Insurance Mathematics and Economics 125 (2025) 103149

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

An observation-driven state-space count model for experience rating

Jae Youn Ahn [a], Himchan Jeong [b], Yang Lu [c], Mario V. Wüthrich [d,*]

[a] Department of Statistics, Ewha Womans University, Seoul, Republic of Korea
[b] Department of Statistics and Actuarial Science, Simon Fraser University, BC, Canada
[c] Department of Mathematics and Statistics, Concordia University, Montreal, QC, Canada
[d] Department of Mathematics, ETH Zurich, Switzerland

ARTICLE INFO

JEL classification:
C32
C53
G22

Keywords:
State-space model
Experience rating
Posterior ratemaking
Bayesian theory
Credibility theory
Empirical Bayes estimation
Conjugate priors
Poisson-Gamma model
Random effects model
Seniority weighting
Bühlmann credibility

ABSTRACT

State-space models are widely used in applications, e.g., in economics, finance and actuarial science. In the domain of count data, one such example is the model proposed by Harvey and Fernandes (1989). Unlike many of its parameter-driven alternatives, this model is observation-driven, and it leads to a closed-form expression for the predictive density. This predictive density takes into account past observations by assigning a seniority weighting to them. This feature makes this model very appealing for general insurance ratemaking. However, the model of Harvey and Fernandes (1989) has the property that the variance diverges in the long-run, which might be an undesirable model feature. In this paper, we extend the model of Harvey and Fernandes (1989) by allowing for flexible variance specifications including non-explosive ones, while keeping the model fully tractable.

1. Introduction

Since the pioneering work of Bichsel (1964) and Dionne and Vanasse (1989), the Poisson-Gamma random effects model has established as a key tool for experience rating in general insurance; see Duval et al. (2024) for an application including telematics data. This random effects model is based on the Poisson-Gamma conjugacy leading to a closed-form pricing formula, that is, providing a fully tractable modeling approach to derive a linear credibility formula, called Bühlmann (1967) credibility. One important shortcoming of this approach is that it uses a common latent random factor Θ that influences all response variables $(Y_t)_{t\geq 1}$ equally. As a consequence, the components of the past claims history $Y_{1:t} = (Y_1, \ldots, Y_t)$ are exchangeable for forecasting the next response Y_{t+1}, meaning that all components of $Y_{1:t}$ receive the same credibility weighting in forecasting Y_{t+1}. Allowing for a seniority weighting of past claims has been advocated in the credibility literature; see, e.g., Gerber and Jones (1973), Sundt (1988), Pinquet et al. (2001), Pinquet (2020a) or Pinquet (2020b). Recently, also statistical tests considering this have been proposed; see Li et al. (2023) and Fung et al. (2024). The issue of seniority weighting has been addressed in the random effects literature by replacing the static (Gamma) random effect by a stochastic process. In particular, Brouhns et al. (2003) and Purcaru and Denuit (2003) assume a log-normal dynamics for this stochastic process. The difficulty with this approach is that this model does not allow for a closed-form log-likelihood function nor an analytical premium updating rule. Lu (2018) suggested to replace the log-normal dynamics by an autoregressive Gamma process, and he derived a closed-form log-likelihood function with an analytical predictive formula. This formula, however, involves the derivative of a matrix function, whose dimension is equal to the number of observed periods t. Thus, although for small t this approach might be more tractable than the approach of Brouhns et al. (2003), it may still become prohibitive for increasing dimensions.

* Corresponding author.
E-mail addresses: jaeyahn@ewha.ac.kr (J.Y. Ahn), himchan_jeong@sfu.ca (H. Jeong), yang.lu@concordia.ca (Y. Lu), mario.wuethrich@math.ethz.ch (M.V. Wüthrich).

https://doi.org/10.1016/j.insmatheco.2025.103149
Received 24 April 2024; Received in revised form 24 June 2025; Accepted 30 June 2025

Available online 20 August 2025
0167-6687/© 2025 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

The computational difficulties of these dynamic random effects models relate to the fact that from a time-series point of view, both the approaches of Brouhns et al. (2003) and Lu (2018) fall into the category of parameter-driven state-space models. These models have the property that the responses are conditionally independent given the path of the latent stochastic process, whose dynamics is exogenous. It is well documented in this literature that besides the Gaussian state-space model of Kalman (1960), most non-Gaussian parameter-driven state-space models do not have tractable log-likelihood functions and are, thus, computationally intensive; see, e.g., Zeger (1988), Henderson and Shimakura (2003), Frühwirth-Schnatter and Wagner (2006), Cui and Lund (2009), Davis and Wu (2009), Jung et al. (2011) and Jia et al. (2023). A notable special case is the class of hidden Markov models (HMMs), which offer closed-form expressions for estimation and prediction due to their discrete latent state-space structure; see Baum and Petrie (1966). Such models have been applied in an insurance ratemaking and telematics data context by Jiang and Shi (2024), Shi et al. (2025) and Chan et al. (2025). The latent state-space in HMMs is restricted to a finite number of discrete values, which can limit the modeling flexibility and accuracy if the number of states is small. Increasing the number of latent states can enhance flexibility and accuracy, but comes at the cost of increased computational complexity, as HMM algorithms typically scale quadratically with the number of hidden states. Moreover, the choice of an appropriate number of states of the HMM model is highly non trivial. Consequently, HMM models are not considered in this paper. Finally, let us also mention the copula approach, see, e.g., Frees and Wang (2005), Shi and Yang (2018) and Yang (2022), who adopt a model without any random effect, which they call the "frequentist" approach. Frees and Wang (2005) mention that "the frequentist perspective avoids assumptions concerning the prior distribution of the latent variables. This may be seen as an advantage or disadvantage, depending on the situation." Furthermore, the use of copulas when count data are involved can cause potential identification issues as explained by Genest and Nešlehová (2007).

Is there a way to avoid the computational burden of parameter-driven state-space models, while keeping their nice interpretation as well as their advantage of allowing for a seniority weighting of past claims? The answer is affirmative. The Harvey and Fernandes (1989) model (HF model) is one of the rare examples of an *observation-driven* state-space model for count data, which enjoys mathematical tractability. More precisely, in contrast to its *parameter-driven* counterparts, the dynamics of the latent process is not exogenous, but endogenous, in the sense that the random effects process $(\Theta_t)_{t\geq 1}$ depends not only on the past values of this state variable, but also on the past values of the observed responses $Y_{1:t}$, i.e., it is based on the assumption

$$\Theta_{t+1}|(\Theta_{1:t}, Y_{1:t}) = \Theta_{t+1}|(\Theta_t, Y_{1:t}) \neq \Theta_{t+1}|\Theta_t, \tag{1}$$

for $\Theta_{1:t} = (\Theta_1, \ldots, \Theta_t)$. This says that the law of the next state Θ_{t+1} depends on the previous state Θ_t *and* the past responses $Y_{1:t}$.

The tractability of the predictive distribution in the HF model is achieved by adding a feedback loop from the observations $Y_{1:t}$ to the next state-space variable Θ_{t+1}. This feedback is of Poisson-Gamma conjugate prior type, leading to tractable models and log-likelihoods. From an interpretation point of view, such a model makes perfect sense in insurance pricing because a (bad) claims experience may result in actions that change the latent factors, e.g., a car driver may change her/his driving behavior and take a skid course after an accident. In the insurance literature, it has been applied by Bolancé et al. (2007) and Boucher and Pigeon (2018) to univariate count processes, by Abdallah et al. (2016) to bivariate count processes and by Ahn et al. (2023) to collective risk modeling.

In this paper, we start by explaining that in the HF model, the state-space variable follows a (multiplicative) random walk, and it has an increasing and exploding variance process as the number of observed periods t increases to infinity. This explosion-in-variance might not be suitable for many applications where t is large. One prime example is telematics insurance, where insurers collect (high-frequency) data of the driving behavior of their policyholders, e.g., a daily or weekly driving behavior signal; see Yanez et al. (2025). Other examples are cyber risk, which has the feature of evolving very fast and, thus, needs frequent updating of the risk development, or health trackers in health insurance. In these examples, we likely result in a large t situation, where the asymptotic behavior of the state-space model is relevant. The HF model has a long-term explosive variance behavior, whereas, e.g., in the telematics example we rather expect a bounded or a decreasing variance behavior. The latter is based on the intuition that increasing experience should lead to decreasing risk and uncertainty. Therefore, the aim of this paper is to extend the HF model to a model class that accommodates various non-explosive variance behaviors, including time-invariant or converging variance processes. As a preview of our results, we are able to classify this extended class of observation-driven state-space models for counts into several groups:

(a) The original HF model, which has an *explosive (non-stationary)* dynamics with increasing variance $\liminf_{t\to\infty} \mathrm{Var}\left(\Theta_t\right) = \infty$.
(b) The second class corresponds to the case where, when t goes to infinity, the latent process $(\Theta_t)_{t\geq 1}$ degenerates to a constant and, hence, the uncertainty related to the non-observability of the latent factor Θ_t asymptotically vanishes.
(c) In the third class, the latent process $(\Theta_t)_{t\geq 1}$ considered in (1) has a variance that remains bounded away from zero and infinity. This includes special cases of time-invariant or converging variance processes. Importantly, variance stationarity within this class provides a practical compromise: general time-series models typically introduce excessive complexity for insurance settings characterized by relatively short observation periods, while static random-effects models fail to capture the temporal dependencies of claims histories. Thus, a variance-stationary model represents a realistic and effective balance, that allows for the incorporation of seniority-weighting of past claims information while keeping the model complexity manageable.

The models in these three classes differ in terms of their (variance) stationarity properties, but they all enjoy the aforementioned properties of providing tractable log-likelihoods and experience rating pricing formulas.

The rest of the paper is organized as follows. Section 2 extends the HF model, by allowing some of the parameters of the HF model to be time-varying. Section 3 discusses various specifications of this extended HF family, and it classifies them into different classes according to their stationarity (or non-stationarity) behavior, see Table 1. Section 4 illustrates the resulting credibility formulas. Section 5 compares these models using a real insurance dataset. Section 6 concludes. Proofs of the mathematical statements are provided in the appendix.

2. The extended HF model

Often, in insurance applications, count responses Y are available as longitudinal data, that is, they are doubly indexed by both the policyholder i and the time period t. In the rest of this section, we will temporarily remove index i in order to focus on the properties of the model in the time-series dimension. The portfolio aspect in policyholder index $1 \leq i \leq n$ will become relevant once we consider empirical Bayes' fitting, we refer to (30), below.

Throughout this paper, we use the following notation.

- Gamma(α, β): Gamma distribution with shape parameter $\alpha > 0$ and rate (scale) parameter $\beta > 0$, having mean α/β and variance α/β^2; by convention, Gamma$(0, \beta)$ denotes a constant zero.

Table 1
Typology of various special cases of Model 1 according to the long-run behavior of the variance process. In this table, all constants p and q lie strictly between 0 and 1. The condition "$(\lambda_t)_{t\geq 1}$ bounded" implies that the process is bounded both above and below by positive constants.

Model	Condition
Classical NB regression model	$p_t = 0$ and q_t defined in (33)
Static random effects	$p_t = q_t = 1$
HF model with increasing (explosive) variance	$p_t = 1,\ q_t = q \in (0,1)$, $(\lambda_t)_{t\geq 1}$ bounded
Decreasing variance	$p_t = p \in [0,1),\ q_t = 1$, $(\lambda_t)_{t\geq 1}$ bounded
Converging variance	$p_t = p \in [0,1),\ q_t = q \in (0,1),\ \lambda_t = \lambda$
Bounded variance	$p_t = p \in (0,1),\ q_t = q \in (0,1)$, $(\lambda_t)_{t\geq 1}$ bounded
Constant variance	p_t, q_t defined in (21)

- Beta(α,β): beta distribution on $(0,1)$ with mean $\frac{\alpha}{\alpha+\beta}$ and variance $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$; by convention, Beta$(\alpha,0)$ denotes a constant one, and Beta$(0,\beta)$ a constant zero.
- Pois(λ): Poisson distribution with mean $\lambda > 0$.
- NB(λ,γ): negative binomial (NB) distribution with mean λ and variance $\lambda + \lambda^2/\gamma$.

We recall the Poisson-Gamma conjugacy (Bühlmann and Gisler, 2006, Section 2.4.1). If Y is Poisson, given Θ, with mean Θ, and Θ follows a Gamma prior distribution, then the posterior distribution of Θ, given Y, is also a Gamma distribution, and the marginal distribution of Y is a NB distribution.

2.1. The model

We provide an observation-driven state-space model with constant mean extending the HF model.

Model 1. Given exogenous processes $(\lambda_t)_{t\geq 1}$, $(p_t)_{t\geq 1}$ and $(q_t)_{t\geq 1}$ satisfying for $t \geq 1$

$$\lambda_t \geq 0, \qquad p_t \in (0,1), \qquad q_t > 0 \quad \text{and} \quad p_t q_t < 1, \tag{2}$$

the response variables $(Y_t)_{t\geq 1}$ and the state-space variables $(\Theta_t)_{t\geq 1}$ satisfy:

i) The conditional distribution of Y_t, given past observations[1] $Y_{1:(t-1)}$ and the state-space variables $\Theta_{1:t}$, is Poisson

$$Y_t | (Y_{1:(t-1)}, \Theta_{1:t}) \sim \text{Pois}\left(\lambda_t \Theta_t\right), \qquad \text{for } t \geq 1. \tag{3}$$

For $\lambda_t = 0$, we set $Y_t = 0$, a.s.

ii) At time $t = 1$, Θ_1 is Gamma distributed as

$$\Theta_1 \sim \text{Gamma}\left(\alpha_{1|0}, \beta_{1|0}\right), \tag{4}$$

with, for identification purposes, we assume equal deterministic parameters $\alpha_{1|0} = \beta_{1|0} > 0$, so that $\mathbb{E}\left[\Theta_1\right] = 1$.

iii) Assume that, given past observations $Y_{1:t}$, the filtering distribution of Θ_t at time $t \geq 1$ is given by

$$\Theta_t | Y_{1:t} \sim \text{Gamma}\left(\alpha_t, \beta_t\right), \tag{5}$$

where $\alpha_t > 0$ and $\beta_t > 0$ are deterministic functions of $Y_{1:t}$ and $\lambda_{1:t}$. Then, the predictive distribution of Θ_{t+1} is given by

$$\Theta_{t+1} | Y_{1:t} \sim \text{Gamma}\left(\alpha_{t+1|t}, \beta_{t+1|t}\right), \tag{6}$$

with for $t \geq 1$

$$\begin{aligned} \beta_{t+1|t} &= q_t \beta_t \ = \ q_t \left(\beta_{t|t-1} + \lambda_t\right), \\ \alpha_{t+1|t} &= p_t q_t \alpha_t + \left(1 - p_t\right) \beta_{t+1|t} \ = \ p_t q_t \left(\alpha_{t|t-1} + Y_t\right) + \left(1 - p_t\right) \beta_{t+1|t}. \end{aligned} \tag{7}$$

We give some explanations to Model 1.

