Dynamic Frailty Count Process in Insurance: A Unified Framework for Estimation, Pricing, and Forecasting

Yang Lu

Dynamic Frailty Count Process in Insurance: A Unified Framework for Estimation, Pricing and Forecasting

Yang Lu (Aix-Marseille University)

September 14, 2016

Abstract

We study count processes in insurance, in which the underlying risk factor is time-varying and unobservable. The factor follows an autoregressive gamma process, and the resulting model generalizes the static Poisson-Gamma model and allows for closed form expression for the posterior Bayes (linear or non-linear) premium. Moreover the estimation and forecasting can be conducted within the same framework in a rather efficient way. An example of automobile insurance pricing illustrates the ability of the model to capture the duration-dependent, non-linear impact of past claims on future ones and the improvement of the Bayes pricing method compared to the linear credibility approach.

Key words: stochastic intensity (dynamic frailty) count process, affine process, non-linear pricing/forecasting, bonus-malus.

1

1 Introduction

We propose a time series model for count variables encountered in Insurance, when the underlying risk factor is time-varying and unobservable. We introduce the autoregressive gamma process for the latent factor dynamics, and show how it provides an integrated framework for the efficient estimation, pricing and forecasting of the risk. One typical application of the model is automobile insurance, for which the insurer holds the history of individual yearly claim counts over several periods. This information should then be taken into account in order to update the future premium regularly, on an individual basis.

Ideally, this latter exercise, also called posterior pricing, should be conducted via the Bayes rule. However, very often, due to the lack of analytic expression of the Bayes premium for most models, the linear credibility method is employed to get an approximation. This practice has some drawbacks. Indeed, the credibility framework is solely based on the first two moments of the count process[1] and does not fully capture the dynamics of the latter, which is clearly non-Gaussian. Thus the linear premium approach induces a systematic pricing bias which has to be evaluated. More importantly, the mean-variance credibility approach is not adapted for a complete risk, which includes non-linear pricing/forecasting, or simulation of the future trajectory of the count process.

In a pioneering work, Dionne and Vanasse (1989) propose a Poisson-Gamma (i.e. negative binomial) model. They show that in a Poisson model with Gamma distributed individual random effect (or frailty, or unobserved heterogeneity), the Bayes premium is linear in past observations and thus can be obtained in closed form. However, one shortcoming of this approach is that all the past observations have the same weight in the premium updating formula, or equivalently, the random effect is time-invariant. There is a growing interest on models that improve this feature [see e.g. Sundt (1988); Brouhns et al. (2003); Purcaru and Denuit (2003); Abdallah et al. (2016); Le Courtois (2015)]. In particular, Pinquet et al. (2001) generalize the negative binomial model by replacing the static frailty with a dynamic one and empirically found evidence of serial correlation of the individual frailty. However, the model suffers from classical computational burdens of models with dynamic unobserved factors, and the only linear credibility premium is available.

Our paper contributes to this literature by proposing a new dynamic frailty model. The static gamma frailty is replaced by a dynamic process with Gamma marginal distribution, called autoregressive gamma process (ARG), which generalizes Dionne and Vanasse (1989)'s model. It belongs to the so-called affine, or compound autoregressive family [see Darolles et al. (2006)]. We show that it provides a tractable, unified approach for estimation, linear or non-linear pricing,

---

[1]Indeed, only Gaussian processes are characterized by its first two moments.

as well as forecasting. More precisely, first, its estimation does not require simulation-based techniques. Second, we show that the *exact* Bayesian forecast formula allows for a closed form expression. Third, this model can be easily extended to higher dimensions, when multiple-risk policies are introduced. We then illustrate the advantage of this new approach via an example of pricing. In particular we show that how the model successfully captures the non-linear, and duration-dependence, and how the credibiilty premium induces a pricing bias with respect to the Bayes premium.

On the contrary to the state-space based, or *parameter-driven* model, the usual count process literature [see e.g. Harvey and Fernandes (1989); Frees and Wang (2005); Gourieroux and Jasiak (2004); Bolancé et al. (2007); Abdallah et al. (2016)] is *observation-driven*, i.e., it specifies directly the dynamics of the observable count variables without serially correlated latent factor. The advantage of parameter-driven models is that it provides a more intuitive interpretation in terms of omitted pricing factors, and can better capture some key features of the observable count data, such as non-linear serial dependence and over-dispersion [see also Davis et al. (2003) for a discussion]. Nevertheless, while state-space models are popular in Finance[2], they have enjoyed a limited success in Insurance. Clearly, one of the reasons is the computational burden involved, since the Bayesian simulation methods typically employed for financial time series are too cumbersome for panel count data[3]. Our paper solves the computational burden by proposing the ARG specification and deriving the exact predictive formulas.

The rest of the paper is organized as follows. Section 2 reviews briefly the negative binomial model and motivates the dynamic frailty framework. Section 3 introduces our model with autoregressive gamma (ARG) based dynamic frailty. We first review the basic properties of this process, and then show how the estimation, Bayes premium updating, as well as non-linear forecasting issues are all rendered easy within the same framework. Section 4 provides a numerical illustration in terms of pricing. Section 5 briefly discusses the prospective bivariate extension of the model. Section 6 concludes. Proofs are gathered in an online Appendix.

2   The static negative binomial model

Consider a population of i.i.d. individuals i, for $i = 1, ..., I$, observed over T periods. Let us denote by $N_{i,1}, N_{i,2}, ..., N_{i,T}$ the yearly number of claims, and a priori risk factors[4] are assumed

---

[2]See for instance the stochastic volatility model for asset prices, and the stochastic intensity model for credit risk occurrences.

[3]One important difference is that in financial applications, the researcher is often interested in a very small number of time series. Thus simulation based methods, albeit time consuming, can still be achieved within a reasonable time. This is not the case in Insurance, where data have a huge cross-sectional dimension accompanied by a (small, yet non negligible) time series dimension. This motivates the need for closed form pricing formulas in insurance applications.

[4]That is, the premium (divided by average cost per claim) of a new customer.

to have the form $(\lambda_{it}) = \exp(\beta' X_{i,t})$, where β is a set of parameters of the model, and $X_{i,t}$ are observable exogenous risk factors. For ease of exposure, in the following, index i is omitted where appropriate. Conditional on a time-invariant, individual-specific random effect U, also called frailty or unobserved heterogeneity, count variables $N_1, N_2, ..., N_T$ are independent, Poisson distributed with respectively parameters $\lambda_t U$, $t = 1, 2, ..., T$. For identification reasons, we assume, without loss of generality, that the mean of the random effect $\mathbb{E}[U]$ is equal to 1.

