Proceedings of the 2022 Winter Simulation Conference
B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C.G. Corlu, L.H. Lee, E.P. Chew, T. Roeder, and P. Lendermann, eds.

A DYNAMIC CREDIBILITY MODEL WITH SELF-EXCITATION AND EXPONENTIAL DECAY

Himchan Jeong

Department of Statistics and Actuarial Science
Simon Fraser University
8888 University Drive
Burnaby, BC V5A 1S6, CANADA

Bin Zou

Department of Mathematics
University of Connecticut
341 Mansfield Road U1009
Storrs, CT 06269, USA

ABSTRACT

This paper proposes a dynamic credibility model for claim count that extends the benchmark Poisson generalized linear models (GLM) by incorporating self-excitation and exponential decay features from Hawkes processes. Under the proposed model, a recent claim has a bigger impact on the credibility premium than an outdated claim. Empirical results show that the proposed model outperforms the Poisson GLM in both in-sample goodness-of-fit and out-of-sample prediction.

1 INTRODUCTION

Credibility theory is widely used in insurance as a tool to adjust individual premium, based on the past claim experience. Two commonly adopted rules in the ratemaking practice are:
Rule (1) past claims increase the credibility premium;
Rule (2) a recent claim has a bigger impact on the credibility premium than an outdated claim.
This paper proposes a dynamic credibility model for claim count that is consistent with the above two rules. To that end, we extend the benchmark Poisson generalized linear models (GLM) by incorporating two essential features of Hawkes processes, self-excitation and (exponential) decay, with the former implying Rule (1) and the latter Rule (2).

Under credibility theory, the credibility premium consists of two components: manual rate or prior premium that reflects observed heterogeneity among different tariff cells, and credibility factor that accounts for unobservable heterogeneity as a function of past claims. To incorporate these fundamental ideas, Dionne and Vanasse (1989) propose the following random effects models: (i) $N_1, N_2, \dots, N_T$ are independent, conditioning on a random parameter θ, where N_t denotes the claim count in the t-th period (e.g., year) and T is the total number of periods in observation; (ii) the conditional expectation of N_t is given by $\mathbb{E}[N_t|\theta] = g(\nu_t^*, \theta)$, where ν_t^* is a function of observable covariates $\mathbf{x}_t$ and g is the so-called link function; (iii) θ is a time-invariant latent variable. ($\mathbf{x}_t$ includes information that may have predictive power in future claims, e.g., age, length of clean driving records, and vehicle type in auto insurance. As the relation between $\mathbf{x}_t$ and $\mathbb{E}[N_t|\theta]$ is most likely nonlinear, the function g serves such a purpose to link the linear term $\eta\mathbf{x}_t$ with $\mathbb{E}[N_t|\theta]$; a popular choice of g is exponential.) Under a similar framework as above, Frangos and Vrontos (2001) obtain the credibility factors for both claim count and severity, by using Poisson-gamma and exponential-inverse gamma random effects models, respectively. Recently, such random effects models have been extended to incorporate dependence between claim frequency and severity, or among different lines of business (see Jeong and Valdez (2020), Cheung et al. (2021), Denuit and Lu (2021), and Oh et al. (2021) for recent contributions).

A major drawback of the above random effects models is that the impact of a past claim on the credibility factor of claim count is *irrelevant* to its arrival time (see, e.g., Frangos and Vrontos (2001) or (5) for the exact credibility formula). As shown in an example of Section 3.3, the random effects models yield the same credibility premium for a policy with one claim in Year 1 and another policy with one claim in Year 5, both over the same past five years. As a consequence, the random effects models do not comply with Rule (2) mentioned in the first paragraph. Moreover, the latent variable θ is *time-invariant* in the random effects models, and hence may fail to account for the possible evolution of the unobserved heterogeneity of a policyholder over time (e.g., improvement in driving skills).

To address the above drawbacks of the static random effects models, several alternative models are proposed in the actuarial literature. In the first line of research, varying weights are assigned in the determination of credibility factors, with the purpose to fit Rule (2). Sundt (1988) introduces geometric weights so that the impact of past claims decays according to a geometric decreasing factor. The Harvey-Fernandes type claim frequency models also follow the exactly same idea (see Bolancé et al. (2007)). In the second line of research, a dynamic generalization of random effects is made to help capture the possible evolution of unobserved heterogeneity over time (see Bolancé et al. (2003) for an excellent work on this topic). The introduction of dynamic random effects can help interpret Rule (2), but often leads to a case in which there does not exist a closed form expression for the posterior expectation of random effects (see, e.g., Brouhns et al. (2003) and Li et al. (2020)), making such dynamic random effects models less practical for ratemaking. In addition, Pinquet (2020) shows that the credibility coefficients obtained under these dynamic random effects models may take *negative* values, i.e., a past claim may be even seen as a 'bonus'. This would easily lead to moral hazard and cause much bigger problems than violating Rule (2) (see Ahn et al. (2021) for detailed discussions).