- We start with (5). Using the Poisson-Gamma conjugacy applied to the tuple (3) and (6) as well as (4), Bayes' rule determines the distribution in (5) and its parameters $\alpha_t > 0$ and $\beta_t > 0$ as deterministic functions of the past observations $Y_{1:t}$ and the exogenous parameters $\lambda_{1:t}$, given by

$$\alpha_t = \alpha_{t|t-1} + Y_t, \tag{8}$$

and

$$\beta_t = \beta_{t|t-1} + \lambda_t. \tag{9}$$

As a result, Model 1 is well defined.
- The critical step, compared to the traditional Poisson-Gamma random effects model, is the evolution step from (5) to (6). The primary motivation for this evolution step in (6) with the specification in (7) is the construction of a Gamma process guided by a thinning property with the following characteristic:

$$\mathbb{E}[\Theta_{t+1} \mid Y_{1:t}] = p_t \, \mathbb{E}[\Theta_t \mid Y_{1:t}] + (1 - p_t). \tag{10}$$

This property immediately guarantees mean stationarity, ensuring that $\mathbb{E}[\Theta_t] = 1$ for all $t \geq 1$; this is based the identification used in (4). However, its variance behavior depends on the choice of the exogenous processes $(p_t)_{t\geq 1}$, $(q_t)_{t\geq 1}$ and $(\lambda_t)_{t\geq 1}$. These aspects will be explained in detail in Section 2.3, below. We note that the linear evolution specification chosen in (7) primarily aims at simplicity and mathematical tractability, particularly for deriving explicit mean and variance behaviors. Non-linear alternatives could also be explored in future research, albeit potentially at the expense of analytical tractability.
- The exogenous parameters $(\lambda_t)_{t\geq 1}$ allow one to integrate time-varying covariates. In our example, we are going to model these parameters using a generalized linear model (GLM) by setting $\lambda_t = \mu(\boldsymbol{x}_t)$ for covariates $\boldsymbol{x}_t$ and a regression function $\mu(\cdot)$ having a GLM structure.

2.2. The HF model as a special case

The HF model of Harvey and Fernandes (1989) is obtained as a special case of Model 1 under the following constraints for $t \geq 1$

$$q_t \equiv q \ \in \ (0,1) \qquad \text{and} \qquad p_t \equiv 1. \tag{11}$$

Harvey and Fernandes (1989)'s original formulation is without covariates, Gamerman et al. (2013) extend their model by introducing a time-varying exogenous heterogeneous sequence $(\lambda_t)_{t\geq 1}$. Now, this type of model is widely recognized across various fields, including economics

[1] By convention, for $t = 1$, the information set $\sigma(Y_{1:(t-1)})$ reduces to the trivial σ-field.

(Fokianos, 2024) and climate studies (Duca et al., 2021). In this model, the evolution step implies $\mathbb{E}\left[\Theta_{t+1}|Y_{1:t}\right] = \mathbb{E}\left[\Theta_t|Y_{1:t}\right]$. In other words, $(\Theta_t)_{t\geq 1}$ is a martingale with respect to the filtration generated by $Y_{1:t}$, $t \geq 1$, this is comparable to the filtering property described in (10).

Lemma 3 in Section 3, below, demonstrates that the HF model's martingale property leads to an explosive variance behavior, which may limit its applicability in many contexts. In contrast, our proposed model offers flexibility in modeling various variance behaviors depending on the specification of the exogenous sequences $(p_t)_{t\geq 1}$, $(q_t)_{t\geq 1}$, and $(\lambda_t)_{t\geq 1}$. This flexibility, achieved through the evolution step detailed in Section 2.3, allows the proposed model to address the limitations of the HF model while maintaining the analytical tractability.

2.3. Stochastic representation of the observation-driven property

Model 1 defines the conditional distributions of the latent variable Θ_t, given the past (or past and current) observations. However, it does not provide a state-space equation directly linking Θ_t with Θ_{t+1}. To work out a state-space equation, we first recall the following lemma.

Lemma 1 *(Lukacs (1955)). Consider two independent random variables*

$$\Theta \sim \text{Gamma}(\alpha, \beta) \qquad \textit{and} \qquad B \sim \text{Beta}\left(p^*\alpha, (1-p^*)\alpha\right), \tag{12}$$

where $\alpha > 0, \beta > 0$ and $p^* \in (0, 1]$ are given constants. Then, their product

$$\Theta B \sim \text{Gamma}(p^*\alpha, \beta).$$

Consequently, if η is independent of Θ and B, and $\eta \sim \text{Gamma}((q - p^*)\beta, q\beta)$ with constant q such that $q \geq p^*$, then we have

$$\frac{\Theta B}{q} + \eta \sim \text{Gamma}\left(p^*\alpha + \left(q - p^*\right)\beta, q\beta\right). \tag{13}$$

Formula (13) implies the following stochastic representation of the latent process $(\Theta_t)_{t\geq 1}$ by setting $p_t^* = p_t q_t \in (0, 1]$, see (2),

$$\Theta_{t+1} = \frac{\Theta_t B_{t+1}}{q_t} + \eta_{t+1}, \tag{14}$$

where Θ_t, given $Y_{1:t}$, satisfies (5), and

$$B_{t+1}|Y_{1:t}, \Theta_{1:t} \sim \text{Beta}(p_t q_t \alpha_t, (1 - p_t q_t)\alpha_t),$$

and

$$\eta_{t+1}|Y_{1:t}, \Theta_{1:t} \sim \text{Gamma}((1 - p_t)q_t\beta_t, q_t\beta_t).$$

Moreover, B_{t+1} and η_{t+1} are assumed to be conditionally independent given $Y_{1:t}$ and $\Theta_{1:t}$. This result provides us with an explicit construction of Model 1, in particular, it gives us a way to simulate under this Model 1.

From the above stochastic representation, the latent process $(\Theta_t)_{t\geq 1}$ satisfies the linear Markov property

$$\mathbb{E}\left[\Theta_{t+1} \mid \Theta_{1:t}, Y_{1:t}\right] = p_t\Theta_t + (1 - p_t). \tag{15}$$

Hence, the conditional expectation of the latent process follows an autoregressive-type recursion of order 1 with coefficient p_t. Note, that this does not imply that the observed process $(Y_t)_{t\geq 1}$ is AR(1) type. Indeed, in Appendix C, we show that process $(Y_t)_{t\geq 1}$ is of ARMA(1,1) type.

3. A classification according to the behavior of the variance process

In this section, we show that Model 1 can allow for various forms of variance behavior, e.g., of increasing, decreasing, constant or stationary type. To analyze more sophisticated cases, let us first investigate some properties of Model 1 concerning the first and second order moments of the processes.

Lemma 2. *In Model 1 we have the following moment behaviors for $t \geq 1$:*

i) $\mathbb{E}\left[\alpha_t\right] = \beta_t$, $\mathbb{E}\left[\Theta_t\right] = 1$ *and* $\mathbb{E}\left[Y_t\right] = \lambda_t$.
ii) $\mathbb{E}\left[\text{Var}\left(\Theta_t \mid Y_{1:t}\right)\right] = \frac{1}{\beta_t}$.
iii) $\mathbb{E}\left[\text{Var}\left(\Theta_{t+1} \mid Y_{1:t}\right)\right] = \frac{1}{q_t}\frac{1}{\beta_t}$.
iv) $\text{Var}\left(\mathbb{E}\left[\Theta_{t+1} \mid Y_{1:t}\right]\right) = p_t^2\,\text{Var}\left(\mathbb{E}\left[\Theta_t \mid Y_{1:t}\right]\right)$.

Proof. See Appendix B.1. □

Property *i)* says that the process $(\Theta_t)_{t\geq 1}$ is mean-stationary. Properties *iii)* and *iv)* allow us to analyze the variance behavior of $(\Theta_t)_{t\geq 1}$. By the total variance decomposition formula, we get

$$\begin{aligned}\text{Var}\left(\Theta_{t+1}\right) &= \mathbb{E}\left[\text{Var}\left(\Theta_{t+1} \mid Y_{1:t}\right)\right] + \text{Var}\left(\mathbb{E}\left[\Theta_{t+1} \mid Y_{1:t}\right]\right) \\ &= \frac{1}{q_t}\frac{1}{\beta_t} + p_t^2\,\text{Var}\left(\mathbb{E}\left[\Theta_t \mid Y_{1:t}\right]\right).\end{aligned} \tag{16}$$

On the other hand, by using the total variance decomposition formula again, we have

$$\begin{aligned}\text{Var}\left(\Theta_t\right) &= \mathbb{E}\left[\text{Var}\left(\Theta_t \mid Y_{1:t}\right)\right] + \text{Var}\left(\mathbb{E}\left[\Theta_t \mid Y_{1:t}\right]\right) \\ &= \frac{1}{\beta_t} + \text{Var}\left(\mathbb{E}\left[\Theta_t \mid Y_{1:t}\right]\right).\end{aligned} \tag{17}$$

Let us compare (16) and (17). Choosing $q_t \leq 1$, the first term in (16) is larger than the first one in (17). Similarly, since by assumption $p_t^2 \leq 1$, the second term in (16) is smaller than the second one in (17). As a result, the variance process $(\text{Var}\left(\Theta_t\right))_{t\geq 1}$ in our model is not necessarily monotone, i.e., depending on the specific choices of the exogenous sequences it may behave differently. Throughout the remainder of this section, we focus on three main types of variance behaviors $(\text{Var}\left(\Theta_t\right))_{t\geq 1}$: increasing variance (Section 3.1), decreasing variance (Section 3.2), and constant variance (Section 3.3). Additional cases with other variance behaviors are provided in Appendix A. We begin by describing the increasing variance case.

3.1. HF model: a model with increasing variance

As demonstrated in Section 2.2, the HF model can be derived as a special case of Model 1 by specifying the exogenous sequences according to (11). The following lemma establishes that the HF model exhibits an explosive variance behavior under certain conditions.

Lemma 3 *(Explosive variance in the HF model). In the HF model, if the exogenous process $(\lambda_t)_{t\geq 1}$ is uniformly bounded both above and below by positive constants,* $\text{Var}\left(\Theta_t\right)$ *and* $\text{Var}\left(Y_t\right)$ *diverge to infinity as $t \to \infty$.*

An advantage of the HF model is that it provides a predictive mean, $\mathbb{E}[Y_{t+1} \mid Y_{1:t}]$, which can take the form of an exponentially weighted moving average (EWMA) of past observations depending on the choice of the exogenous sequences; see Section 4, below, for the specific formulation. The literature on EWMA forecasting of count variables has traditionally focused on *non-stationary* models only; see, e.g., Hyndman et al. (2008), Chapter 17. It is seen in the next subsection, however, that it is not the only possible specification leading to EWMA predictors.

3.2. A model with decreasing variance

In this subsection, we consider the special case of Model 1 with constraints for $t \geq 1$

$$p_t = p \in [0, 1] \quad \text{and} \quad q_t = 1.$$

Especially, when p_t is strictly less than 1,

$$p_t = p \in [0, 1) \quad \text{and} \quad q_t = 1, \tag{18}$$

we have updates

$$\alpha_t = p\alpha_{t-1} + (1-p)\beta_{t-1} + Y_t \qquad \text{and} \qquad \beta_t = \beta_{t-1} + \lambda_t,$$

for $t \geq 2$. If $(\lambda_t)_{t\geq 1}$ is uniformly bounded from below, then both processes $(\beta_t)_{t\geq 1}$ and $(\alpha_t)_{t\geq 1}$ go to infinity when t increases to infinity.

Let us study the variance of the resulting state-space process. By comparing (16) and (17) with the condition in (18), we get

$$\begin{aligned}\mathrm{Var}\left(\Theta_{t+1}\right) - \mathrm{Var}\left(\Theta_t\right) &= \left(p^2-1\right)\mathrm{Var}\left(\mathbb{E}\left[\Theta_t \,|\, Y_{1:t}\right]\right) \\ &= \left(p^2-1\right)\mathrm{Var}\left(\frac{\alpha_t}{\beta_t}\right) < 0.\end{aligned} \tag{19}$$

Thus, the latent process $(\Theta_t)_{t\geq 1}$, and hence $(Y_t)_{t\geq 1}$, have both a decreasing variance (for the latter it is sufficient to assume that $(\lambda_t)_{t\geq 1}$ is bounded). Moreover, we have the following stronger result.

Lemma 4. *Under Model 1 and constraint (18), if the exogenous process $(\lambda_t)_{t\geq 1}$ is uniformly bounded both above and below by positive constants, then* $\mathrm{Var}\left(\Theta_t\right)$ *converges to zero for $t \to \infty$, and* $\mathrm{Var}\left(Y_t\right)$ *converges to λ, under the additional assumption of the following time-invariance property*

$$\lambda_t = \lambda > 0, \quad \textit{for all} \quad t \geq 1.$$

Proof. See Appendix B.3. □

3.3. A model with constant variance

In this subsection, we discuss a special case of Model 1 for which the variance of Θ_t is constant in $t \geq 1$. Note that we do not require the process to be covariance stationary, i.e., the covariances $\mathrm{Cov}(\Theta_t, \Theta_{t+h})$ may still depend on t. That is, we study the case

$$1 = \mathbb{E}\left[\Theta_t\right] \qquad \text{and} \qquad \frac{1}{\beta_{1|0}} = \mathrm{Var}\left(\Theta_1\right) = \mathrm{Var}\left(\Theta_t\right), \qquad \text{for all } t \geq 1, \tag{20}$$

where the mean stationarity is given by Lemma 1. By comparing (16) and (17), an immediate consequence is the following result.