Dionne and Vanasse (1989) then show that, conditional on the past claim history $\underline{N_T} = N_1, N_2, ..., N_T$, the p.d.f. of U is:

$$f(U \mid \underline{N_T}) \propto f_0(U) U^{\overline{N_T}} e^{-\Lambda_T U},$$

and the predicted number of future claim is:

$$\lambda_{T+1} \mathbb{E}[U \mid N_1, N_2, ..., N_T] = \lambda_{T+1} \frac{\mathbb{E}[U^{\overline{N_T}+1} e^{-\Lambda_T U}]}{\mathbb{E}[U^{\overline{N_T}} e^{-\Lambda_T U}]}. \tag{1}$$

This expression is generically non-linear in $\overline{N_T}$ and has a quite complicated expression[5]. Therefore, it has to be approximated, by, say, a linear function of the past claims, called the credibility approach [see Bühlmann (1967)], except when U follows a gamma distribution with shape parameter k and scale parameter $1/k$, that is: $f(u) \propto u^{k-1} e^{-ku}$. Indeed, Dionne and Vanasse (1989); Lemaire (1995)] show that in this case, we have:

$$\mathbb{E}[U \mid \underline{N_T}] = \frac{k + \overline{N_T}}{k + \overline{\Lambda}}, \tag{2}$$

which is linear in $\overline{N_T}$. This explains the popularity of the negative binomial model.

3 The model with dynamic frailty

One shortcoming of the static frailty assumption is that the conditional mean $\mathbb{E}[U|\underline{N_T}]$ depends on $N_1, N_2, ..., N_T$ only via the sum, hence, past observations have the same weight on the premium, regardless of their respective duration. The literature has proposed evolutionary credibility premium formula that put different weights to claim counts according to their age (or duration) [see e.g. Gerber and Jones (1975); Sundt (1988)]. However, most of these premium formulas are not associated with a full specification of the underlying risk and thus are not appropriate for an integrated risk analysis.

Another alternative is to replace the static frailty U_i with a dynamic one $(U_{it})_t$ [see e.g.

---

[5]See e.g. Jewell (1974); Tremblay (1992); Najafabadi (2010) for a discussion.

Pinquet et al. (2001), Purcaru and Denuit (2003), Englund et al. (2008)], to capture individual *effort*. These processes are i.i.d. across the population but feature serial correlation for the same individual. Moreover $(U_{it})_t$ is exogenous, i.e., the individual unobservable risk factor evolves due to exogenous reasons, that is, the (unconditional) dynamics of (U_t) is independent[6] of the realizations of $N_1, N_2, ..., N_T$. As usual, we assume that given (U_t), count variables (N_t) are independent and Poisson distributed with parameters $\lambda_t U_t$, respectively. Process (U_t) is also called stochastic intensity, or dynamic frailty, and (N_t) is said to have a state-space representation, where (U_t) is the state variable. Let us now complete the model specification by proposing an appropriate exogenous dynamics for the the dynamic frailty process, called (first-order) autoregressive gamma process (ARG).

3.1   The autoregressive gamma process

Let us assume that process (U_t) is Markov, and its dynamics is defined via an intermediate discrete variable[7] Z_t:

- conditional on the past $\underline{U_t}$, variable Z_t follows Poisson distribution with parameter βU_t.
- conditional on $\underline{U_t}$ and $\underline{Z_t}$, variable U_{t+1} follows gamma distribution with shape parameter $\delta + Z_t$ and scale parameter c.

The ARG process has been applied to inter-trade durations [see Gouriéroux and Jasiak (2006)] and to biostatistical count data [see Henderson and Shimakura (2003)]. The specification differs from the autoregressive log-normal process proposed in Insurance [see e.g. Pinquet et al. (2001), Bolancé et al. (2003), Purcaru and Denuit (2003)]. This latter assumes that the log-scaled process $(\log U_t)$ is autoregressive Gaussian, and does not lead to tractable expressions for estimation and Bayes premium updating. The strength of the ARG process is that its dynamics can be easily characterized by the conditional Laplace transform:

Proposition 1. *For any nonnegative argument s, we have:*

$$\mathbb{E}[\exp(-sU_{t+1}) \mid U_t] = \frac{1}{(1+cs)^\delta} \exp\Big(-\frac{\rho s}{1+cs} U_t\Big), \tag{3}$$

where $\rho := \beta c$.

The proof is available in Gouriéroux and Jasiak (2006). For the sake of completeness, it is also included in the online Appendix. We can remark that the conditional Laplace transform

---

[6]However, the conditional dynamics of (U_t) is, by definition, a function of the past claim numbers $N_1, N_2, ..., N_T$.

[7]The intermediate, unobservable Poisson variable Z_t is not equal to the observable count variable. Rather, it is introduced to provide an interpretation of the dynamics of the process, as well as a simple means of simulation.

is an exponential affine function of U_t. Processes possessing this property are called compound autoregressive processes (CaR) [see Darolles et al. (2006)]; they are the discrete time counterparts of the (continuous time) affine processes, which is widely used in Finance and Insurance [see e.g. Duffie et al. (2003); Schrager (2006)]. The simple Laplace transform formula has two advantages. First, we can derive, by taking successive derivatives of the conditional Laplace transform, the conditional (centered, then non-centered) moments of the process. For instance we have:

Proposition 2 (See Gouriéroux and Jasiak (2006)).

$$\mathbb{E}[U_{t+1} \mid U_t] = c\delta + \rho U_t \tag{4}$$

$$\mathbb{V}[U_{t+1} \mid U_t] = c^2\delta + 2\rho c U_t. \tag{5}$$

Proof. See Appendix. □

Equation (4) says the ARG process has a (weak) AR(1) representation, with conditional heteroscedasticity [(5)], which we alternatively rewrite as:

$$\Delta U_t := U_{t+1} - U_t = (1-\rho)\Big(\frac{c\delta}{1-\rho} - U_t\Big) + \sqrt{c^2\delta + 2\rho c U_t}\,\epsilon_t,$$

where ϵ_t is a (weak, and non Gaussian) white noise with unitary variance: $\mathbb{E}[\epsilon_t|U_t] = 0$, $\mathbb{V}[\epsilon_t|U_t] = 1$. This is similar as continuous-time square root, or CIR process. Indeed in Property 5 it will be shown that the ARG arises as the exact time discretization of a CIR process.

Second, equation (3) can be used recursively to obtain the conditional distribution at longer horizons:

Proposition 3 (See Gourieroux and Jasiak (2004), Proposition 1). *For all integer horizon h and time t, we have:*

$$\mathbb{E}[\exp(-sU_{t+h})|U_t] = \frac{1}{(1+c_h s)^\delta} \exp\Big(-\frac{\rho_h s}{1+c_h s}U_t\Big), \tag{6}$$

where:

$$\rho_h = \rho^h, \qquad c_h = c\frac{1-\rho^h}{1-\rho}. \tag{7}$$

Proof. See Appendix. □

Thus the conditional distribution of $U_{t+h}|U_t$ stays within the same parametric family[8] when h varies.

---

[8]This family is called non-centered gamma distribution.

Let us now discuss the strong stationarity of the ARG. The existence (and tractability) of a stationary distribution is vital for the analysis of the financial balance and ruin probabilities of a bonus-malus system [see e.g. Baione et al. (2002)].