In this article, we propose a dynamic credibility model that naturally implies both Rule (1) and Rule (2), and is dramatically different from the existing models with varying weights or dynamic random effects. To achieve this objective, we "borrow" two essential features from Hawkes processes (see Hawkes (1971)), self-excitation and exponential decay, and incorporate them into the benchmark Poisson GLM. Under the proposed model, the intensity excites by a level $\beta > 0$ upon the arrival of a new claim, which implies Rule (1), and such an excitation effect decays exponentially at speed α over time, which leads to the consistence with Rule (2). Further, we apply a different exponential factor to model the possible evolution of the unobservable heterogeneity. Through an empirical analysis, we test the performance of our proposed model, with the standard Poisson GLM (also termed the naïve model) and the static credibility model introduced in Frangos and Vrontos (2001) and Jeong (2020). Our numerical results show that the proposed model performs the best in terms of in-sample goodness-of-fit, and also outperforms the both benchmark Poisson GLM and the static credibility model in mean absolute error (MAE) and root-mean-square error (RMSE) in the out-of-sample prediction. We also conduct a sensitivity analysis to investigate the impact of self-excitation and exponential decay parameters (β and α) on the credibility premium, and find a positive (increasing) relation for the former and a negative (decreasing) relation for the latter.

The rest of the paper is organized as follows. Section 2 proposes our dynamic credibility model. Actuarial applications of the proposed model are provided in Section 3. Section 4 concludes the paper. Appendices A and B collect additional empirical results.

2 THE MODEL

Let us consider a standard data structure for claim count in non-life insurance. We observe the claim counts of an insurance business for I policyholders, indexed by $i = 1, 2, \ldots, I$, over a total of T time periods (e.g., years). Throughout the paper, we assume, without loss of generality, that a unit period is one year. Denote the data structure by $\mathscr{D} = \left\{ \left(\mathbf{N} = (N_{it}); \mathbf{x} = (\mathbf{x}_{it})\right) \middle| i = 1, \ldots, I \text{ and } t = 1, \ldots, T \right\}$, where N_{it} denotes the number of claims and $\mathbf{x}_{it}$ captures the observable covariates of the i-th policyholder in the t-th year. In what follows, we will drop the index i from N_{it} and $\mathbf{x}_{it}$ for notation simplicity and use N_t and $\mathbf{x}_t$. Let $\mathscr{F}_t$ denote the available information up to time t, generated by $\{N_1, \ldots, N_t; \mathbf{x}_1, \ldots, \mathbf{x}_t\}$ for all $t = 1, 2, \ldots$.

Recall that, under the standard Poisson GLM, we have

$$N_{t+1} | \mathscr{F}_t \sim \text{Poisson}\left(\nu^*_{t+1}\right), \quad \text{with } \nu^*_{t+1} = \mathrm{e}^{\eta\, \mathbf{x}_{t+1}}, \tag{1}$$

where η is the regression coefficients associated with the observable policy characteristics $\mathbf{x}_{t+1}$. According to (1), the claim count in the next period N_{t+1}, conditioning on the past claim history $\{N_1,\ldots,N_t\}$ and observable characteristics $\{\mathbf{x}_1,\ldots,\mathbf{x}_t\}$, follows a Poisson distribution, in which the logarithm of its intensity is a linear function of $\mathbf{x}_{t+1}$. While this standard model can reflect the impacts of observable heterogeneity in $\mathbf{x}_{t+1}$, it does not consider unobservable heterogeneity in the ratemaking procedure.