Lemma 5. *In Model 1, the variance process* $\left(\mathrm{Var}\left(\Theta_t\right)\right)_{t\geq 1}$ *is constant, if and only if p_t and q_t satisfy the following equation for all $t \geq 1$*

$$\frac{1}{q_t}\frac{1}{\beta_t} + p_t^2\left(\frac{1}{\beta_{1|0}} - \frac{1}{\beta_t}\right) = \frac{1}{\beta_{1|0}}. \tag{21}$$

Consequently, there are infinitely many possible combinations of $(p_t)_{t\geq 1}$ and $(q_t)_{t\geq 1}$ in order for $\left(\mathrm{Var}\left(\Theta_t\right)\right)_{t\geq 1}$ to remain constant. Among such choices, motivated by the conditional linear auto-regressive structure (CLAR(1)) of Grunwald et al. (2000), we may assume the following updating rule

$$\mathbb{E}\left[\Theta_{t+1} \,|\, Y_{1:t}\right] = p\,\mathbb{E}\left[\Theta_t \,|\, Y_{1:t}\right] + 1 - p, \tag{22}$$

equivalently saying $p_t \equiv p \in (0,1)$ is time-invariant. Then, under this additional assumption (22), requirement (21) becomes for $t \geq 1$

$$q_t = \frac{\beta_{1|0}}{p^2\beta_{1|0} + \left(1-p^2\right)\beta_t}, \tag{23}$$

which can be calculated recursively from

$$\beta_t = q_{t-1}\beta_{t-1} + \lambda_t, \quad t \geq 2,$$

with initialization

$$\beta_1 = \beta_{1|0} + \lambda_1.$$

When $p_t = 1$ and $q_t = 1$, equivalently $B_{t+1} \equiv 1$ and $\eta_{t+1} \equiv 0$, Model 1 becomes the static random effects Poisson-Gamma model. This model (with a static random effect) assumes that there is a given time-invariant latent variable $\Theta_t \equiv \Theta$, a.s., for all $t \geq 1$. Given $\Theta \sim \mathrm{Gamma}(\alpha_{1|0}, \beta_{1|0})$, the counts Y_t are conditionally independent with a $\mathrm{Pois}(\lambda_t\Theta)$ distribution. For $\lambda_t \equiv \lambda > 0$, this is a special case of the Bühlmann and Straub (1970) credibility model.

3.4. A typology of models according to the variance process

The previous Sections 3.1, 3.2 and 3.3 have presented the three special cases of a diverging variance behavior, an asymptotically vanishing variance behavior and a constant one, respectively, of the state-space process. Naturally, there are many other interesting cases for practical applications, e.g., a converging variance or bounded variance behavior. To not overload this outline, other interesting cases are deferred to Appendix A. Table 1 summarizes the variance behavior of various models considered in this section as well as in Appendix A, along with the independent case (Classical NB regression model) for reference.

The evolution step defined in (6) is the key element governing the dynamics of the evolutionary credibility, see Section 4, below. Depending on the value of p_t, the evolution step of the models in Table 1 can be classified into three distinct types:

i. Models with $p_t = 0$: In this case, the predictive distribution is fully determined by the exogenous process, and the predictive mean is given by

$$\mathbb{E}\left[Y_{t+1} \middle| Y_{1:t}\right] = \lambda_{t+1}.$$

ii. Models with $p_t = 1$: Here, the predictive mean exhibits an invariance property, relying entirely on the conditional expectation of the latent state-space variable

$$\mathbb{E}\left[Y_{t+1} \middle| Y_{1:t}\right] = \lambda_t\,\mathbb{E}\left[\Theta_t \middle| Y_{1:t}\right].$$

iii. Models with $p_t \in (0,1)$: In this intermediate case, the predictive distribution is determined as a weighted average of two components, the posterior mean and the exogenous mean

$$\mathbb{E}\left[Y_{t+1} \middle| Y_{1:t}\right] = p_t\lambda_{t+1}\mathbb{E}\left[\Theta_t \middle| Y_{1:t}\right] + (1-p_t)\lambda_{t+1}.$$

4. Evolutionary credibility and credibility formulas

The above discussion has been focusing on characterizing model types with different long-term variance behaviors. In this section, we verify that these models lead to tractable log-likelihoods and to evolutionary credibility formulas for the responses Y_{t+1}, given $Y_{1:t}$. The concept of evolutionary credibility plays a crucial role in the derivation of experience rating pricing schemes in insurance; see Chapter 9 in Bühlmann and Gisler (2006).

To begin, from Model 1, we derive the explicit formulation of the joint probability mass function in an autoregressive form

$$f(y_{1:t}) = \prod_{s=1}^{t} f(y_s \,|\, y_{1:s-1}),$$

where each term $f(y_s \,|\, y_{1:s-1})$ for $s \geq 2$ denotes the conditional probability mass function of Y_s, given $Y_{1:s-1} = y_{1:s-1}$, and the marginal distribution of Y_1 is given by $f(y_1) = f(y_1 \,|\, y_{1:0})$.

These conditional probability mass functions follow negative binomial distributions derived from

$$Y_{t+1} \,|\, Y_{1:t} \sim \mathrm{NB}\left(\lambda_{t+1}\frac{\alpha_{t+1|t}}{\beta_{t+1|t}}, \alpha_{t+1|t}\right), \tag{24}$$

and

$$Y_1 \sim \mathrm{NB}\left(\lambda_1\frac{\alpha_{1|0}}{\beta_{1|0}}, \alpha_{1|0}\right).$$

As a result, the log-likelihood is fully tractable and decomposable across time, and the model parameters can be estimated efficiently using standard GLM routines and gradient descent methods. This feature offers a significant computational advantage over more complex alternatives such as hidden Markov models (HMMs).

Directly derived from (24), we have predictive mean

$$\mathbb{E}\left[Y_{t+1}\middle|Y_{1:t}\right]=\lambda_{t+1}\mathbb{E}\left[\Theta_{t+1}\middle|Y_{1:t}\right],$$

which indicates that deriving a recursive formula for the predictive mean of the latent variables is sufficient to construct the evolutionary credibility formula.

Now, we derive the recursive formula for the predictive mean of the latent variables. Using (6), we have

$$\mathbb{E}\left[\Theta_{t+1}\middle|Y_{1:t}\right]=p_t\mathbb{E}\left[\Theta_t\,|\,Y_{1:t}\right]+\left(1-p_t\right)\cdot 1, \tag{25}$$

we add the "1" to indicate that this is a weighted average between the posterior mean $\mathbb{E}\left[\Theta_t\,|\,Y_{1:t}\right]$ and the prior mean $\mathbb{E}\left[\Theta_{t+1}\right]=1$. Furthermore, for the derivation of the posterior mean we have

$$\mathbb{E}\left[\Theta_t\,|\,Y_{1:t}\right]\;=\;\frac{\alpha_t}{\beta_t}=\frac{\alpha_{t|t-1}+Y_t}{\beta_{t|t-1}+\lambda_t}.$$

To simplify notation, for $\lambda_t=0$, we use the conventions

$$\frac{Y_t}{\lambda_t}:=0\qquad\text{and}\qquad\frac{\mathbb{E}\left[Y_t\,|\,Y_{1:t-1}\right]}{\lambda_t}:=\mathbb{E}\left[\Theta_t\,|\,Y_{1:t-1}\right].$$

With these definitions, the recursion can be expressed as a weighted average of the following form

$$\mathbb{E}\left[\Theta_t\,|\,Y_{1:t}\right]=z_t\,\frac{Y_t}{\lambda_t}+(1-z_t)\mathbb{E}\left[\Theta_t\,|\,Y_{1:t-1}\right],$$

with credibility weight

$$z_t=\frac{\lambda_t}{\beta_{t|t-1}+\lambda_t}\;\in\;[0,1).$$

Combining this with (25), we obtain the following recursion of the predictive mean of the latent variables

$$\mathbb{E}\left[\Theta_{t+1}\middle|Y_{1:t}\right]=p_tz_t\,\frac{Y_t}{\lambda_t}+p_t(1-z_t)\mathbb{E}\left[\Theta_t\middle|Y_{1:t-1}\right]+\left(1-p_t\right). \tag{26}$$

Especially assuming a positive $(\lambda_t)_{t\geq 1}\subset(0,\infty)$, this recursion implies evolutionary credibility

$$\begin{aligned}\mathbb{E}\left[Y_{t+1}\middle|Y_{1:t}\right]&=\lambda_{t+1}\mathbb{E}\left[R_{t+1}\middle|Y_{1:t}\right]\\&=\lambda_{t+1}\left(m_{1,t}\frac{Y_t}{\lambda_t}+m_{2,t}\frac{\mathbb{E}\left[Y_t\middle|Y_{1:t-1}\right]}{\lambda_t}+m_{3,t}\cdot 1\right),\end{aligned}$$

where the weights $m_{1,t},m_{2,t},m_{3,t}\geq 0$ satisfy $m_{1,t}+m_{2,t}+m_{3,t}=1$ and are given by

$$m_{1,t}=p_tz_t,\quad m_{2,t}=p_t(1-z_t),\quad\text{and}\quad m_{3,t}=1-p_t. \tag{27}$$

By recursively expanding the result in (26), we arrive at the following corollary.

Corollary 1. *Under Model 1, we have*

$$\mathbb{E}\left[Y_{t+1}\middle|Y_{1:t}\right]=\lambda_{t+1}\left(\sum_{s=1}^{t}m_s^{[t]}\frac{Y_s}{\lambda_s}+m_0^{[t]}\cdot 1\right), \tag{28}$$

for $t\geq 0$, where an empty product is set equal to 1 and an empty sum equal to zero, and

$$m_s^{[t]}=\omega_{1,s}\prod_{k=s+1}^{t}\omega_{2,k}\quad\text{and}\quad m_0^{[t]}=\prod_{s=1}^{t}\omega_{2,s}+\sum_{s=1}^{t}\left(\prod_{k=s+1}^{t}\omega_{2,k}\right)\omega_{3,s}.$$

Here, $m_0^{[t]},\ldots,m_t^{[t]}\geq 0$ are proper weights satisfying the normalization $\sum_{s=0}^{t}m_s^{[t]}=1$.

Corollary 1 provides a credibility weighted average that combines the observations $Y_{1:t}$ with the prior mean $\mathbb{E}\left[\Theta_{t+1}\right]=1$. In particular, we would like to highlight that the weights $(m_s^{[t]})_{s=0}^{t}$ are deterministic and they solely depend on the exogenous sequences $(\lambda_t)_{t\geq 1}$, $(p_t)_{t\geq 1}$ and $(q_t)_{t\geq 1}$, and the initial parameter $\alpha_{1|0}=\beta_{1|0}>0$. Thus, from Corollary 1 we observe that past observations $Y_{1:t}$ enter the predictive mean of Y_{t+1} only linearly, i.e., (28) provides us with a *linear credibility formula* for pricing the next claim Y_{t+1}, given past claims experience $Y_{1:t}$.

The specific choices of the exogenous sequences can give rise to a seniority weighting. For example, assuming constant parameters $\lambda_t\equiv\lambda>0$, $p_t\equiv p\in(0,1)$, and $q_t\equiv q\in(0,1)$, the following relationship holds

$$\frac{m_s^{[t]}}{m_{s-1}^{[t]}}=\frac{1}{p}>1, \tag{29}$$

which demonstrates the presence of a seniority weighting. Furthermore, (29) indicates an exponential decay of the credibility weights, showing the relevance of Model 1 with the classical credibility model with geometric credibility factors; see Gerber and Jones (1973, 1975) and Sundt (1988).

Note that for all the models discussed in Table 1, the predictive formula share the same form outlined in Corollary 1. However, since models with different variance patterns have different specifications of $(p_t)_{t\geq 1}$ and $(q_t)_{t\geq 1}$, and the coefficients $(m_s^{[t]})_{s=0}^{t}$ in Corollary 1 depend on $(p_t)_{t\geq 1}$ and $(q_t)_{t\geq 1}$, these models will have different predictive formulas.

5. Real data analysis

Obviously, it would be exciting to present a high-frequency data example at this stage, see, e.g., Yanez et al. (2025) on weekly telematics data. Unfortunately, there is no such publicly available data in the insurance domain, therefore, we can only present a small scale example that has 6 observation periods. The different variance behaviors are still relevant, though we do not consider a long-term example, as they directly impact the credibility weights, see Corollary 1. From a modeling and interpretability perspective, an actuary would probably select a model with a non-increasing variance behavior for the individual insurance policies, otherwise the heterogeneity within the insurance portfolio would increase over time which, intuitively, would be difficult to justify, see discussion in Bühlmann and Gisler (2006), Section 9.5, page 235.

5.1. Description of data

We use the Local Government Property Insurance Fund (LGPIF) data from the US state of Wisconsin, which is a useful resource with insurance claim and individual policy information over time; see Frees et al. (2016). The portfolio used considers Building & Contents (BC) insurance contracts. After some data cleaning, our dataset consists of 1,229 insurance policies suffering 3,636 claims, observed longitudinally for the period of the calendar years 2006-2011.