Proposition 4 (Stationarity). *1. If $\rho = \beta c < 1$, then process (U_t) is strongly ergodic, and the stationary distribution is gamma with shape parameter δ and scale parameter $\tilde{c} = c/(1-\rho)$.*

2. *If U_0 follows the stationary gamma distribution, then (U_t) is strongly stationary. In particular, for each t, U_t follows the same gamma distribution. From now on let us assume that this two condition is satisfied[9].*

3. *The joint Laplace transform of the pair (U_t, U_{t+1}) is:*

$$\mathcal{L}(s_1, s_2) = \mathbb{E}[e^{-s_1 U_t - s_2 U_{t+1}}] = \frac{1}{(1 + \tilde{c}s_1 + \tilde{c}s_2 + \tilde{c}^2(1-\rho)s_1 s_2)^\delta}, \qquad \forall s_1, s_2 \geq 0, \quad (8)$$

or more generally, for each $h \geq 1$, the pair (U_t, U_{t+h}) has a joint Laplace transform:

$$\mathbb{E}[e^{-s_1 U_t - s_2 U_{t+h}}] = \frac{1}{(1 + \tilde{c}s_1 + \tilde{c}s_2 + \tilde{c}^2(1-\rho_h)s_1 s_2)^\delta}, \qquad \forall s_1, s_2 \geq 0. \quad (9)$$

4. *For all $h \geq 0$, $corr(U_t, U_{t+h}) = \rho^h$.*

5. *(U_t) can be obtained as the time-discretization of a CIR process.*

Proof. See Appendix. $\qquad \square$

Property (1) and (2) explain the terminology of ARG process. From a Bayesian perspective, property (2) can be explained by the causality chain:

$$\cdots \to U_t \to Z_t \to U_{t+1} \to Z_{t+1} \to \cdots$$

In this chain, the stationary distribution of U_t, which is gamma, is conjugate prior to the conditional likelihood function of Z_t, which conditionally follows Poisson distribution. This conjugacy property greatly simplifies the updating formula and explains why the distribution of U_t can remain gamma for all t.

In Property (3), if the correlation coefficient ρ goes to 1, and c goes to zero accordingly such that $\tilde{c} = 1/\delta = \frac{c}{1-\rho}$ remains constant, then $\tilde{c}^2(1-\rho)$ goes to zero and we get a limit Laplace transform: $\mathcal{L}_{\text{limit, } 1}(s_1, s_2) = \frac{1}{(1+s_1/\delta+s_2/\delta)^\delta}$, which corresponds to the singular bivariate distribution (U, U). Therefore the ARG process allows for the static gamma process $U_t = U$

---

[9]Such a choice is rather popular in the econometric literature [see e.g. Heckman (1987) for discussions].

as a special case. On the other hand, when ρ goes to 0, we get another limit Laplace transform: $\mathcal{L}_{\text{limit, 2}}(s_1, s_2) = \frac{1}{(1+s_1/\delta+s_2/\delta+s_1 s_2/\delta^2)^\delta}$, which corresponds to two independent gamma components.

Property (4) implies that U_t and U_{t+h} are nonnegatively correlated, which reflects the *habit inertia* of the policy-holder. Moreover, the autocorrelation function decreases in a geometric rate, similarly as stationary ARMA models.

We close this subsection by giving the joint Laplace transform of the finite dimensional distribution of the process.

Proposition 5. *Denote by* $R = (r_{i,j})_{1 \leq i,j \leq T} = (\rho^{|i-j|/2})_{1 \leq i,j \leq T}$ *the matrix of square-rooted correlation coefficients of* $(U_1, U_2, ..., U_T)$, *then the joint Laplace transform is:*

$$\mathbb{E}[\exp(-s_1 U_1 - s_2 U_2 - ... - s_T U_t)] = \frac{1}{\left\{ \det \left[I + R/\delta \ Diag(s_1, s_2, ..., s_T) \right] \right\}^\delta}. \qquad (10)$$

Proof. See Sim (1990). $\qquad\qquad\square$

In particular when $T = 2$, we get

$$\det[I + \tilde{c}R \ \text{Diag}(s_1, s_2, ..., s_T)] = \begin{vmatrix} 1 + s_1/\delta & \sqrt{\rho}s_2/\delta \\ \sqrt{\rho}s_1/\delta & 1 + s_2/\delta \end{vmatrix} = 1 + s_1/\delta + s_2/\delta + (1-\rho)s_1 s_2/\delta^2,$$

which, up to a power transformation, coincides with the RHS of equation (3).

For insurance applications, the number of claims is usually small(generically smaller than 4, for a panel of up to 10 years), thus these lower-order cross partial derivatives are easily computable using a symbolic computation language such as *Mathematica*.

Finally, let us note that we can multiply U_t any positive constant and dividing λ_t by the same constant, without changing the dynamics of the observable variable N_t. Thus for identification reasons, we assume, from now on, $\mathbb{E}[U_1] = 1$, that is, $\frac{c}{1-\rho}\delta = 1$.

3.2 Statistical inference

The model can be estimated using a simple two-stage approach [see e.g. Zeger (1988)]. In the first stage, parameters α and δ are estimated without taking into account the serial dependence. In other words, one estimates the mis-specified i.i.d. negative binomial model:

$$\ell_2(\alpha, \delta) = \sum_{i=1}^{I} \sum_{t=1}^{T} \log \mathbb{P}[N_{i,t} = n_{i,t}] \propto \sum_{i=1}^{I} \sum_{t=1}^{T} \log \mathbb{E}\left[U_{i,t}^{n_{i,t}} \exp(-e^{\alpha' X_{i,t}} U_{i,t}) \right], \qquad (11)$$

Such an estimator can be obtained with a standard statistical software, such as the "glm.nb" function in the R language. In the second step, the autocorrelation parameter $\rho = \beta c$ is estimated by matching the empirical counterpart of the autocorrelation functions of (N_t) with its theoretical values.

Finally, the model can be also estimated by other two competing methods, that are maximum likelihood estimation and composite likelihood estimation. These two methods are more efficient than the method we describe above, but are also more computationally intensive. In order to save space, the details concerning these two methods are left in Section 1 of the online Appendix.

3.3 Bayesian pure premium updating and forecasting

Given the observation of $N_1, N_2, ..., N_T$, let us provide the conditional mean $\lambda_{T+1}\mathbb{E}[U_{T+1} \mid \underline{N_T}]$, that is the new (actuarial) premium at period $T + 1$, as well as the the conditional variance of the future claim :

$$v_{T+1}(\underline{N_T}) = \mathbb{V}[N_{T+1} \mid \underline{N_T}] = \lambda_{T+1}\mathbb{E}[U_{T+1} \mid \underline{N_T}] + \lambda_{T+1}^2\mathbb{V}[U_{T+1} \mid \underline{N_T}]. \tag{12}$$

We have:

$$\mathbb{E}[U_{T+1} \mid \underline{N_T}] = \mathbb{E}\Big[\mathbb{E}[U_{T+1} \mid U_T] \mid \underline{N_T}\Big] = c\delta + \rho\mathbb{E}[U_T \mid \underline{N_T}] \tag{13}$$

$$= c\delta + \rho\frac{\mathbb{E}[U_T \prod_{t=1}^{T} U_t^{N_t} e^{-\lambda_t N_t}]}{\mathbb{E}[\prod_{t=1}^{T} U_t^{N_t} e^{-\lambda_t N_t}]} = c\delta - \rho\frac{\frac{\partial^{\overline{N}_T+1}}{\partial_1^{n_1}\partial_2^{n_2}\cdots\partial_T^{n_T+1}}\mathbb{E}[\exp(-\lambda_1 U_1 - \lambda_2 U_2 - \cdots - \lambda_T U_T)]}{\frac{\partial^{\overline{N}_T}}{\partial_1^{n_1}\partial_2^{n_2}\cdots\partial_T^{n_T}}\mathbb{E}[\exp(-\lambda_1 U_1 - \lambda_2 U_2 - \cdots - \lambda_T U_T)]},$$
$$\tag{14}$$

which can be computed[10] by Proposition 5. As stated by Dionne and Vanasse (1989), the ability to compute Bayes premium, compared to the credibility premium, is that first, it is fair for policyholders since it takes into account individual risk characteristics, as well as individual claim history. Second, such a premium principle is automatically financially balanced.

