, M., Denuit, M., and Pinquet, J. (2003). Bonus-malus scales in segmented tariffs with stochastic migration between segments. *Journal of Risk and Insurance*, 70(4):577–599.

Coble, K. H., Miller, M. F., Rejesus, R. M., Boyles, R., Knight, T. O., and Goodwin, B. K. (2011). Methodology analysis for weighting of historical experience. *USDA Risk Management Agency Technical Report*.

Côté, M.-P. and Genest, C. (2018). Dependence in a background risk model. *Journal of Multivariate Analysis*.

Cox, J. C., Ross, S. A., Rubinstein, M., et al. (1979). Option pricing: A simplified approach. *Journal of financial Economics*, 7(3):229–263.

Czado, C., Kastenmeier, R., Brechmann, E. C., and Min, A. (2012). A mixed copula model for insurance claims and claim sizes. *Scandinavian Actuarial Journal*, 2012(4):278–305.

de Alba, E. and Nieto-Barajas, L. E. (2008). Claims reserving: a correlated Bayesian model. *Insurance: Mathematics and Economics*, 43(3):368–376.

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

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

Duan, J.-C. and Simonato, J.-G. (2001). American option pricing under GARCH by a Markov chain approximation. *Journal of Economic Dynamics and Control*, 25(11):1689–1718.

Englund, M., Guillén, M., Gustafsson, J., Nielsen, L. H., and Nielsen, J. P. (2008). Multivariate latent risk: A credibility approach. *ASTIN Bulletin*, 38(1):137–146.

Englund, M., Gustafsson, J., Nielsen, J. P., and Thuring, F. (2009). Multidimensional credibility with time effects: An application to commercial business lines. *Journal of Risk and Insurance*, 76(2):443–453.

Farmer, L. E. and Toda, A. A. (2017). Discretizing nonlinear, non-gaussian markov processes with exact conditional moments. *Quantitative Economics*, 8(2):651–683.

Frangos, N. E. and 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*, 31(01):1–22.

Frees, E., Carriere, J., and Valdez, E. (1996). Annuity Valuation with Dependent Mortality. *Journal of Risk and Insurance*, 63(2):229–261.

Frees, E. W., Gao, J., and Rosenberg, M. A. (2011). Predicting the frequency and amount of health care expenditures. *North American Actuarial Journal*, 15(3):377–392.

Galindev, R. and Lkhagvasuren, D. (2010). Discretization of highly persistent correlated AR (1) shocks. *Journal of Economic Dynamics and Control*, 34(7):1260–1276.

Garrido, J., Genest, C., and Schulz, J. (2016). Generalized linear models for dependent frequency and severity of insurance claims. *Insurance: Mathematics and Economics*, 70:205–215.

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

Gospodinov, N. and Lkhagvasuren, D. (2014). A moment-matching method for approximating vector autoregressive processes by finite-state Markov chains. *Journal of Applied Econometrics*, 29(5):843–859.

Gouriéroux, C. and Jasiak, J. (2006). Autoregressive Gamma Processes. *Journal of Forecasting*, 25(2):129–152.

Gschlößl, S. and Czado, C. (2007). Spatial modelling of claim frequency and claim size in non-life insurance. *Scandinavian Actuarial Journal*, 2007(3):202–225.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384.

Kitagawa, G. (1981). A nonstationary time series model and its fitting by a recursive filter. *Journal of Time Series Analysis*, 2(2):103–116.

Lemaire, J. and Zi, H. (1994). A comparative analysis of 30 bonus-malus systems. *ASTIN Bulletin*, 24(2):287–309.

Li, H. and Lu, Y. (2018). A Bayesian non-parametric model for small population mortality. *Scandinavian Actuarial Journal*, 2018(7):605–628.

Lu, Y. (2018). Dynamic frailty count process in insurance: A unified framework for estimation, pricing, and forecasting. *Journal of Risk and Insurance*, 85(4):1083–1102.

Lu, Y. (2019). Flexible (panel) regression models for bivariate count-continuous data with an insurance application. *forthcoming in Journal of the Royal Statistical Society: Series A (Statistics in Society)*.

Luciano, E., Spreeuw, J., and Vigna, E. (2008). Modelling Stochastic Mortality for Dependent Lives. *Insurance: Mathematics and Economics*, 43(2):234–244.

Oakes, D. (1989). Bivariate Survival Models Induced by Frailties. *Journal of the American Statistical Association*, 84(406):487–493.

Pinquet, J. (1997). Allowance for cost of claims in bonus-malus systems. *Astin Bulletin*, 27(01):33–57.

Pinquet, J. (2018). Poisson models with dynamic random effects and nonnegative credibilities per period. *Researchgate preprint*.

Pinquet, J., Guillén, M., and Bolancé, C. (2001). Allowance for the age of claims in bonus-malus systems. *Astin Bulletin*, 31(02):337–348.

Pitt, M. K. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94(446):590–599.

Porth, L., Zhu, W., and Seng Tan, K. (2014). A credibility-based Erlang mixture model for pricing crop reinsurance. *Agricultural Finance Review*, 74(2):162–187.

Seneta, E. (2006). *Non-Negative Matrices and Markov Chains*. Springer Science & Business Media.

Shi, P., Feng, X., and Ivantsova, A. (2015). Dependent frequency–severity modeling of insurance claims. *Insurance: Mathematics and Economics*, 64:417–428.

Shi, P. and Valdez, E. A. (2014). Multivariate negative binomial models for insurance claim counts. *Insurance: Mathematics and Economics*, 55:18–29.

Tan, C. I., Li, J., Li, J. S.-H., and Balasooriya, U. (2015). Optimal relativities and transition rules of a bonus–malus system. *Insurance: Mathematics and Economics*, 61:255–263.

Tauchen, G. (1986). Finite state Markov-chain approximations to univariate and vector autoregressions. *Economics Letters*, 20(2):177–181.

Tauchen, G. and Hussey, R. (1991). Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models. *Econometrica*, 59(2):371–396.

Taylor, G. (1997). Setting a bonus-malus scale in the presence of other rating factors. *ASTIN Bulletin*, 27(2):319–327.

Wen, L., Wu, X., and Zhou, X. (2009). The credibility premiums for models with dependence induced by common effects. *Insurance: Mathematics and Economics*, 44(1):19–25.

Woodard, J. (2014). A conditional distribution approach to assessing the impact of weather sample heterogeneity on yield risk estimation and crop insurance ratemaking. *North American Actuarial Journal*, 18(2):279–93.

Woodard, J. D., Sherrick, B. J., and Schnitkey, G. D. (2011). Actuarial impacts of loss cost ratio ratemaking in us crop insurance programs. *Journal of Agricultural and Resource Economics*, 36(1):211–228.

Yang, Y., Qian, W., and Zou, H. (2018). Insurance premium prediction via gradient tree-boosted tweedie compound poisson models. *Journal of Business & Economic Statistics*, 36(3):456–470.

Yip, K. C. and Yau, K. K. (2005). On modeling claim frequency data in general insurance with extra zeros. *Insurance: Mathematics and Economics*, 36(2):153–163.

Zhang, J., Qiu, C., and Wu, X. (2018). Bayesian ratemaking with common effects modeled by mixture of polya tree processes. *Insurance: Mathematics and Economics*, 82:87–94.

Zhu, W., Tan, K. S., and Porth, L. (2018). Agricultural insurance ratemaking: Development of a new premium principle. *North American Actuarial Journal*, forthcoming.