We use the observations of calendar years 2006-2010 for model fitting, while the observations from calendar year 2011 are set aside for out-of-sample forecasting and model validation; there are 3,006 claims from 2006 to 2010. We illustrate an excerpt of the data in Table 2. `PolicyNum` is a unique identifier of each insurance policyholder which allows us to follow an insurance contract over the calendar years 2006 to 2011, and `Year` gives the corresponding calendar year. The insurance policies concern 6 different types of entities, indicated by the categorical variable `Type`, these are of institutional type and read as `City`, `County`, `Misc`, `School`, `Town`, `Village`. `Coverage` gives the logarithm of the covered amount (censored at 300 due to scarcity of data above), and `Deductible` is an insurance deductible which takes values between 500 and 100,000 (due to scarcity of large deductibles, we censor this variable at 25,000). Finally, `ClaimNb` gives the observed numbers of claims in each calendar year. Table 3 provides a brief summary statistics of the observed policy characteristics.

Table 2
Excerpt of the LGPIF building and contents (BC) data used.

PolicyNum	Year	Type	Coverage	Deductible	ClaimNb
integer	integer	categorical (6)	real	real	integer
120002	2006	County	22.7	1000	0
120002	2007	County	25.0	1000	0
120002	2008	County	20.9	1000	0
120002	2009	County	21.9	1000	0
120002	2010	County	23.5	1000	1

Table 3
Policy characteristics used as covariates over all years 2006-2011.

Categorical levels	Description		Proportions	
City	Level for city entity:		13.72%	
County	Level for county entity:		5.65%	
Misc	Level for miscellaneous entity:		11.01%	
School	Level for school entity:		28.24%	
Town	Level for town entity:		17.36%	
Village	Level for village entity:		24.01%	
Continuous variables		**Minimum**	**Mean**	**Maximum**
Coverage	Scaled covered log-amount	0.0035	30.97	300
Deductible	Deductible amount	500	2,893	25,000

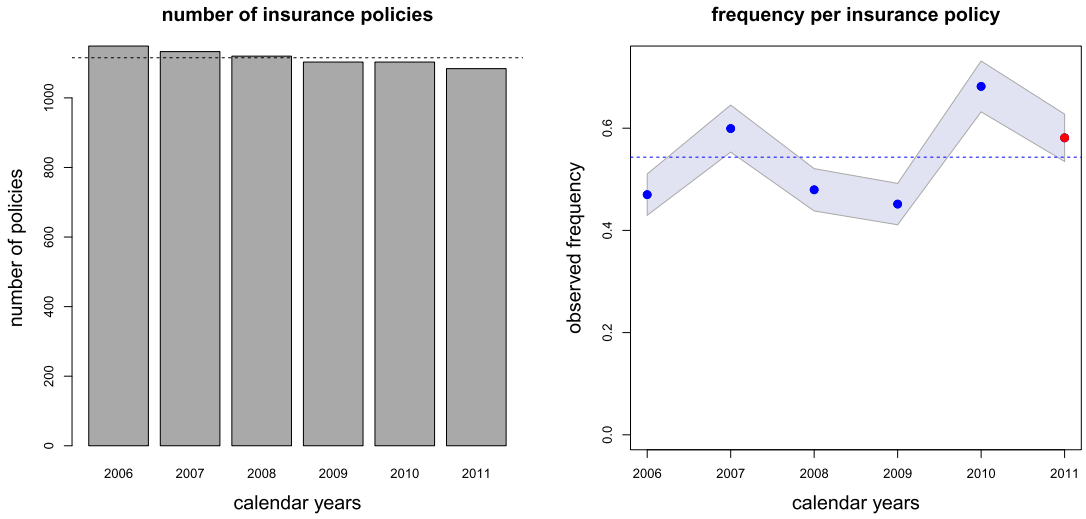


Fig. 1. (lhs) Observed numbers of annual insurance policies, and (rhs) observed annual claim frequencies. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

Fig. 1 (lhs) shows the observed numbers of annual insurance policies. The volume is slightly decreasing from 1,149 to 1,084 (over the years 2006-2011), i.e., we have a slight decrease of volume in this insurance portfolio. Fig. 1 (rhs) shows the observed (empirical) frequencies over the calendar years 2006-2011. We observe a fluctuating time-series; the shaded area shows (annual) confidence bounds of two standard deviations estimated under a Poisson model assumption. The observations underlying the blue dots (2006-2010) in Fig. 1 (rhs) are going to be used for model fitting, and the red dot (2011) collects the claims that we would like to forecast and price using experience rating.

Fig. 2 provides the empirical statistics for the different covariate levels of the variables Type, Coverage (discretized) and Deductible; the top row shows the exposure plots and the bottom row the corresponding observed frequencies. We see significant differences in these frequency plots. For the covariate Type, the levels County and City stand out with much bigger empirical frequencies. Empirical frequencies are increasing in Coverage, and we are going to model this covariate log-linearly. The picture of Deductible is more difficult to interpret. Naturally we would expect that the frequency is decreasing in an increasing deductible, however, to understand the empirical frequency plot of Deductibles one needs to understand its interaction with other variables, especially with the variable Type, which typically has higher deductibles for the level County.

5.2. Log-likelihood function under a GLM assumption

We are going to fit Model 1 to this data using different specifications for the parameters p_t and q_t. We choose these parameters time-invariant $p_t \equiv p$ and $q_t \equiv q$ to study some of the models summarized in Table 1. We denote by $1 \le i \le n = 1,229$ the different insurance policyholders, and we assume that their claims experiences are independent. Let ϑ collect all parameters that need to be estimated. From (24), it follows that the joint log-likelihood function is given by

$$\vartheta \mapsto \ell(\vartheta) := \sum_{i=1}^{n} \sum_{t=1}^{T_i} \log \left(\frac{\Gamma\left(Y_{i,t} + \alpha_{i,t|t-1}\right)}{Y_{i,t}!\,\Gamma\left(\alpha_{i,t|t-1}\right)} \left(\frac{\lambda_{i,t}}{\lambda_{i,t} + \beta_{i,t|t-1}} \right)^{Y_{i,t}} \times \left(\frac{\beta_{i,t|t-1}}{\lambda_{i,t} + \beta_{i,t|t-1}} \right)^{\alpha_{i,t|t-1}} \right), \tag{30}$$

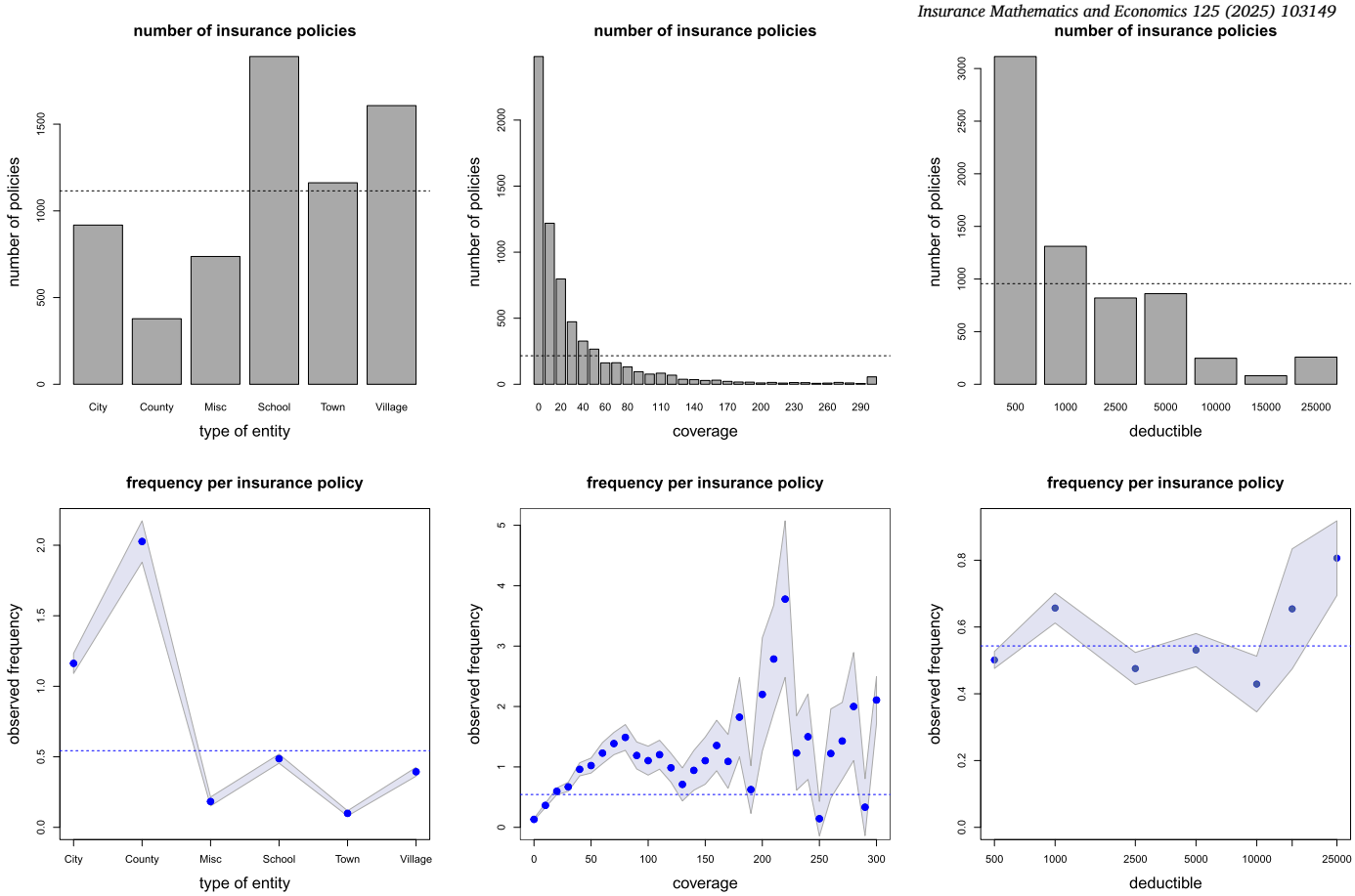


Fig. 2. The top shows the observed total exposures in the covariate levels of `Type` (lhs), `Coverage` (middle) and `Deductible` (rhs), and the bottom shows the observed frequencies in these covariate levels.

where we have to be careful with the time index $t = 1, \dots, T_i$. First, $t = 1$ is initialized individually to the inception dates S_i of the insurance contracts of the policyholders i, which can be any of the calendar years 2006-2011, and $T_i \in \{0, \dots, 5\}$ is the total number of observations within the calendar years 2006-2010; note that we exclude the latest year 2011, as this is going to be forecast by the fitted model.

Let $\boldsymbol{x}_{i,t} \in \mathbb{R}^d$ denote the covariates in calendar `Year` $= S_i + t - 1$ where, again, $S_i \in \{2006, \dots, 2011\}$ denotes the inception year of contract i. For our analysis, we choose the covariates `Type` (categorical with 6 levels and implemented by dummy coding), `Coverage` (continuous on log-scale), `Deductible` (continuous but implemented categorically). This results in a covariate of dimension $d = 12$. We then choose a model with log-link and GLM regression function

$$\boldsymbol{x}_{i,t} \in \mathbb{R}^d \;\mapsto\; \log \lambda_{i,t} = \log \lambda(\boldsymbol{x}_{i,t}) = b_0 + \sum_{j=1}^{d} b_j x_{i,j,t}, \tag{31}$$

with regression parameter $\boldsymbol{b} = (b_0, \dots, b_d)^\top \in \mathbb{R}^{d+1}$. In summary, there are $d + 1 = 13$ regression parameters $\boldsymbol{b} \in \mathbb{R}^{13}$, an initial value $\alpha_{1|0} = \beta_{1|0}$ and two parameters p and q to be estimated in Model 1 with time-invariant p and q choices. That is, we have a $d + 4 = 16$ dimensional parameter

$$\vartheta = \left(\boldsymbol{b}, \beta_{1|0}, p, q\right) = \left(b_0, \dots, b_d, \beta_{1|0}, p, q\right), \tag{32}$$

to be estimated in Model 1 with time-invariant parameters. These parameters are estimated with maximum likelihood estimation (MLE) which is doable because we have a fully tractable log-likelihood function (31).

Remark 1. In the above model, we have assumed that the sequences $p_t \equiv p$ and $q_t \equiv q$ are time-invariant. This allows us to directly estimated p and q with MLE, and extrapolation beyond the latest observed time point is not an issue. One may question this approach, and one may want to let these sequences be time-dependent. This allows us to still estimate p_t and q_t non-parametrically, but the issue with this non-parametric approach is that the extrapolation is not straightforward, but it requires further assumptions (or a parametric model). However, such a non-parametric approach can still be exploited to see whether a time-invariance assumption is justified, we come back to this in Section 5.6, below.

5.3. The classical NB-regression model

We start by considering the classical NB-regression model, this corresponds to the first case in Table 1 with $p_{i,t} \equiv 0$. Moreover, we choose

$$q_{i,t} = \beta_{1|0}/(\beta_{1|0} + \lambda_{i,t}) > 0, \tag{33}$$

implying $\beta_{i,t+1|t} = \beta_{1|0}$ for all $t \geq 1$ and $1 \leq i \leq n$, i.e., the random effects are assumed to be i.i.d. Gamma distributed with mean 1 and variance $1/\beta_{1|0}$. This implies for the variances of the responses

$$\mathrm{Var}\left(Y_{i,t}\right) = \lambda_{i,t} + \lambda_{i,t}^2/\beta_{1|0}.$$

This classical NB-regression model involves the parameters $\vartheta = (\boldsymbol{b}, \beta_{1|0})$, see (32), and we estimate these with MLE. We therefore use the R function `glm.nb` from the R package `MASS`; Venables and Ripley (2002).

The results are presented on the first line of Table 4. The first three columns show the estimated parameters: p is set to zero, and $q_{i,t} = \beta_{1|0}/(\beta_{1|0} + \lambda_{i,t})$ is adjusted so that we have i.i.d. Gamma random

Table 4
Regression results.