Similarly, we have:

$$\mathbb{E}[U_{T+1}^2 \mid \underline{N_T}] = \mathbb{E}\Big[\mathbb{E}[U_{T+1}^2|U_T] \mid \underline{N_T}\Big] = \mathbb{E}\Big[\mathbb{E}[U_{T+1}|U_T]^2 + \mathbb{V}[U_{T+1}|U_T] \mid \underline{N_T}\Big]$$
$$= \mathbb{E}\Big[(c\delta + \rho U_T)^2 + c^2\delta + 2\rho c U_T \mid \underline{N_T}\Big]$$
$$= c^2\delta(1 + \delta) + \rho^2\mathbb{E}[U_T^2|\underline{N_T}] + 2\rho c(1 + \delta)\mathbb{E}[U_T|\underline{N_T}],$$

---

[10]The RHS of the equation (14) does not have an analytical expression in $N_1, ..., N_T$, however for given values of (λ_t) and (N_t), its value can be easily computed without approximation by *Mathematica*.

thus

$$\mathbb{V}[U_{T+1} \mid \underline{N_T}] = c^2\delta + \rho^2(\mathbb{E}[U_T^2|\underline{N_T}] - \mathbb{E}[U_T|\underline{N_T}]^2) + 2\rho c\mathbb{E}[U_T|\underline{N_T}].$$

where $\mathbb{E}[U_T^2|\underline{N_T}] = \dfrac{\frac{\partial^{\overline{N}_T+2}}{\partial_1^{n_1}\partial_2^{n_2}\ldots\partial_T^{n_T+2}}\mathbb{E}[\exp(-\lambda_1 U_1 - \lambda_2 U_2 - \cdots - \lambda_T U_T)]}{\frac{\partial^{\overline{N}_T}}{\partial_1^{n_1}\partial_2^{n_2}\ldots\partial_T^{n_T}}\mathbb{E}[\exp(-\lambda_1 U_1 - \lambda_2 U_2 - \cdots - \lambda_T U_T)]}$.

Remark 1. When T is very large, computing the Laplace transform, as well as all its partial derivatives would become computationally intensive[11]. However, in this case, we can use the fact that the process (N_t) has short memory to approximate the Bayes updating formula by:

$$\mathbb{E}[U_{T+1} \mid \underline{N_T}] \approx \mathbb{E}[U_{T+1} \mid N_{T-l}, N_{T-l+1}, ..., N_T], \tag{15}$$

where the integer l denotes the number of lagged values retained for the approximation.

Finally, the prediction can also be conducted at any horizon $h > 1$. This analysis is provided in Section 7 of the online Appendix.

Illustration. For $T = 1$, by using $\mathbb{E}[e^{-tU_1}] = \frac{1}{(1+\frac{c}{1-\rho}t)^\delta}$ we can obtain:

$$\mathbb{E}[U_2|N_1] = c\delta + \rho\frac{\mathbb{E}[U_1^{N_1+1}e^{-\lambda U_1}]}{\mathbb{E}[U_1^{N_1}e^{-\lambda U_1}]} = c\delta + \rho\frac{\delta + N_1}{(1 + \frac{c}{1-\rho}\lambda)}\frac{c}{1-\rho}, \tag{16}$$

$$\mathbb{V}[U_2|N_1] = c^2\delta + c^2\frac{\rho(2-\rho)(\delta + N_1)}{(1 + \frac{c}{1-\rho}\lambda)^2(1-\rho)^2}. \tag{17}$$

Thus the conditional variance is also increasing in N_1. In order to understand its implication, let us assume that the insurer uses the mean-variance[12], instead of pure premium pricing method. In this case, riskier policyholders are charged more because of both the higher expected cost, and the higher uncertainty.

As a comparison, the traditional credibility based forecast is the best linear predictor of the future claim, that is: $\mathbb{EL}[N_{T+1}|\underline{N_T}]$. By the definition of the Bayes forecast[13], the credibility forecast is associated with a larger mean square error:

$$\mathbb{E}\Big[(N_{T+1} - \mathbb{EL}[N_{T+1}|\underline{N_T}])^2|\underline{N_T}\Big] > \mathbb{E}\Big[(N_{T+1} - \mathbb{E}[N_{T+1}|\underline{N_T}])^2|\underline{N_T}\Big] = \mathbb{V}[N_{T+1}|\underline{N_T}].$$

Moreover, the credibility approach does not account for the heteroscedasticity of the count process

---

[11]Note, however, that a similar computational difficult exists for the credibility model, which involves the inversion of a $(T + 1) \times (T + 1)$ dimensional matrix [see Albrecht (1985) for details].

[12]That is: premium=expected cost+ a risk loading that is a multiple of the variance. A non-linear, Esscher-transform based premium principle will be discussed later.

[13]That is the best forecast in terms of mean square error.

[see (12)]. In other words, it is assumed that the conditional variance of the forecast error $N_{T+1} - \mathbb{EL}[N_{T+1}|\underline{N_T}]$ does not depend on $\underline{N_T}$. As a consequence, when an insurer uses the mean-variance pricing principle along with the credibility forecast, low-risk individuals are unfairly charged the same risk loading as high-risk individuals.

3.4 Non-linear pricing

The previous subsection has derived the actuarial posterior premium. In practice, however, the insurer needs to charge more than the actuarial premium, due to the risk aversion, the cost of the capital, and the profit, etc. In a milestone paper, Bühlmann (1984) provided the economic foundation for the Esscher transform as a general premium principle. Let us now show that our dynamic specification of the frailty leads to closed form formula for such non-linear, posterior premium. This possibility is due to the simple form of the Laplace transform of a Poisson distribution, as well as that of the finite dimensional distribution of the ARG process.

For a positive risk-aversion parameter α, the Esscher transform of the conditional distribution of N_{T+1} is defined as:

$$p_{2,T+1}(\underline{N_T}) := \frac{\mathbb{E}[N_{T+1}e^{\alpha N_{T+1}} \mid \underline{N_T}]}{\mathbb{E}[e^{\alpha N_{T+1}} \mid \underline{N_T}]} = \frac{\partial}{\partial \alpha} \log \mathbb{E}[e^{\alpha N_{T+1}} \mid \underline{N_T}]. \tag{18}$$

Therefore it suffices to compute $\mathbb{E}[e^{\alpha N_{T+1}} \mid \underline{N_T}]$, which is equal to:

$$\mathbb{E}[e^{\alpha N_{T+1}} \mid \underline{N_T}] = \mathbb{E}\Big(\mathbb{E}[e^{\alpha N_{T+1}}|U_{T+1}, \underline{N_T}] \mid \underline{N_T} \Big)$$
$$= \mathbb{E}[e^{-\lambda_{T+1}(1-e^\alpha)U_{T+1}} \mid \underline{N_T}] = \frac{\mathbb{E}[e^{-\lambda_{T+1}(1-e^\alpha)U_{T+1}} \prod_{t=1}^{T} U_t^{N_t} e^{-\lambda_t U_t}]}{\mathbb{E}[\prod_{t=1}^{T} U_t^{N_t} e^{-\lambda_t U_t}]},$$

which can be obtained explicitly by Proposition 5.