	Parameters			In-sample 2006-2010		Out-of-sample 2011		
	$\widehat{p}$	$\widehat{q}$	$\widehat{\beta}_{1\|0}$	AIC	$\ell_{2006:2010}$	ℓ_{2011}	RMSE	PDL
(1) Classical NB-regression	0	$q_{i,t}$	1.543	9,159.2	-4,565.6	-918.1	1.0359	1.0782
(2) Static random effects	1	1	2.216	8,976.0	-4,474.0	-874.1	0.9113	0.9255
(3) Harvey and Fernandes	1	0.84	2.624	8,969.1	-4,469.5	**-868.5**	**0.9049**	**0.9203**
(4) Bounded variance	0.79	0.84	1.470	**8,938.3**	-4,453.1	-869.7	0.9068	0.9264
(5) Variable $(p_t)_t$ and $(q_t)_t$	*0.77*	*0.82*	1.512	8,948.0	-4,450.0	na	na	na

effects $\Theta_{i,t}$. The in-sample figures show Akaike (1974)'s information criterion (AIC) and the in-sample log-likelihood $\ell_{2006:2010}(\widehat{\vartheta})$ of the years 2006-2010, evaluated in the MLE $\widehat{\vartheta}$, see (30).

The out-of-sample figures give the (out-of-sample) log-likelihood $\ell_{2011}(\widehat{\vartheta})$ (the MLE $\widehat{\vartheta}$ has only been estimated on data prior to 2011), the out-of-sample rooted mean squared error (RMSE) and the out-of-sample Poisson deviance loss (PDL) given by, respectively,

$$\ell_{2011} = \sum_{i\in\mathcal{T}} \log\left(\frac{\Gamma\left(Y_{i,T_i+1} + \widehat{\alpha}_{i,T_i+1|T_i}\right)}{Y_{i,T_i+1}!\,\Gamma\left(\widehat{\alpha}_{i,T_i+1|T_i}\right)} \left(\frac{\widehat{\lambda}_{i,T_i+1}}{\widehat{\lambda}_{i,T_i+1} + \widehat{\beta}_{i,T_i+1|T_i}} \right)^{Y_{i,T_i+1}} \times \left(\frac{\widehat{\beta}_{i,T_i+1|T_i}}{\widehat{\lambda}_{i,T_i+1} + \widehat{\beta}_{i,T_i+1|T_i}} \right)^{\widehat{\alpha}_{i,T_i+1|T_i}} \right),$$

$$\text{RMSE} = \sqrt{\frac{1}{|\mathcal{T}|} \sum_{i\in\mathcal{T}} \left(Y_{i,T_i+1} - \widehat{Y}_{i,T_i+1}\right)^2},$$

$$\text{PDL} = \frac{1}{|\mathcal{T}|} \sum_{i\in\mathcal{T}} 2\left(\widehat{Y}_{i,T_i+1} - Y_{i,T_i+1} - Y_{i,T_i+1} \log\left(\frac{\widehat{Y}_{i,T_i+1}}{Y_{i,T_i+1}} \right) \right),$$

where $\mathcal{T}$ collects all insurance policies that have an exposure in 2011, the quantities with hats $\widehat{\ }$ use the MLE $\widehat{\vartheta}$ (estimated on the years 2006-2010) for the unknown parameters, and $\widehat{Y}_{i,T_i+1}$ is the estimated predictor of Y_{i,T_i+1}. We use these results as a benchmark for the random effects models that induce dependence along the time axis.

5.4. Static random effects model

Next, we study the static random effects case with $p_{i,t} = q_{i,t} = 1$ and $\beta_{i,1|0} = \beta_{1|0}$ for all $t \geq 1$ and $1 \leq i \leq n$; this model has been studied, e.g., in Dionne and Vanasse (1989). We have log-likelihood function in the static random effects case

$$\ell(\vartheta) = \sum_{i=1}^{n} \log\left(\frac{\Gamma\left(\sum_{t=1}^{T_i} Y_{i,t} + \beta_{1|0}\right)}{\Gamma\left(\beta_{1|0}\right) \prod_{t=1}^{T_i} Y_{i,t}!} \left(\frac{\beta_{1|0}}{\sum_{t=1}^{T_i} \lambda_{i,t} + \beta_{1|0}} \right)^{\beta_{1|0}} \times \prod_{t=1}^{T_i} \left(\frac{\lambda_{i,t}}{\sum_{s=1}^{T_i} \lambda_{i,s} + \beta_{1|0}} \right)^{Y_{i,t}} \right). \tag{34}$$

The parameters can again be estimated with MLE, which is equivalent here to an empirical Bayes' estimation in a common random effects model with covariates.

We have implemented the log-likelihood (34) using the R package `Keras`, see Chollet et al. (2017). In particular, we bring the data into a tensor structure, and then apply a time-distributed GLM layer to this data tensor, see Chapter 8 of Wüthrich and Merz (2023) for time-distributed layers. The code is provided in Listing 1 in the appendix. The empirical Bayes' estimators of (34) are then obtained using gradient descent fitting; for more explanation we also refer to Remark 2, below.

The results are presented on the second line of Table 4. From these results we conclude that the static random effects model outperforms the NB-regression model in all considered loss statistics. This tells us that there is dependence between the claim counts in different calendar years, and we can learn from past experience to receive better forecasts, thus, experience rating provides more accurate prices than prior information pricing which only takes into account the covariates $\boldsymbol{x}_{i,t}$.

5.5. Harvey-Fernandes model and bounded variance model

We now consider Model 1 with time-invariant parameters $p_t \equiv p$ and $q_t \equiv q$. For the Harvey and Fernandes (1989) increasing variance model, we select $p = 1$ and $q \in (0, 1)$. For the observation-driven bounded variance model, we need to choose $p, q \in (0, 1)$, see Table 1. These parameters p and q are part of the model parameter ϑ, given in (32). This model parameter ϑ can be estimated with MLE by maximizing the log-likelihood function (30). The implementation of this log-likelihood function (30) of Model 1 is a bit sophisticated because the observations $Y_{i,1:t}$ and the covariates $\boldsymbol{x}_{i,1:t}$ enter the parameters $\alpha_{i,t|t-1}$ and $\beta_{i,t|t-1}$ recursively, see Model 1. We have implemented this in the R package `Keras` [Chollet et al. (2017)] using a tailor-made recurrent layer, the code is given in Listing 2 in the appendix. This recurrent layer takes as inputs $(q_{1:t}, \lambda_{1:t}, p_{1:t}, Y_{1:t})$, and it outputs the sequence $(\alpha_{s|s-1}, \beta_{s|s-1})_{s=1}^{t}$. Based on this recurrent layer, the log-likelihood (30) can be implemented using again a time-distributed GLM layer for the regression function (31), see Listing 3 in the appendix. The MLE $\widehat{\vartheta}$ of model parameter (32) is then found numerically by applying the gradient descent algorithm to the negative log-likelihood $-\ell(\vartheta)$.

The estimation results are presented on lines (3) and (4) of Table 4 for the Harvey and Fernandes (1989) model and the observation-driven bounded variance model $p, q \in (0, 1)$, respectively. We observe that both of these models outperform the classical NB model and the static random effects model. Comparing the two observation-driven models on lines (3) and (4), the situation becomes more difficult. First, based on the in-sample figures we give preference to the bounded variance model, this preference is w.r.t. the lower AIC value (in bold face), which adjusts for the number of parameters involved. Out-of-sample (2011), the Harvey and Fernandes (1989) model with an increasing variance has a better performance in the corresponding loss figures (in bold face). Of course, from a practical point of view, this is an unpleasant situation that we would like to discuss and understand a better:

- In-sample, we give preference to the observation-driven bounded variance model with estimated parameters $\widehat{p} = 0.78$ and $\widehat{q} = 0.84$. These parameters are clearly bounded away from zero and one, which supports the bounded variance case. Theoretically, one could verify this null hypothesis with a likelihood ratio or a Wald test. Unfortunately, practically, this is not feasible as the corresponding Hessian is intractable. Therefore, we perform model selection using AIC, which is also a valid in-sample model selection criterion here (as we work with MLE estimated models). In a typically insurance pricing situation, this is the available information for pricing the next calendar year, and this determines the selection of the pricing model.
- Out-of-sample, we observe that the Harvey and Fernandes (1989) model with parameters $p = 1$ and $\widehat{q} = 0.84$ has a superior performance. Unfortunately, this information is *not* available for pricing the next calendar year 2011, i.e., we do not select the best possi-

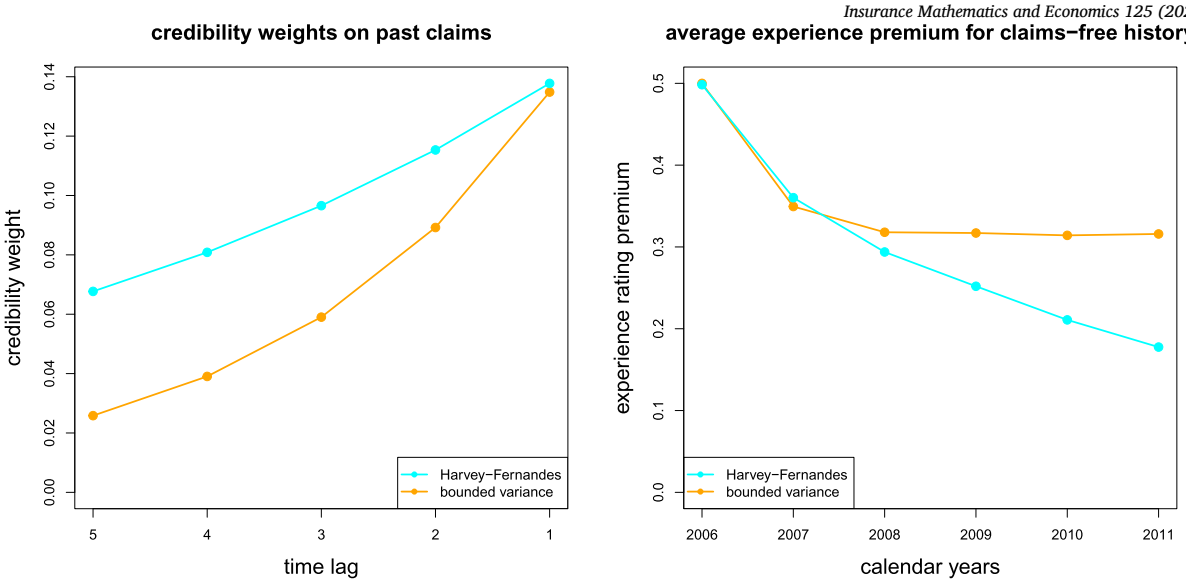


Fig. 3. (lhs) Weights on past observations $Y_{i,1}, \ldots, Y_{i,T_i}$ to predict Y_{i,T_i+1}, and (rhs) evolution of experience premium $\mathbb{E}\left[Y_{i,t+1}|Y_{i,1} = \ldots = Y_{i,t} = 0\right]$ for $t \geq 0$.

ble model in this case if only the information at the end of calendar year 2010 is available. In fact, this out-of-sample validation is sometimes called an *out-of-time* validation which may also be impacted by non-stationarity of the process. By trying to understand this result, we should also question the mean stationarity $\mathbb{E}\left[\Theta_t\right] = 1$, $t \geq 1$. In view of the empirical average frequencies, see Fig. 1, it is not clear whether mean stationarity holds in this example, furthermore, there may also exist global (annual) risk factors that simultaneously affect multiple insurance policyholders, violating the independence assumption between policyholders. The latter would require a non-trivial change of the model that will drop out of the presented modeling framework.

This fitted Harvey and Fernandes (1989) and bounded variance models allow us to study how past claims $Y_{i,1}, \ldots, Y_{i,T_i}$ are weighted (contribute) to predict Y_{i,T_i+1}; see Corollary 1. The resulting predictive mean $\mathbb{E}\left[Y_{i,T_i+1}|Y_{i,1:T_i}\right]$ corresponds to the experience rating premium that is charged to the corresponding insurance policyholder i. Fig. 3 (lhs) shows the contributions of past claims $Y_{i,s}$, $1 \leq s \leq T_i$, as a function of time lag to the next claim Y_{i,T_i+1} to be predicted; the plot is an average over the available portfolio. The cyan graph shows the results of the fitted Harvey and Fernandes (1989) model and the orange graph of the bounded variance model, given in Table 4 on lines (3) and (4). We observe that past claims have a seniority (fading memory property) that is fading faster in the bounded variance case in the parametrization of Table 4, the most recent claim Y_{i,T_i} receiving a weight of roughly 13.1% in both cases. Fig. 3 (rhs) shows the evolution of the experience rating prices $\mathbb{E}\left[Y_{i,t+1}|Y_{i,1} = \ldots = Y_{i,t} = 0\right]$ for $t \geq 0$, i.e., this plot shows how the premium decays for an increasing claims history being free of accidents. If there is no history available the price is 0.53, and after the first year of not having an accident, $Y_{i,1} = 0$, it decays to 0.37 (decay of 30%). This behavior coincides in both models, however, the further development differs between the two models. The bounded variance case stays roughly constant at an experience rating price of 0.32, whereas in the Harvey and Fernandes (1989) model it further decreases for an accident-free claims history. It is now up to the pricing actuary which pricing mechanism she/he prefers, this will depend on many factors, e.g., the line of business involved, further testing, hopefully richer data, etc. The cyan pricing scheme in Fig. 3 (rhs) is more frequently used for car insurance pricing, whereas private property and liability insurance leads to no-claims-discounts that rather have the shape of the orange curve in Fig. 3 (rhs). This closes the example.

Remark 2. The log-likelihood optimizations in the static random effects case (34) and in the bounded variance case (30) have both been implemented in `Keras`, based on TensorFlow, though we do not use a neural network regression model, but only a GLM for $\lambda_{i,t}$, see (31). The former model could be fitted with the iterated reweighted least squares (IRLS) algorithm, see Nelder and Wedderburn (1972); McCullagh and Nelder (1989), but computing the score equations and the corresponding Hessian is a bit cumbersome. The latter bounded variance case is more complicated because of the recursive structure of the parameters $(\alpha_{s|s-1}, \beta_{s|s-1})_{s=1}^{t}$. A recurrent neural network layer is a straightforward and convenient solution to this optimization problem, see Listing 2. The fact, that $\lambda_{i,t}$ uses a time-invariant GLM (31) is then implemented by a time-distributed GLM layer. Using the code of Listings 1 and 3, it is straightforward to lift these models to neural network regressions. This only requires that the time-distributed GLM layers on line 13 of Listing 1 and line 36 of Listing 3, respectively, are replaced by time-distributed neural network layers. We did not study such a neural network extension because our dataset is too small to fit a neural network model (we have only 1,229 insurance policyholders).

5.6. Time-dependent observation-driven bounded variance model

A straightforward extension of the observation-driven bounded variance model in the previous section is to let the parameters $(p_t)_{t\geq 1}$ and $(q_t)_{t\geq 1}$ be time-dependent. The implementation only needs a small adaption of the code in Listing 3, see Listing 4; the main change is that we equip the sequences with a position index, `Time` $= t - 1$, to include a time-varying modeling of $(p_t)_{t\geq 1}$ and $(q_t)_{t\geq 1}$. The remaining parts of the model architecture and fitting are identical to the models discussed above. Line (5) of Table 4 presents the results of the time-varying bounded variance model. The italic p- and q-values are the averages of the estimated values $(\widehat{p}_t)_{t=1}^{5}$ and $(\widehat{q}_t)_{t=1}^{5}$.

The estimates $(\widehat{p}_t)_{t=1}^{5}$ and $(\widehat{q}_t)_{t=1}^{5}$ are plotted in Fig. 4. The graphs in Fig. 4 show that these estimates fluctuate, i.e., they do not look like constants. However, the resulting AIC value in this time-varying bounded variance model case does not support this additional model flexibility, see Table 4, i.e., AIC suggests to select constant parameters $p_t \equiv p$ and $q_t \equiv q$.

Another deficiency of this non-parametric modeling approach of the sequences $(p_t)_{t\geq 1}$ and $(q_t)_{t\geq 1}$ is that we cannot easily extrapolate the estimates $(\widehat{p}_t)_{t=1}^{5}$ and $(\widehat{q}_t)_{t=1}^{5}$ to the next calendar year $\widehat{p}_6$ and $\widehat{q}_6$. This is the reason for not giving an out-of-sample (out-of-time) analysis in this

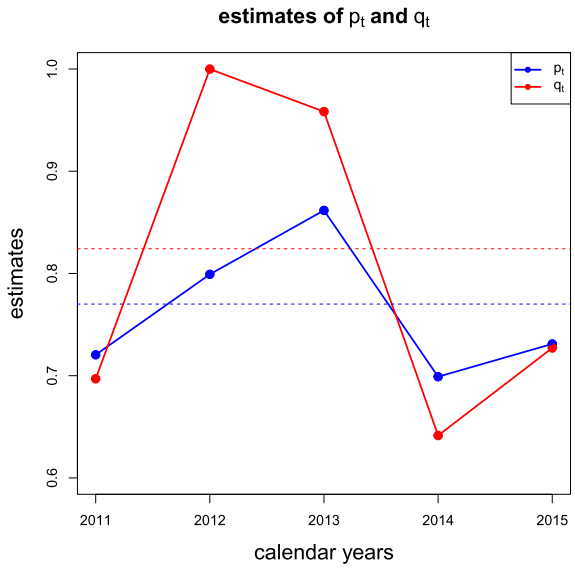


Fig. 4. Estimates $(\widehat{p}_t)_{t=1}^5$ and $(\widehat{q}_t)_{t=1}^5$ in the time-varying bounded variance model; the horizontal dotted lines give the averages of the two sequences.

time-varying case, see Table 4. An extrapolation would need further assumptions, e.g., in the form of a parametric model, if longer time-series would be available, Fig. 4 may for instance indicate business cycles. This closes our example.

6. Conclusion

In this paper, we expanded the observation-driven state-space model of Harvey and Fernandes (1989) to a broader spectrum of specifications characterized by various variance process behaviors. They are suitable for count processes with a constant mean, but with an increasing, decreasing, constant, converging, or bounded variance process. These models inherit most of the major advantages of state-space models, but they are more tractable for regression modeling than their parameter-driven counterparts. In particular, these models have closed-form log-likelihoods that allow for empirical Bayes' estimation, they allow for closed-form experience rating prices, and our real data example has provided a seniority weighting to past claims in these experience rating prices.

Our example has been based on a GLM assumption for the regression function and fitting was done by MLE using a gradient descent algorithm. This GLM regression function can easily be replaced by other regression functions such as neural networks. Such an extension is not complicated, only model fitting will be more time consuming.

For two reasons, we did not study the long-term behavior of our model, such as ergodicity. First, usually (on an annual scale) we often do not have a longer history than 20 past observations. Second, to study long-term behavior such as ergodicity, one needs to restrict to stationary covariates; see, e.g., Debaly and Truquet (2021). Though mathematically interesting, we did not study this because it seems less relevant in insurance applications. In insurance applications, policyholder information is naturally aging, e.g., the age of policyholder increases every year by 1, making stationarity in covariates not a realistic assumption.

CRediT authorship contribution statement

Jae Youn Ahn: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Himchan Jeong:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Yang Lu:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Mario V. Wüthrich:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Additional variance behavior types

A.1. A model with converging variance

In this subsection, we discuss a special case of Model 1 which is asymptotically strongly stationary. By *strongly* stationarity, we mean that the conditional distribution of Y_{t+h}, given Y_t, converges to a non-degenerate distribution. In other words, in the long-run, the process $(Y_t)_{t\geq 1}$ evolves in a "steady state", i.e., in an "equilibrium". This implies, in particular, that the variance and the mean of the process converge to positive constants.[2] Note also that stationarity does not mean "time-invariant". Indeed, for large time horizons h, the variance of Y_{t+h} converges to a positive constant.

Specifically, we assume, in Model 1, that p_t, q_t and λ_t are all time-invariant for $t \geq 1$ with

$$p_t = p \in (0,1), \qquad q_t = q \in (0,1), \tag{A.1}$$

and

$$\lambda_t = \lambda > 0. \tag{A.2}$$

Then, we have for $t \geq 1$

$$\begin{aligned} \alpha_t &= \alpha_{t|t-1} + Y_t = pq\alpha_{t-1} + (1-p)\,q\beta_{t-1} + Y_t, \\ \beta_t &= \beta_{t|t-1} + \lambda_t = q\beta_{t-1} + \lambda, \end{aligned} \tag{A.3}$$

where the second identities require $t \geq 2$.

Because our model has a NB distribution as predictive distribution, with the number of trial parameter satisfying a linear recursion, it is useful to compare this model with the literature on NB-INGARCH models. Most of these papers assume that the conditional distribution of Y_t given $Y_{1:t-1}$ is NB, and there are two different approaches for this conditional distribution[3]:

- The first approach [Zhu (2011); Christou and Fokianos (2014); Davis and Liu (2016); Ahmad and Francq (2016)] assumes that the conditional NB distribution at time t has a time-invariant size parameter r, and a time-varying probability parameter ϕ_t. Moreover, the transformed parameter $\frac{1-\phi_t}{\phi_t}$ (or equivalently the conditional expectation) is assumed to satisfy a linear recursion.
- The second approach assumes that the conditional NB distribution has a time-varying size parameter r_t, and a time-invariant probability parameter ϕ. Moreover, the sequence of size parameter (or equivalently the sequence of conditional mean parameter) is updated by a linear recursion. Such a model is, for instance, considered in Aknouche and Francq (2021), Section 2.4, who assume a NB-INGARCH(p,q) specification, and also in Xu et al. (2012) and Gouriéroux and Lu (2019), who assume a finite order NB-INARCH(p) specification.

[2] However, the strong stationarity is stronger than the second order convergence of the variance and mean of the process. Indeed, the stationarity implies the convergence of any moments, whenever they exist.

[3] For the following discussion we use the parametrization of the NB distribution with probability parameter ϕ and size parameter r so that if $N \sim \text{NB}(\phi, r)$ then $\mathbb{E}[N] = \frac{r(1-\phi)}{\phi}$ and $\text{Var}[N] = \frac{r(1-\phi)}{\phi^2}$.

Both of these approaches enter into the general framework of Gonçalves et al. (2015); see their Observation 2, page 505, for the discussion of the above two different approaches.

In our model under conditions (A.1) and (A.2), the predictive mean, denoted by

$$m_t := \mathbb{E}\left[Y_t \mid Y_{1:t-1}\right] = \lambda \frac{\alpha_{t|t-1}}{\beta_{t|t-1}},$$

satisfies the recursion

$$\frac{\beta_{t|t-1}}{\lambda} m_t = \alpha_{t|t-1} = pq \frac{\beta_{t-1|t-2}}{\lambda} m_{t-1} + pqY_{t-1} + (1-p)q,$$

or equivalently

$$m_t = pq \frac{\beta_{t-1|t-2}}{\beta_{t|t-1}} m_{t-1} + \frac{\lambda pq}{\beta_{t|t-1}} Y_{t-1} + \lambda(1-p)q/\beta_{t|t-1}. \tag{A.4}$$

This is also a linear type recursion for the sequence $(m_t)_{t\geq 1}$, similar as (1.1) of Aknouche and Francq (2021). When t goes to infinity, $\beta_{t|t-1} = q\beta_t$ goes to $\lambda q/(1-q)$, and the ratio $\beta_{t-1|t-2}/\beta_{t|t-1}$ converges to 1. Thus, the two coefficients in the above linear recursion (A.4) converge to pq and $p(1-q)$, respectively. Then, we can apply Aknouche and Francq (2021), Corollary 3.1 or Gonçalves et al. (2015), Theorem 3,[4] which state that the sum of these two coefficients needs to be smaller than 1 for achieving strong stationarity. As a consequence we get the following statement.

Lemma 6. *If in Model 1 the parameters satisfy (A.1) and (A.2), then the process $(Y_t)_{t\geq 1}$ is asymptotically strongly stationary. In particular, $\left(\mathrm{Var}\left(\Theta_t\right)\right)_{t\geq 1}$ and $\left(\mathrm{Var}\left(Y_t\right)\right)_{t\geq 1}$ converge.*

Under assumptions (A.1) and (A.2), the predictive mean m_{t+1} is once again an EWMA of the past observations $Y_{1:t}$. In other words, it is an extension of the standard EWMA forecasting literature; see Hyndman et al. (2008).

Remark 3. There are also other stationary count process models allowing for EWMA predictive mean formulas, such as the INGARCH(1,1), see Ferland et al. (2006), or the NB-INGARCH, see Zhu (2011). These other models, however, do not admit a state-space representation and, therefore, do not possess the three properties mentioned in the introduction. In other words, among all the stationary count models with an EWMA predictive mean, the model of Aknouche and Francq (2021) and Gonçalves et al. (2015) has the advantage of having a state-space representation.

A.2. A model with bounded variance

For many applications, the assumption of a time-invariant $(\lambda_t)_{t\geq 1}$ imposed in the previous subsections is too restrictive because insurance policies are aging which naturally leads to an evolution in regression function (if they consider time). In this subsection, we relax this assumption, and consider a model satisfying (A.1) only, but not (A.2). We get the following result.

[4] This theorem requires the process to satisfy assumption H2 in Section 3.2 of Gonçalves et al. (2015), which states that the ratio between the conditional variance and the conditional expectation should be a linear function of the conditional expectation. In our model, this ratio is equal to $1 + \frac{\lambda}{\beta_{t|t-1}}$, where λ is constant and $\beta_{t|t-1}$ satisfies the recursion $\beta_{t+1|t} = q(\beta_{t|t-1} + \lambda)$, and thus converges to the constant $\frac{\lambda}{1-q}$. As a consequence, the ratio between the conditional variance and conditional mean converges to a constant as t increases to infinity. In other words, Condition H2 is asymptotically satisfied.

Lemma 7. *If in Model 1, (A.1) is satisfied, and if the process $(\lambda_t)_{t\geq 1}$ is uniformly bounded from both above and below by positive constants, then the variance process $\left(\mathrm{Var}\left(Y_t\right)\right)_{t\geq 1}$ is bounded from above.*

Proof. See Appendix B.4. □

Appendix B. Proofs

B.1. Proof of Lemma 2

Part i) We proceed by induction. From (4), we get $\mathbb{E}\left[\Theta_1\right] = 1$ and $\mathbb{E}\left[Y_1\right] = \lambda_1$. Hence, $\mathbb{E}\left[\alpha_1\right] = \beta_1$. Assume that $\mathbb{E}\left[\Theta_t\right] = 1$ and $\mathbb{E}\left[\alpha_t\right] = \beta_t$ for some $t \geq 1$. First, we have $\mathbb{E}\left[\Theta_{t+1}\right] = 1$ from the following consideration, in view of (6),

$$\mathbb{E}\left[\Theta_{t+1}\right] = \mathbb{E}\left[\mathbb{E}\left[\Theta_{t+1} \mid Y_{1:t}\right]\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{\alpha_{t+1|t}}{\beta_{t+1|t}} \middle| Y_{1:t}\right]\right]$$
$$= \mathbb{E}\left[\frac{p_t q_t \alpha_t + \left(1-p_t\right) q_t \beta_t}{q_t \beta_t}\right] = p_t \mathbb{E}\left[\frac{\alpha_t}{\beta_t}\right] + (1-p_t) = 1.$$

This also implies $\mathbb{E}\left[Y_t\right] = \lambda_t$ for all $t \geq 1$. Then, from the updates in (8) and (9) along with the induction assumption, we have for $t \geq 1$

$$\mathbb{E}\left[\alpha_{t+1}\right] = \mathbb{E}\left[p_t q_t \alpha_t + \left(1-p_t\right) q_t \beta_t + Y_{t+1}\right]$$
$$= p_t q_t \mathbb{E}\left[\alpha_t\right] + \left(1-p_t\right) q_t \beta_t + \mathbb{E}\left[Y_{t+1}\right]$$
$$= q_t \beta_t + \lambda_{t+1} = \beta_{t+1}.$$

Part ii) This is a direct consequence of Part *i*).