4 Pricing implications

Let us now demonstrate the advantage of our dynamic frailty model, as well as the associated Bayes premium computation method, via an example of pricing. We compute the a posteriori premium for individuals with different claim history and compare it to the linear credibility premium, as well as the (Bayes) premium for the nested, static negative binomial model. Let us first remind the computation methodology for these three types of premium. For ease of exposition, we assume, without loss of generality, that $(\lambda_t) = (\lambda)$ is constant. Time-varying λ_t can be easily accommodated without increasing the complexity of calculation.

Bayes premium for the ARG model. For a given T and a sequence of claim history, $(N_t)_{1 \leq t \leq T}$, we compute the Laplace transform, then its partial derivatives of order $(N_1, N_2, ..., N_T)$, then of order $(N_1, N_2, ..., N_{T-1}, N_T + 1)$ by Mathematica[14].

To set the parameters of the model, we remark that the model can be estimated using the estimation equation approach proposed by Bolancé et al. (2003), thus we take the parameter estimates from their paper. We have: $\frac{c}{1-\rho}\delta = \mathbb{E}[U_t] = 1$ for normalization, and we set $\left(\frac{c}{1-\rho}\right)^2 \delta = \mathbb{V}[U_t] = 1.366$ [see equation (22) in Bolancé et al. (2003)], and $\rho = \text{corr}[U_t, U_{t+1}] = 0.73$ according to their Table[15] 1. Then we take $\lambda_t = \lambda = 0.07$ as their paper. In other words, the premium (assuming, for expository purpose, a unitary cost per claim) of a new customer is $\lambda = 0.07$.

Static negative binomial model. For the static model, the premium formula is given by (2). As this model is a special case of the ARG model, we can use the same parameter δ as before, and choose correlation parameter $\rho = 0$.

Credibility premium. The premium for the ARG model is obtained [see e.g. Albrecht (1985); Pinquet et al. (2001) for details] by fitting the linear model:

$$N_{T+1} = \underbrace{a + \sum_{t=1}^{T} b_{t,T} N_t}_{\text{Credibility premium}} + \epsilon_{T+1},$$

where a and $b_{t,T}, t = 1, ..., T$ are regression coefficients, and the residual ϵ_{T+1} is uncorrelated with the regressors $1, N_1, ..., N_T$: $\mathbb{E}[\epsilon_T] = \mathbb{E}[\epsilon_T N_1] = \cdots = \mathbb{E}[\epsilon_T N_T] = 0$.

Case $T = 1$. Table 1 illustrates the posterior premium at period 2 according to the number of claims (which we limit to 2) in the first period. For ease of comparison, each premium is expressed as a multiple of λ. Nevertheless, the reader should keep in mind that these multiplicative coefficients depend also on the value of λ for both models and both premium principle.

	ARG, Bayes	Negative Binomial	ARG, Credibility
$p_2(N_1 = 0)$	$0.93\ \lambda$	$0.91\ \lambda$	$0.93\ \lambda$
$p_2(N_1 = 1)$	$1.84\ \lambda$	$2.16\ \lambda$	$1.84\ \lambda$
$p_2(N_1 = 2)$	$2.75\ \lambda$	$3.48\ \lambda$	$2.75\ \lambda$

Table 1: Premium at the second period, according to the number of claims in the first period. The second column reports the Bayes premium obtained with the ARG model, the third column reports the Bayes premium obtained with the negative binomial model, and the last one reports the credibility premium of the ARG model.

---

[14] A copy of the programme is available from the author upon request.

[15] In a previous version of the paper available online, we also explain how to fit a dynamic frailty model with a more flexible autocorrelation function than the first order ARG process.

As expected, one or more claim results in an increase of the premium in each of the three columns. However, the size of the increase differs. For instance, if there is one (resp. two) claim at the first year, then the premium of the next year is given by (16) and is 1.84 λ (resp. 2.75λ), that is an 84 % (resp. 175%) increase. As a comparison, in the classical negative binomial model, the premium would have been: $\frac{0.733+1}{0.733+0.07}\lambda = 2.16\lambda$ ($\frac{0.733+2}{0.733+0.07}\lambda = 3.48\lambda$) that is more than 100 % (resp. 250 %) more expensive. This is due to the mean-reverting property of the dynamic frailty.

On the contrary, if there is no claim in the first year, the next year premium is 0.93λ, compared to 0.91λ in the negative binomial model. As a summary, our model provides a less severe post-claim malus, as well as a less generous non-claim bonus than that induced by the negative binomial model [a similar result is obtained for the linear credibility premium, see also Pinquet et al. (2001)].

Finally, let us remark that in Table 1, the credibility and Bayes premia are equal. This is expected, but is only true in the special case where $T = 1$. Indeed, equation (16) shows that the conditional expectation $\mathbb{E}[U_2|N_1 = n_1]$ is a linear function of n_1.

Case $T = 2$. We now move to the posterior premium at period 3, according to the claim history of the two first periods. In particular, we will see that from now on, the credibility premium is generically biased with respect to the Bayes premium and the bias is path-dependent. For illustration purpose, in Table 2, we restrict ourselves to no more than 3 claims during the first two periods.

	ARG, Bayes	Negative Binomial	ARG, Credibility
$p_3(0,0)$	0.89 λ	0.84 λ	0.90 λ
$p_3(0,1)$	1.75 λ	1.99 λ	1.76 λ
$p_3(1,0)$	1.49 λ	1.99 λ	1.50 λ
$p_3(1,1)$	2.47 λ	3.13 λ	2.37 λ
$p_3(0,2)$	2.60 λ	3.13 λ	2.63 λ
$p_3(2,0)$	2.08 λ	3.13 λ	2.11 λ
$p_3(3,0)$	2.67 λ	4.27 λ	2.72 λ
$p_3(2,1)$	3.10 λ	4.27 λ	2.98 λ
$p_3(1,2)$	3.36 λ	4.27 λ	3.24 λ
$p_3(0,3)$	3.46 λ	4.27 λ	3.50 λ

Table 2: Premium for the third period, according to the number of claims in the first and second periods. For expository purpose we use the abbreviation $p_3(1,0) = p_3(N_1 = 1, N_2 = 0)$, say.