Part iii) We have

$$\mathbb{E}\left[\mathrm{Var}\left(\Theta_{t+1} \mid Y_{1:t}\right)\right] = \frac{1}{\beta_{t+1|t}} \mathbb{E}\left[\mathbb{E}\left[\Theta_{t+1} \mid Y_{1:t}\right]\right] = \frac{1}{q_t \beta_t},$$

where the first equation follows from (6).

Part iv) (6) implies

$$\mathbb{E}\left[\Theta_{t+1} \mid Y_{1:t}\right] = \frac{\alpha_{t+1|t}}{\beta_{t+1|t}} = p_t \frac{\alpha_t}{\beta_t} + (1-p_t),$$

which implies

$$\mathrm{Var}\left(\mathbb{E}\left[\Theta_{t+1} \mid Y_{1:t}\right]\right) = p_t^2 \,\mathrm{Var}\left(\frac{\alpha_t}{\beta_t}\right) = p_t^2 \,\mathrm{Var}\left(\mathbb{E}\left[\Theta_t \mid Y_{1:t}\right]\right).$$

This proves Lemma 2.

B.2. Proof of Lemma 3

By comparing equations (16) and (17), and using $p_t \equiv 1$ and $q_t \equiv q \in (0,1)$, we have

$$\mathrm{Var}\left(\Theta_{t+1}\right) - \mathrm{Var}\left(\Theta_t\right) = \left(\frac{1}{q} - 1\right) \frac{1}{\beta_t} > 0. \tag{B.1}$$

If the exogenous process $(\lambda_t)_{t\geq 1}$ is bounded from above by $M > 0$, then β_t, which satisfies the recursion $\beta_t = q\beta_{t-1} + \lambda_t$, is bounded from above by $M/(1-q)$. Hence, $1/\beta_t \geq (1-q)/M > 0$ is bounded from below. Therefore, the sequence of variances $\left(\mathrm{Var}\left(\Theta_t\right)\right)_{t\geq 1}$ increases to infinity when t goes to infinity, because we add uniformly lower bounded constants to the variances by increasing t, see (B.1).

Furthermore, if λ_t is bounded from below by $m > 0$, then by the total variance decomposition formula, $\mathrm{Var}\left(Y_t\right)$ increases to infinity as well as $t \to \infty$.

B.3. Proof of Lemma 4

Because the sequence of variances $(\mathrm{Var}\left(\Theta_t\right))_{t\geq 1}$ is decreasing but positive, it converges to a non-negative constant. In other words, the left-hand side of equation (19) goes to zero as $t \to \infty$. As a consequence, the right-hand side of (19) also goes to zero. Thus, $\mathrm{Var}\left(\frac{\alpha_t}{\beta_t}\right) = \frac{1}{\beta_t^2}\mathrm{Var}\left(\alpha_t\right)$ converges to zero. Hence, the second term in (17) goes to zero when t goes to infinity. As for the first term in (17), we have that $\mathbb{E}\left[\frac{\alpha_t}{\beta_t^2}\right] = \frac{1}{\beta_t^2}\mathbb{E}\left[\alpha_t\right] = \frac{1}{\beta_t^2}\beta_t$ goes to zero as well since β_t goes to infinity as t goes to infinity. As a consequence, we deduce that $(\mathrm{Var}\left(\Theta_t\right))_{t\geq 1}$ converges to zero when t goes to infinity.

Finally, by the total variance decomposition we have

$$\mathrm{Var}\left(Y_t\right) = \mathbb{E}\Big[\mathrm{Var}\left(Y_t|\Theta_t\right)\Big] + \mathrm{Var}\left(\mathbb{E}[Y_t|\Theta_t]\right) = \lambda_t\mathbb{E}[\Theta_t] + \lambda_t^2\mathrm{Var}\left(\Theta_t\right). \tag{B.2}$$

Under the additional assumption (A.2), the latter converges to $\lambda\mathbb{E}[\Theta_t] = \lambda\mathbb{E}[\Theta_1] = \lambda$ as t goes to infinity.

B.4. Proof of Lemma 7

First, since $(\lambda_t)_{t\geq 1}$ is bounded, $(\beta_t)_{t\geq 1}$ is also bounded from above and below by positive constants. Then, by (B.2) and applying the total variance decomposition formula once more

$$\mathrm{Var}\left(Y_t\right) = \lambda_t\mathbb{E}[\Theta_t] + \lambda_t^2\Big(\mathrm{Var}\left(\alpha_t/\beta_t\right) + \mathbb{E}[\alpha_t/\beta_t^2]\Big).$$

Thus it suffices to show that $\mathrm{Var}\left(\alpha_t\right)$ is bounded. We have

$$\begin{aligned}
\mathrm{Var}\left(\alpha_t\right) &= \mathrm{Var}\left(pq\alpha_{t-1} + \beta_{t-1}q(1-p) + Y_t\right)\\
&= \mathrm{Var}\left(pq\alpha_{t-1} + Y_t\right)\\
&= \mathbb{E}\Big[\mathrm{Var}\left(pq\alpha_{t-1} + Y_t|Y_{1:(t-1)}\right)\Big] + \mathrm{Var}\Big(\mathbb{E}[pq\alpha_{t-1} + Y_t|Y_{1:(t-1)}]\Big)\\
&= \mathbb{E}\Big[\mathrm{Var}\left(Y_t|Y_{1:(t-1)}\right)\Big] + \mathrm{Var}\Big(pq\alpha_{t-1} + \mathbb{E}[Y_t|Y_{1:(t-1)}]\Big) \qquad \text{(B.3)}\\
&= \mathbb{E}\left[\lambda_t\frac{\alpha_{t|t-1}}{\beta_{t|t-1}} + \lambda_t^2\frac{\alpha_{t|t-1}}{\beta_{t|t-1}^2}\right] + \mathrm{Var}\left(pq\alpha_{t-1} + \lambda_t\frac{\alpha_{t|t-1}}{\beta_{t|t-1}}\right), \qquad \text{(B.4)}\\
&= \left(\frac{\lambda_t}{q\beta_{t-1}} + \frac{\lambda_t^2}{q^2\beta_{t-1}^2}\right)\mathbb{E}\left[pq\alpha_{t-1} + (1-p)q\beta_{t-1}\right]\\
&\quad + \mathrm{Var}\left(pq\alpha_{t-1} + \lambda_t\frac{pq\alpha_{t-1} + (1-p)q\beta_{t-1}}{q\beta_{t-1}}\right),
\end{aligned}$$

where from (B.3) to (B.4) we have used the mean and variance formula of the conditional distribution of Y_t, given $Y_{1:(t-1)}$, which is NB.

The first term on the right-hand side is uniformly upper bounded, since $\mathbb{E}[\alpha_{t-1}] = \beta_{t-1}$ is uniformly bounded. The last term can be written as, note β_{t-1} is deterministic,

$$\mathrm{Var}\Big(\alpha_{t-1}(pq + \lambda_t p/\beta_{t-1})\Big) = \mathrm{Var}\Big(p\frac{q\beta_{t-1} + \lambda_t}{\beta_{t-1}}\alpha_{t-1}\Big) = \mathrm{Var}\left(p\frac{\beta_t}{\beta_{t-1}}\alpha_{t-1}\right).$$

Thus, by dividing both sides of (B.4) by β_t^2, which is both upper and lower bounded by positive constants, we get

$$\mathrm{Var}\left(\frac{\alpha_t}{\beta_t}\right) = p^2\mathrm{Var}\left(\frac{\alpha_{t-1}}{\beta_{t-1}}\right) + \text{uniformly bounded term.}$$

By recurrence, we deduce that $\mathrm{Var}\left(\frac{\alpha_t}{\beta_t}\right)$ is uniformly upper bounded. Therefore, $\mathrm{Var}\left(\alpha_t\right)$ is also upper bounded. Hence, $\mathrm{Var}\left(Y_t\right)$ is also uniformly upper bounded.

Appendix C. Weak ARMA(1,1) representation of (Y_t) in the time homogeneous case

In this section, we show that the autocorrelation function of the process $(Y_t)_{t\geq 1}$, when it is well defined (i.e., time-invariant), is of ARMA(1,1) type. To derive the autocorrelation, one has to ensure that the process $(Y_t)_{t\geq 1}$ is (asymptotically) stationary, that is, p_t, q_t, λ_t are all constant. In the following, we focus on this case.

Let us first recall that in a linear Kalman filter

$$W_t = \rho W_{t-1} + \epsilon_t, \tag{C.1}$$

$$Z_t = \beta W_t + \eta_t, \tag{C.2}$$

where the bivariate process $(\epsilon_t, \eta_t)_{t\geq 1}$ is uncorrelated with its past, $(W_t)_{t\geq 1}$ is the latent process (also called the state variable), $(Z_t)_{t\geq 1}$ is the observable process (also called the measurement variable), ρ, β are parameters such that $\rho \in (-1, 1)$. Then we can show that $(Z_t)_{t\geq 1}$ is a (weak) ARMA(1,1) process since

$$Z_t - \rho Z_{t-1} = \beta\epsilon_t + \eta_t - \rho\eta_{t-1},$$

and the right hand side $\beta\epsilon_t + \eta_t - \rho\eta_{t-1}$ is an MA(1) process since its autocorrelation terms are zero beyond lag 1.

In our paper, the count process $(Y_t)_{t\geq 1}$ with constant p_t, q_t, λ_t has also a (weak) ARMA(1,1), since:

- first, by equation (15) of the paper, we have

  $$\mathbb{E}[\Theta_{t+1} \mid \Theta_t, Y_{1:t}] = pq\Theta_t + (1-q),$$

  thus, the latent variable $(\Theta_t)_{t\geq 1}$ has the same AR(1) form as in equation (C.1) above, with ϵ_t having a mean of $(1-q)$;
- second, since $Y_{t+1} \mid Y_{1:t}, \Theta_{1:(t+1)}$ follows Poisson distribution, we have

  $$\mathbb{E}[Y_{t+1} \mid Y_{1:t}, \Theta_{1:(t+1)}] = \lambda\Theta_{t+1}.$$

  Thus, $(Y_t)_{t\geq 1}$ satisfies equation (C.2) above with η_t having a zero mean.

Appendix D. Code

Listing 1: Implementation of the static random effects model.

```
library(keras)
library(tensorflow)
########
GLM <- function(seed, lookback, q0, output.acti){
    k_clear_session()
    set.seed(seed)
    set_random_seed(seed)
    Design    <- layer_input(shape = c(lookback, q0[1]), dtype = 'float32')
    Exposure <- layer_input(shape = c(lookback), dtype = 'float32')
    Ones      <- k_ones_like(x=Exposure[,1]) %>% layer_flatten()
    #
    Network = Design %>%
          time_distributed(layer_dense(units=1, activation=output.acti)) %>%
```

```
          layer_flatten()
    # estimates the shape parameter
    Beta = Ones %>% layer_dense(units=1, activation="exponential",
                                use_bias=FALSE, weights=list(array(0, dim=c(1,1))))
    # estimates the mean
    GLM = list(Network, Exposure) %>% layer_multiply()
    #
    Response = list(GLM, Beta) %>% layer_concatenate()
    #
    keras_model(inputs = c(Design, Exposure), outputs = c(Response))
    }
```

Listing 2: Recurrent layer for Model 1 fitting.

```
library(keras)
library(tensorflow)
########
# This code is based on "Working with RNNs", 2022-12-16
# Authors Scott Zhu, Francois Chollet, Tomasz Kalinowski
# https://tensorflow.rstudio.com/guides/keras/working_with_rnns
########
BetaAlphaCell(keras$layers$Layer) %py_class% {
  initialize <- function(unit_1, unit_2, ...) {
    self$unit_1      <- unit_1
    self$unit_2      <- unit_2
    self$state_size  <- list(shape(unit_1), shape(unit_2))
    self$output_size <- list(shape(unit_1), shape(unit_2))
    super$initialize(...)
  }
  call <- function(inputs, states) {
    c(input_1, input_2, input_3, input_4) %<-% tf$nest$flatten(inputs)
    c(s1, s2) %<-% states
    # beta sequence
    output_1   <- input_1 * (s1 + input_2)
    state_1    <- output_1
    # alpha sequence
    output_2   <- input_3*input_1*(s2 + input_4) + (1-input_3)*output_1
    state_2    <- output_2
    #
    output     <- tuple(output_1, output_2)
    new_states <- tuple(state_1, state_2)
    tuple(output, new_states)
  }
  get_config <- function(){list("unit_1" = self$unit_1, "unit_2" = self$unit_2)}
}
########
cell      <- BetaAlphaCell(1, 1)
RNN       <- layer_rnn(cell = cell, return_sequence=TRUE)
BetaAlpha <- RNN(tuple(q, mu, p, N))
```