We can see that the pricing error of the credibility approach with respect to the Bayes approach is large, so long as both N_1 and N_2 are non zero. In order to better understand the dependence of p_3 in the two variables N_1 and N_2, we introduce the first and second order

difference operators[16]:

$$\Delta_x p_3(x, y) := p_3(x + 1, y) - p_3(x, y), \quad \Delta_y p_3(x, y) := p_3(x, y + 1) - p_3(x, y),$$

$$\Delta_{xx}^2 p_3(x, y) := \Delta_x(\Delta_x p_3(x, y)) = p_3(x + 2, y) + p_3(x, y) - 2p_3(x + 1, y),$$

$$\Delta_{xy}^2 p_3(x, y) := \Delta_x(\Delta_y p_3(x, y)) = p_3(x + 1, y + 1) + p_3(x + 1, y) - p_3(x, y + 1) + p_3(x, y),$$

$$\Delta_{yy}^2 p_3(x, y) := \Delta_y(\Delta_y p_3(x, y)) = p_3(x, y + 2) + p_3(x, y) - 2p_3(x, y + 1), \quad \forall x, y \in \mathbb{N}.$$

In particular, $\Delta_y p_3(x, 0)$ is the bonus-malus factor of a non-claim second period. In order to save space, the value of these finite differences are reported in Tables 1-5 of the online Appendix. From Tables 1 and 2 we can remark that, for both the negative binomial premium and the credibility premium, $\Delta_x p_3$ and $\Delta_y p_3$ are constant and therefore $\Delta_{xx}^2 p_3 = \Delta_{xy}^2 p_3 = \Delta_{yy}^2 p_3 = 0$. This is expected, since under these two pricing principles, p_3 is a linear function of the two arguments x, y.

Then we can remark that, for both the credibility and Bayes ARG model, we have always $\Delta_x p_3 < \Delta_y p_3$, in other words, the more recent claim has a larger impact on the posterior premium. This is in contrast with the negative binomial model in which p_3 depends only on the total number of claims counts, and shows the interest of having a time-varying heterogeneity term.

Let us now discuss the second-order differences for the Bayes ARG model. We can see that $\Delta_{xx}^2 p_3$ and $\Delta_{yy}^2 p_3$ are generically different from zero, but are approximately zero when $x = y = 0$, whereas $\Delta_{xy}^2 p_3$ is significantly positive, for all $x + y \leq 1$. For instance, we have:

$$\Delta_{xy}^2 p_3(0, 0) = 0.12 = \Big(p_3(1, 1) - p_3(0, 0) \Big) - \Big(p_3(1, 0) - p_3(0, 0) \Big) - \Big(p_3(0, 1) - p_3(0, 0) \Big),$$

in other words the malus of having accidents at both the first and second period, that is $p_3(1, 1) - p_3(0, 0)$, is larger than the sum of the malus of having just one accident during the first and second period, respectively. Equivalently, in terms of the bonus-malus factor, we have

$$p_3(1, 1) - p_3(1, 0) = 0.72\lambda > p_3(0, 1) - p_3(0, 0) = 0.60\lambda, \tag{19}$$

which depends on the past claim history, and is increasing in the other variable, compared to the credibility premium formula which assumes it to be independent of x and y. This can be interpreted as the heteroscedasticity of the process (U_t) (and hence of process (N_t)). More precisely, the conditional expectation $\mathbb{E}[U_1 | N_1]$ is larger if an individual has a claim at period $t = 1$. Thus we also expect that when the information of N_2 is incorporated, the updated

---

[16]We thank an anonymous referee for this suggestion.

expectation $\mathbb{E}[U_2|N_1, N_2]$, has a larger sensitivity in N_2, if N_1 is equal to one.

This finding illustrates the bias, if the usual linear approximation in the credibility approach is to be used:

$$p_3(x, y) = p_3(0, 0) + \Delta_x p_3(0, 0)x + \Delta_y p_3(0, 0)y, \qquad \forall x, y \in \mathbb{N}, \tag{20}$$

Indeed when x, y are small but both are non zero, a *convexity adjustment* is needed in the previous expansion:

$$p_3(x, y) \approx p_3(0, 0) + \Delta_x p_3(0, 0)x + \Delta_y p_3(0, 0)y + \Delta_{xy}^2 p_3(0, 0)\frac{xy}{2}. \tag{21}$$

Of course, the approximation (21) becomes less precise when at least one of x, y is large, since in that case the higher order terms $\Delta_{xx}^2 p_3(0, 0)\frac{x^2}{2}$ and $\Delta_{yy}^2 p_3(0, 0)\frac{y^2}{2}$, or even third-order cross terms $\Delta_{xxy}^3 p_3(0, 0)\frac{x^2 y}{6}$, $\Delta_{xyy}^3 p_3(0, 0)\frac{xy^2}{6}$ become non negligible.

This analysis is also linked to a recent work of Le Courtois (2015), who proposes to improve the linear credibility formula by including quadratic terms x^2, y^2 into the regression (20). Indeed, (21) shows that at least in the ARG dynamic frailty model, the cross terms are more important than the squared ones. To some extent, this could be interpreted as the following. The squared terms are mainly intended to capture the non-Gaussian feature of the *marginal* distribution of each N_t, whereas the cross terms capture the non-linear, that is heteroscedastic feature of the *serial dependence*.

Finally, is $\Delta_{xy}^2 p_3(x, y)$ always positive for all x, y, and/or for all values of λ_1, λ_2? The answer is negative. In Figure 1 of the online Appendix, we plot the iso-contour of $\Delta_{xy}^2 p_3(0, 0)$ as a function of λ_1 and λ_2, which we allow, momentarily, to be different[17]. We find that, for instance, when $\lambda_1 = 0.07$ and $\lambda_2 = 0.2$, that is when there is a sudden increase the value of the a priori premium λ in the second period, we have $\Delta_{xy}^2 p_3(0, 0) = -0.03 < 0$. However, when λ_1, λ_2 are close to each other, the sign of $\Delta_{xy}^2 p_3(0, 0)$ is always positive. Moreover, the value of $\Delta_{xy}^2 p_3(x, y)$ also depends on parameters of the ARG, that are (ρ, δ). In Figure 2 of the online Appendix, we plot the iso-contour of $\Delta_{xy}^2 p_3(0, 0)$ as a function of ρ and δ, when λ_1, λ_2 are fixed to be 0.07. We can see in particular that when δ is small, the value of $\Delta_{xy}^2 p_3(0, 0)$ is very large, and thus the convexity adjustment term in equation (21) is very important.

Case $T = 3$. Let us now move to the posterior premium for period $T + 1 = 4$, based on the claim history of the three previous periods. Table 3 reports individuals with at most one claim

---

[17]In this exercise, we have kept the same values of δ and ρ. Of course, the numerical results would be different when different values of δ and ρ are used.

during each period[18].

	ARG, Bayes	Negative Binomial	ARG, Credibility
$p_4(0,0,0)$	0.87 λ	0.77 λ	0.87 λ
$p_4(0,0,1)$	1.69 λ	1.84 λ	1.72 λ
$p_4(0,1,0)$	1.43 λ	1.84 λ	1.45 λ
$p_4(0,1,1)$	2.38 λ	2.90 λ	2.30 λ
$p_4(1,0,0)$	1.25 λ	1.84 λ	1.26 λ
$p_4(1,0,1)$	2.25 λ	2.90 λ	2.11 λ
$p_4(1,1,0)$	1.90 λ	2.90 λ	1.85 λ
$p_4(1,1,1)$	2.91 λ	3.97 λ	2.69 λ

Table 3: Premium for the fourth period, according to the numbers of claims in the previous periods.

Similarly, for the Bayes ARG model, we can compute the first-order differences Δ_x, Δ_y, Δ_z of the function $p_4(x,y,z)$, as well as the second-order differences Δ^2_{xx}, Δ^2_{yy}, Δ^2_{zz}, Δ^2_{xy}, Δ^2_{yz}, Δ^2_{zx}. Their values are reported in Tables 6-9 of the online Appendix. As expected, we obtain similar findings as in the case $T = 2$, namely:

- $\Delta_x p_4 < \Delta_y p_4 < \Delta_z p_4$, that is the importance of the observations in the premium function decreases as the duration increases.
- At point $(x, y, z) = (0, 0, 0)$, we have approximately $\Delta^2_{xx} p_4 = \Delta^2_{yy} p_4 = \Delta^2_{zz} p_4 = 0$, whereas the cross-differences $\Delta^2_{xy} p_4$, $\Delta^2_{yz} p_4$, $\Delta^2_{zx} p_4$ are significantly positive. This explains why the pricing bias of the credibility approach is generically larger whenever there are more than two claims during the previous periods. For instance, if there are one claim during each period, then the Bayes premium is 2.91 λ, compared to the credibility premium 2.69 λ.

Evolution of the premium. Let us now illustrate the path-dependency of the Bayes premium. In Table 4 we report the evolution of the Bayes premium an individual obtains at periods $t = 2, 3, ..., 7$, before reporting its second claim. We denote by τ the timing of the first claim and distinguish seven cases: $\tau = 1, 2, ..., 6$, or $\tau > 6$, corresponding the case of no claim in the first six years.

---

[18]A more detailed report is available in Table 6 of the online Appendix.

	$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$	$\tau = 5$	$\tau = 6$	$\tau > 6$
$t = 2$	1.84	0.93	0.93	0.93	0.93	0.93	0.93
$t = 3$	1.49	1.75	0.89	0.89	0.89	0.89	0.89
$t = 4$	1.25	1.43	1.69	0.87	0.87	0.87	0.87
$t = 5$	1.10	1.22	1.39	1.66	0.86	0.86	0.86
$t = 6$	1.01	1.08	1.20	1.37	1.64	0.84	0.84
$t = 7$	0.94	0.94	1.07	1.18	1.36	1.62	0.84

Table 4: Evolution of the Bayes premium (divided by λ) for an individual before reporting the second claim.

For instance, column $\tau = 1$ reports the sequence of at periods $2 - 7$, if this individual has a first claim at period $\tau = 1$, but does not report any claim in the subsequent periods. In other words, the six entries of this column are respectively: $p_2(1)$, $p_3(1, 0)$, $p_4(1, 0, 0)$, $p_5(1, 0, 0, 0)$, $p_6(1, 0, 0, 0, 0)$, and $p_7(1, 0, 0, 0, 0, 0)$, divided by λ. On the other hand, column $\tau > 6$ reports the premium for an individual who has never reported claims, that is, the entries of that column are respectively: $p_2(0)$, $p_3(0, 0)$, $p_4(0, 0, 0)$, $p_5(0, 0, 0, 0)$, $p_6(0, 0, 0, 0, 0)$, and $p_7(0, 0, 0, 0, 0, 0)$.

For each column, the sequence of premium is smaller than 1 and decreasing so long as there is no claim yet. In particular, in the last column, the premium of an individual who has never reported claims decreases and converges to a constant (0.84). This is the best "bonus premium" one can obtain. As a comparison, for the negative binomial model, the premium would have decreased towards zero if the individual never reports claim [see equation (2)]. However, this convergence rate is hyperbolic, thus much slower than in the dynamic frailty model. For instance, after six non-claim periods, the premium at $t = 7$ obtained by the negative binomial model is 0.63. This is because our model assigns more weights to recent claims than the static model.

After the first claim, the premium jumps back to above 1. This new "malus premium" depends on the timing of the first claim and ranges from 1.84 for $\tau = 1$ to 1.62 to $\tau = 6$. The larger τ, that is the larger the number of consecutive non-claim periods, the lower this first "malus premium". In particular, for an individual who previously benefits the best premium 0.84, the malus premium is 1.62 (which corresponds to the entry $t = 7, \tau = 6$.)

From the column $\tau = 1$, we observe that after the first claim at $t = 1$, it takes the individual 4 consecutive periods of non-claim in order to get the premium 1.01 at the period $t = 6$, which is approximately the initial, a priori premium. This waiting time is also decreasing in τ. For instance, column $\tau = 2$ says that for an individual who has one non-claim period before the first claim, 4 consecutive non-claim periods after the first claim leads to a premium equal to 0.94.

5   A simple common factor bivariate extension

There is a growing interest in the literature to extend the standard univariate analysis to multiple count processes [see e.g. Frees (2003); Pinquet (1998); Englund et al. (2008, 2009); Abdallah et al. (2016); Bermúdez and Karlis (2015)]. For instance, a car insurance policy involves in general two types of coverage: the third-party liability and the (body and vehicle) casualty. This gives rise to a bivariate count process $(N_{1,t}, N_{2,t})_t$, which generically features positive correlation between the two components, as well as serial correlation between different periods. The aim of this section is to briefly discuss a bivariate extension of our model. Let us assume that i), the unobservable factor (U_t) is, stationary ARG. ii), conditional on the entire trajectory of (U_t), pairs $(N_{1,t}, N_{2,t})_t$ are independent , and have the following stochastic representation:

$$N_{1,t} = Z_{0,t} + Z_{1,t}, \qquad N_{2,t} = Z_{0,t} + Z_{2,t}, \tag{22}$$

where $Z_{0,t}, Z_{1,t}, Z_{2,t}$ are mutually independent, Poisson distributed conditionally on (U_t), with parameters $\lambda_{0,t}U_t, \lambda_{1,t}U_t$ and $\lambda_{2,t}U_t$ respectively, and $\lambda_{0,t}, \lambda_{1,t}, \lambda_{2,t}$ are observable up to a finite number of parameters α.

This decomposition states that the occurrence of claims can be either due to third-party liability alone, or due to casualty alone, or due to both, although for each claim, this information is not necessarily available to the actuary[19]. This is usually called trivariate Poisson reduction [see Gourieroux et al. (1984), Bermúdez and Karlis (2012)].

Thus the unobservable common factor process (U_t) introduces positive dependence between $N_{1,t}$ and $N_{2,t}$, as well as serial correlation. As a consequence, the past observations of both count variables have prediction power.

Let us also remark that $N_{1,t}$ (resp. $N_{2,t}$) are conditionally Poisson distributed with parameter $(\lambda_{0,t} + \lambda_{1,t})U_t$ (resp. $(\lambda_{0,t} + \lambda_{2,t})U_t$). Thus the marginal distributions of $N_{1,t}$ and $N_{2,t}$ are conditionally Poisson (and marginally negative binomial), with a dynamic latent factor that is a autoregressive gamma process. In other words, in this bivariate model, each component count process is compatible with the univariate model described above. This compatibility is quite appealing, since usually the actuary has a portfolio of policyholders in which some choose two (or more) types of coverage, whereas others choose only one type. Thus a bivariate model that is not compatible with the marginal model is difficult to use on such a mixed portfolio.