Listing 3: Implementation of Model 1 with the regression setting.

```
library(keras)
library(tensorflow)
########
GLM.dynamic <- function(seed, q0, init0){
    k_clear_session()
    set.seed(seed)
    set_random_seed(seed)
    YY      <- layer_input(shape = c(q0[1]), dtype = 'float32')
    XX      <- layer_input(shape = c(q0[1],q0[2]), dtype = 'float32')
    Obs     <- layer_input(shape = c(q0[1]), dtype = 'float32')
    Ones    <- k_ones_like(x=Obs[,1]) %>% layer_flatten()
    Zeros   <- k_zeros_like(x=Obs[,1]) %>% layer_flatten()

    # alpha10 = beta10 initialization
    alpha1  = Ones %>% layer_dense(units=1, use_bias=FALSE, activation="exponential",
                                   weights=list(array(init0[2], dim=c(1,1))))
    YY1     = list(alpha1, YY) %>% layer_concatenate() %>%
                                  layer_reshape(target_shape=c(q0[1]+1,1))
    #
    pp      = Zeros %>% layer_embedding(input_dim = 1, output_dim = 1, input_length = 1,
                        weights=list(array(init0[3], dim=c(1,1)))) %>% layer_flatten() %>%
                        layer_lambda(function(x) k_sigmoid(x)) %>%
                        layer_repeat_vector(c(q0[1])) %>%
                        layer_reshape(target_shape=c(q0[1]))
    p1      = list(Ones, pp) %>% layer_concatenate() %>%
                                 layer_reshape(target_shape=c(q0[1]+1,1))
    #
    qq      = Zeros %>% layer_embedding(input_dim = 1, output_dim = 1, input_length = 1,
                        weights=list(array(init0[4], dim=c(1,1)))) %>% layer_flatten() %>%
                        layer_lambda(function(x) k_sigmoid(x)) %>%
                        layer_repeat_vector(c(q0[1])) %>%
                        layer_reshape(target_shape=c(q0[1]))
    q1      = list(Ones, qq) %>% layer_concatenate() %>%
                                 layer_reshape(target_shape=c(q0[1]+1,1))
```

```
    #
    lambda = XX %>% time_distributed(layer_dense(units=1, activation='exponential',
                  weights=list(array(0,dim=c(q0[2],1)), array(init0[1], dim=c(1)))))) %>%
                  layer_reshape(target_shape=c(q0[1]))
    # only consider lambda's where have exposure
    lambda0 <- list(lambda, Obs) %>% layer_multiply()
    #
    lambda1 = list(alpha1, lambda0) %>% layer_concatenate() %>%
                                        layer_reshape(target_shape=c(q0[1]+1,1))
    #
    cell <- BetaAlphaCell(1, 1)
    rnn <- layer_rnn(cell = cell, return_sequence=TRUE)
    Network <- rnn(tuple(q1, lambda1, p1, YY1)) %>% layer_concatenate()
    outputs <- list(lambda1, p1, q1, Network) %>% layer_concatenate()
    #
    keras_model(inputs = c(YY, XX, Obs), outputs = list(outputs))
    }
```

Listing 4: Implementation of Model 1 with the regression setting and time-varying p_t and q_t.

```
library(keras)
library(tensorflow)
########
GLM.dynamic <- function(seed, q0, init0){
    k_clear_session()
    set.seed(seed)
    set_random_seed(seed)
    YY      <- layer_input(shape = c(q0[1]), dtype = 'float32')
    XX      <- layer_input(shape = c(q0[1],q0[2]), dtype = 'float32')
    Obs     <- layer_input(shape = c(q0[1]), dtype = 'float32')
    Time    <- layer_input(shape = c(q0[1],1), dtype = 'float32')
    Ones    <- k_ones_like(x=Obs[,1]) %>% layer_flatten()

    # alpha10 = beta10 initialization
    alpha1  = Ones %>% layer_dense(units=1, use_bias=FALSE, activation="exponential",
                                   weights=list(array(init0[2], dim=c(1,1))))
    YY1     = list(alpha1, YY) %>% layer_concatenate() %>%
                                  layer_reshape(target_shape=c(q0[1]+1,1))
    #
    pp      = Time %>% layer_embedding(input_dim = q0[1], output_dim = 1, input_length = 1,
                       weights=list(array(init0[3], dim=c(q0[1],1)))) %>%
                       layer_flatten() %>%
                       layer_lambda(function(x) k_sigmoid(x)) %>%
                       layer_reshape(target_shape=c(q0[1]))
    p1      = list(Ones, pp) %>% layer_concatenate() %>%
                                 layer_reshape(target_shape=c(q0[1]+1,1))
    #
    qq      = Time %>% layer_embedding(input_dim = q0[1], output_dim = 1, input_length = 1,
                       weights=list(array(init0[4], dim=c(q0[1],1))))
                       %>% layer_flatten() %>%
                       layer_lambda(function(x) k_sigmoid(x)) %>%
                       layer_reshape(target_shape=c(q0[1]))
    q1      = list(Ones, qq) %>% layer_concatenate() %>%
                                 layer_reshape(target_shape=c(q0[1]+1,1))
    #
    lambda = XX %>% time_distributed(layer_dense(units=1, activation='exponential',
                  weights=list(array(0,dim=c(q0[2],1)), array(init0[1], dim=c(1)))))) %>%
                  layer_reshape(target_shape=c(q0[1]))
    # only consider lambda's where have exposure
    lambda0 <- list(lambda, Obs) %>% layer_multiply()
    #
    lambda1 = list(alpha1, lambda0) %>% layer_concatenate() %>%
                                        layer_reshape(target_shape=c(q0[1]+1,1))
    #
    cell <- BetaAlphaCell(1, 1)
    rnn <- layer_rnn(cell = cell, return_sequence=TRUE)
    Network <- rnn(tuple(q1, lambda1, p1, YY1)) %>% layer_concatenate()
    outputs <- list(lambda1, p1, q1, Network) %>% layer_concatenate()
    #
    keras_model(inputs = c(YY, XX, Time, Obs), outputs = list(outputs))
    }
```

Data availability

The authors do not have permission to share data.

References

Abdallah, A., Boucher, J.-P., Cossette, H., 2016. Sarmanov family of multivariate distributions for bivariate dynamic claim counts model. Insur. Math. Econ. 68, 120–133.

Ahmad, A., Francq, C., 2016. Poisson QMLE of count time series models. J. Time Ser. Anal. 37 (3), 291–314.

Ahn, J.Y., Jeong, H., Lu, Y., 2023. A simple Bayesian state-space approach to the collective risk models. Scand. Actuar. J. 2023 (5), 509–529.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19 (6), 716–723.

Aknouche, A., Francq, C., 2021. Count and duration time series with equal conditional stochastic and mean orders. Econom. Theory 37 (2), 248–280.

Baum, L.E., Petrie, T., 1966. Statistical inference for probabilistic functions of finite state Markov chains. Ann. Math. Stat. 37 (6), 1554–1563.

Bichsel, F., 1964. Erfahrungstarifierung in der Motorfahrzeug-Haftpflicht-Versicherung. Bull. - Swiss Assoc. Actuar. 64, 119–130.

Bolancé, C., Denuit, M., Guillén, M., Lambert, P., 2007. Greatest accuracy credibility with dynamic heterogeneity: the Harvey-Fernandes model. Belg. Actuar. Bull. 7 (1), 14–18.
Boucher, J.-P., Pigeon, M., 2018. A claim score for dynamic claim counts modeling. arXiv preprint. arXiv:1812.06157.
Brouhns, N., Guillén, M., Denuit, M., Pinquet, J., 2003. Bonus-malus scales in segmented tariffs with stochastic migration between segments. J. Risk Insur. 70 (4), 577–599.
Bühlmann, H., 1967. Experience rating and credibility. ASTIN Bull. 4 (3), 199–207.
Bühlmann, H., Gisler, A., 2006. A Course in Credibility Theory and Its Applications. Springer Science & Business Media.
Bühlmann, H., Straub, E., 1970. Glaubwürdigkeit für Schadensätze. Bull. - Swiss Assoc. Actuar. 70, 111–131.
Chan, I.W.C., Badescu, A.L., Lin, X.S., 2025. Assessing driving risk through unsupervised detection of anomalies in telematics time series data. ASTIN Bull. 55 (2), 205–241.
Chollet, F., Allaire, J., et al., 2017. R interface to keras. https://github.com/rstudio/keras.
Christou, V., Fokianos, K., 2014. Quasi-likelihood inference for negative binomial time series models. J. Time Ser. Anal. 35 (1), 55–78.
Cui, Y., Lund, R., 2009. A new look at time series of counts. Biometrika 96 (4), 781–792.
Davis, R.A., Liu, H., 2016. Theory and inference for a class of nonlinear models with application to time series of counts. Stat. Sin. 26, 1673–1707.
Davis, R.A., Wu, R., 2009. A negative binomial model for time series of counts. Biometrika 96 (3), 735–749.
Debaly, Z.M., Truquet, L., 2021. Iterations of dependent random maps and exogeneity in nonlinear dynamics. Econom. Theory 37 (6), 1135–1172.
Dionne, G., Vanasse, C., 1989. A generalization of automobile insurance rating models: the negative binomial distribution with a regression component. ASTIN Bull. 19 (2), 199–212.
Duca, V.E., Fonseca, T.C., Oliveira, F.L.C., 2021. A generalized dynamical model for wind speed forecasting. Renew. Sustain. Energy Rev. 136, 110421.
Duval, F., Boucher, J.-P., Pigeon, M., 2024. Telematics combined actuarial neural networks for cross-sectional and longitudinal claim count data. ASTIN Bull. 54 (2), 239–262.
Ferland, R., Latour, A., Oraichi, D., 2006. Integer-valued GARCH process. J. Time Ser. Anal. 27 (6), 923–942.
Fokianos, K., 2024. Multivariate count time series modelling. Econom. Stat. 31, 100–116.
Frees, E.W., Lee, G., Yang, L., 2016. Multivariate frequency-severity regression models in insurance. Risks 4 (1), 4.
Frees, E.W., Wang, P., 2005. Credibility using copulas. N. Am. Actuar. J. 9 (2), 31–48.
Frühwirth-Schnatter, S., Wagner, H., 2006. Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state space modelling. Biometrika 93 (4), 827–841.
Fung, T.C., Li, Y., Peng, L., Qian, L., 2024. Testing constant serial dynamics in two-step risk inference for longitudinal actuarial data. N. Am. Actuar. J. 28 (4), 861–881.
Gamerman, D., dos Santos, T.R., Franco, G.C., 2013. A non-Gaussian family of state-space models with exact marginal likelihood. J. Time Ser. Anal. 34 (6), 625–645.
Genest, C., Nešlehová, J., 2007. A primer on copulas for count data. ASTIN Bull. 37 (2), 475–515.
Gerber, H.-U., Jones, D., 1975. Credibility formulas of the updating type. Trans. Soc. Actuar. 27 (1), 31–52.
Gerber, H.U., Jones, D.A., 1973. Credibility formulae with geometric weights. In: Proceedings of the Business and Economic Section, vol. 229. American Statistical Association, p. 230.
Gonçalves, E., Mendes-Lopes, N., Silva, F., 2015. Infinitely divisible distributions in integer-valued GARCH models. J. Time Ser. Anal. 36 (4), 503–527.
Gouriéroux, C., Lu, Y., 2019. Negative binomial autoregressive process with stochastic intensity. J. Time Ser. Anal. 40 (2), 225–247.
Grunwald, G.K., Hyndman, R.J., Tedesco, L., Tweedie, R.L., 2000. Theory & methods: non-Gaussian conditional linear AR(1) models. Aust. N. Z. J. Stat. 42 (4), 479–495.
Harvey, A.C., Fernandes, C., 1989. Time series models for count or qualitative observations. J. Bus. Econ. Stat. 7 (4), 407–417.
Henderson, R., Shimakura, S., 2003. A serially correlated gamma frailty model for longitudinal count data. Biometrika 90 (2), 355–366.
Hyndman, R., Koehler, A.B., Ord, J.K., Snyder, R.D., 2008. Forecasting with Exponential Smoothing: the State Space Approach. Springer Science & Business Media.
Jia, Y., Kechagias, S., Livsey, J., Lund, R., Pipiras, V., 2023. Latent Gaussian count time series. J. Am. Stat. Assoc. 118 (541), 596–606.
Jiang, Q., Shi, T., 2024. Auto insurance pricing using telematics data: application of a hidden Markov model. N. Am. Actuar. J. 28 (4), 822–839.
Jung, R.C., Liesenfeld, R., Richard, J.-F., 2011. Dynamic factor models for multivariate count data: an application to stock-market trading activity. J. Bus. Econ. Stat. 29 (1), 73–85.
Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. J. Basic Eng. 82, 35–45.
Li, Y., Fung, T.C., Peng, L., Qian, L., 2023. Diagnostic tests before modeling longitudinal actuarial data. Insur. Math. Econ. 113, 310–325.
Lu, Y., 2018. Dynamic frailty count process in insurance: a unified framework for estimation, pricing, and forecasting. J. Risk Insur. 85 (4), 1083–1102.
Lukacs, E., 1955. A characterization of the gamma distribution. Ann. Math. Stat. 26 (2), 319–324.
McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. Chapman & Hall.
Nelder, J.A., Wedderburn, R.W., 1972. Generalized linear models. J. R. Stat. Soc. A, General 135 (3), 370–384.
Pinquet, J., 2020a. Poisson models with dynamic random effects and nonnegative credibilities per period. ASTIN Bull. 50 (2), 585–618.
Pinquet, J., 2020b. Positivity properties of the ARFIMA (0, d, 0) specifications and credibility analysis of frequency risks. Insur. Math. Econ. 95, 159–165.
Pinquet, J., Guillén, M., Bolancé, C., 2001. Allowance for the age of claims in bonus-malus systems. ASTIN Bull. 31 (2), 337–348.
Purcaru, O., Denuit, M., 2003. Dependence in dynamic claim frequency credibility models. ASTIN Bull. 33 (1), 23–40.
Shi, P., Yang, L., 2018. Pair copula constructions for insurance experience rating. J. Am. Stat. Assoc. 113 (521), 122–133.
Shi, Y., Punzo, A., Otneim, H., Maruotti, A., 2025. Hidden semi-Markov models for rainfall-related insurance claims. Insur. Math. Econ. 120, 91–106.
Sundt, B., 1988. Credibility estimators with geometric weights. Insur. Math. Econ. 7 (2), 113–122.
Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, fourth edition. Springer, New York.
Wüthrich, M.V., Merz, M., 2023. Statistical Foundations of Actuarial Learning and Its Applications. Springer, Cham.
Xu, H.-Y., Xie, M., Goh, T.N., Fu, X., 2012. A model for integer-valued time series with conditional overdispersion. Comput. Stat. Data Anal. 56 (12), 4229–4242.
Yanez, J.S., Guillén, M., Nielsen, J.P., 2025. Weekly dynamic motor insurance ratemaking with a telematics signals bonus-malus score. ASTIN Bull. 55 (1), 1–28.
Yang, L., 2022. Nonparametric copula estimation for mixed insurance claim data. J. Bus. Econ. Stat. 40 (2), 537–546.
Zeger, S.L., 1988. A regression model for time series of counts. Biometrika 75 (4), 621–629.
Zhu, F., 2011. A negative binomial integer-valued GARCH model. J. Time Ser. Anal. 32 (1), 54–67.