---

[19]Indeed, if this information is available, then we have a trivariate count variable.

6 Concluding remarks

This paper proposed a dynamic frailty model for count processes in Insurance, and extends the previous work by Pinquet et al. (2001); Bolancé et al. (2003); Brouhns et al. (2003). We showed how the estimation, premium updating as well as non-linear risk forecasting are greatly simplified within the same framework. In particular, in terms of pricing, we derived closed form formulas for the *exact* Bayes posterior premium. Using a numerical example on automobile insurance pricing, we illustrated how the posterior premium captures the non-linear, duration-dependent impact of past claim history on future ones, and evaluated the potential pricing bias, if a credibility approach were to be used instead. More precisely, the (linear) credibility approach mis-prices future risks by failing to account for a convexity adjustment, that is the surcharge of having accidents at several different periods, compared to the sum of marginal malus coefficients of individual accidents. As a by-product, our model suggests an improved quadratic credibility formula, when the modeller wants to use a specification of the underlying dynamic frailty that is different from ARG (and thus when the closed-form Bayes premium is not available):

$$p_{T+1}(N_1, N_2, ..., N_T) = \sum_{t=1}^{T} a_t N_t + \sum_{1 \leq t_1 < t_2 \leq T} b_{t_1, t_2} N_{t_1} N_{t_2}.$$

This formula suggests several avenues for future research. First, is it possible to ensure the positivity of the regression coefficients? Second, when T is large, this formula involves the inversion of a high-dimensional (indeed $\frac{T(T+1)}{2} \times \frac{T(T+1)}{2}$) matrix. Thus is it possible to impose some sparsity constraints on the matrix $(b_{i,j})$, say, $b_{i,j} = 0$ so long as $|i - j|$ is large enough?

Our key technical contribution is the introduction of the ARG dynamic frailty model. Indeed, besides the tractability it provides for estimation, pricing and forecasting in our specific application, this process has also the great advantage of belonging to the family of (discrete/continuous time) affine processes, which are very popular in Finance. Thus this framework is also quite appealing for other insurance applications which involves financial risks, thanks to the compatibility it can provide. These situations have become quite standard nowadays due to the interconnectedness between the two sectors. One typical example concerns the modelling of life-insurance lapses. In this case we can imagine an unobservable common risk factor which drives the process of monthly total number of lapses of an insurer. This avoids the use of *ad hoc* observable risk factors and thus provides more flexibility for risk analysis.

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*.

Albrecht, P. (1985). An evolutionary credibility model for claim numbers. *ASTIN Bulletin*, 15(01):1–17.

Baione, F., Levantesi, S., and Menzietti, M. (2002). The development of an optimal bonus-malus system in a competitive market. *Astin Bulletin*, 32(01):159–170.

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

Bermúdez, L. and Karlis, D. (2015). A posteriori ratemaking using bivariate poisson models. *Scandinavian Actuarial Journal*, pages 1–11.

Bolancé, C., Denuit, M., Guillén, M., and Lambert, P. (2007). Greatest accuracy credibility with dynamic heterogeneity: the harvey-fernandes model. *Belgian Actuarial Bulletin*, 7(1):14–18.

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.

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.

Bühlmann, H. (1967). Experience rating and credibility. *Astin Bulletin*, 4(03):199–207.

Bühlmann, H. (1984). The general economic premium principle. *Astin Bulletin*, 14(01):13–21.

Darolles, S., Gourieroux, C., and Jasiak, J. (2006). Structural laplace transform and compound autoregressive models. *Journal of Time Series Analysis*, 27(4):477–503.

Davis, R. A., Dunsmuir, W. T., and Streett, S. B. (2003). Observation-driven models for poisson counts. *Biometrika*, 90(4):777–790.

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.

Duffie, D., Filipović, D., and Schachermayer, W. (2003). Affine Processes and Applications in Finance. *Annals of Applied Probability*, 13(3):984–1053.

Englund, M., Guillén, M., Gustafsson, J., Nielsen, L. H., and Nielsen, J. P. (2008). Multivariate latent risk: A credibility approach. *Astin Bulletin*, 38(01):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.

Frees, E. W. (2003). Multivariate credibility for aggregate loss models. *North American Actuarial Journal*, 7(1):13–37.

Frees, E. W. and Wang, P. (2005). Credibility using copulas. *North American Actuarial Journal*, 9(2):31–48.

Gerber, H. and Jones, D. (1975). Credibility formulas of the updating type. *Transactions of the Society of Actuaries*, 27:31–52.

Gourieroux, C. and Jasiak, J. (2004). Heterogeneous inar (1) model with application to car insurance. *Insurance: Mathematics and Economics*, 34(2):177–192.

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

Gourieroux, C. and Monfort, A. (1997). *Simulation-based econometric methods*. OUP Oxford.

Gourieroux, C., Monfort, A., and Trognon, A. (1984). Pseudo maximum likelihood methods: Applications to poisson models. *Econometrica*, pages 701–720.

Harvey, A. C. and Fernandes, C. (1989). Time series models for count or qualitative observations. *Journal of Business & Economic Statistics*, 7(4):407–417.

Heckman, J. J. (1987). The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process and some monte carlo evidence. *in Structural Analysis of Discrete Data, (ed) Charles F. Manski, Daniel Mcfadden*.

Henderson, R. and Shimakura, S. (2003). A serially correlated gamma frailty model for longitudinal count data. *Biometrika*, 90(2):355–366.

Jewell, W. S. (1974). Credible means are exact bayesian for exponential families. *Astin Bulletin*, 8(01):77–90.

Le Courtois, O. (2015). q-credibility. *Available at SSRN 2571195*.

Lemaire, J. (1995). Bonus-malus systems in automobile insurance. *Insurance Mathematics and Economics*, 3(16):277.

Najafabadi, A. T. P. (2010). A new approach to the credibility formula. *Insurance: Mathematics and Economics*, 46(2):334–338.

Pinquet, J. (1998). Designing optimal bonus-malus systems from different types of claims. *Astin Bulletin*, 28(02):205–220.

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

Purcaru, O. and Denuit, M. (2003). Dependence in dynamic claim frequency credibility models. *Astin Bulletin*, 33(01):23–40.

Schrager, D. F. (2006). Affine stochastic mortality. *Insurance: mathematics and economics*, 38(1):81–97.

Sim, C.-H. (1990). First-order autoregressive models for gamma and exponential processes. *Journal of Applied Probability*, pages 325–332.

Sundt, B. (1988). Credibility estimators with geometric weights. *Insurance: Mathematics and Economics*, 7(2):113–122.

Tremblay, L. (1992). Using the poisson inverse gaussian in bonus-malus systems. *Astin Bulletin*, 22(01):97–106.

Varin, C., Reid, N., and Firth, D. (2011). An Overview of Composite Likelihood Methods. *Statistica Sinica*, 21(1):5–42.

Zeger, S. L. (1988). A regression model for time series of counts. *Biometrika*, 75(4):621–629.