As motivated in Section 1, we now propose the following dynamic credibility model:

$$N_{t+1}|\mathscr{F}_t \sim \text{Poisson}(\lambda_{t+1}), \quad \text{with } \lambda_{t+1} = \mathrm{e}^{-\gamma t}\nu^*_{t+1} + \sum_{\tau_j<t} \beta\,\mathrm{e}^{-\alpha(t-\tau_j)} \text{ and } \nu^*_{t+1} = \mathrm{e}^{\eta\,\mathbf{x}_{t+1}}, \tag{2}$$

where τ_j denotes the arrival time of the j-th claim during the observation period, $\alpha,\beta > 0$, and $\gamma \in \mathbb{R}$. Several important remarks on the above proposed model (2) are due as follows:

- Upon the arrival of a claim at time $\tau_j = s$, the *instantaneous* intensity λ_s excites up by $\beta > 0$, leading to an increased probability of reporting claims in the future. As time evolves, the impact of this particular claim on future credibility decays exponentially at speed $\alpha > 0$; as a result, we model such an impact at time $t > \tau_j$ by $\beta\mathrm{e}^{-\alpha(t-\tau_j)}$. The motivation of incorporating both features comes from Hawkes processes and their applications in actuarial science (see Jang and Oh (2021)).
- The proposed model (2) takes into account the impact of the exogenous and observable covariates $\mathbf{x}$ on ratemaking, as shown in $\nu^*_{t+1} = \mathrm{e}^{\eta\,\mathbf{x}_{t+1}}$ inherited from (1), which is a standard practice in insurance ratemaking.
- To capture the possible evolution of the *unobservable* heterogeneity of a policy, we apply another exponential term $\mathrm{e}^{-\gamma t}$ to ν^*_{t+1}, in which a positive (resp. negative) γ corresponds to improvement (resp. deterioration) in the rating factor. An example of such dynamic heterogeneity is the improvement in driving skills of policyholders over time in the study of auto insurance. Since such evolution is *unobservable*, the proposed exponential factor may not be the best choice to model it. Finding the "best" choice, on the other hand, is a challenging task and likely varies by types of insurance.
- In the limit case of $\alpha,\beta \downarrow 0$ and $\gamma = 0$, the proposed model (2) reduces to the Poisson GLM (1).

In time series theory and also in many applications, the stationary of a process (or a time series) is often a highly preferred property. To ensure the stationary of a one-dimensional Hawkes process with exponential kernel, we impose the following sufficient condition (see Bacry et al. (2015)):

$$\alpha > \beta. \tag{3}$$

3 EMPIRICAL ANALYSIS

In this section, we conduct an empirical analysis to investigate the performance of our dynamic credibility model (2). In particular, we are interested in two tasks:

1. comparing the proposed model (2) with the Poisson GLM (1) and the static credibility model in Frangos and Vrontos (2001);
2. studying the impact of α (exponential decay) and β (self-excitation) on ratemaking.

3.1 Data and Estimations

In the empirical analysis, we use the LGPIF (Wisconsin Local Government Property Insurance Fund) dataset, which consists of policy and claim information of the local government units in Wisconsin (US) and is publicly available. (Please refer to Frees et al. (2016) for a comprehensive introduction on the LGPIF dataset.) While claim information on multiple lines of business is available, we focus on IM (inland marine) claims and their associated policy characteristics from the selected dataset. We then extract 5,240 of such policies observed during the 2006–2010 period from the LGPIF dataset. The policy characteristics used here are summarized in Table 1.

Table 1: Observable policy characteristics (covariates). The two continuous variables (CoverageIM and lnDeductIM) in the above table have the unit of million US dollars.

Categorical	Description			Proportions
TypeCity	Indicator for city entity:	Y=1		14.51 %
TypeCounty	Indicator for county entity:	Y=1		5.92 %
TypeMisc	Indicator for miscellaneous entity:	Y=1		10.78 %
TypeSchool	Indicator for school entity:	Y=1		29.10 %
TypeTown	Indicator for town entity:	Y=1		16.60 %
TypeVillage	Indicator for village entity:	Y=1		23.09 %
Continuous		Minimum	Mean	Maximum
CoverageIM	Log coverage amount of IM claim	0.00	0.87	46.75
lnDeductIM	Log deductible amount of IM claim	0.00	5.34	9.21

Note that in order to fit the proposed model (2), we need to know the exact time when each claim arrives (i.e., τ_j in (2)). However, the claim arrival times are often *not* recorded in the publicly available property and casualty insurance dataset, which is indeed the case for the LGPIF dataset used in our empirical study. (Note that some proprietary insurance dataset may include the exact arrival times of claims, e.g., the Dutch fire insurance data used in Albrecher et al. (2021) contains such information; however, we do not access to any of those datasets.) To proceed, we impose the following assumption on the arrival times of claims: (In Appendix A, we consider a different assumption where all claims arrive according to uniform distributions.)

Assumption 1 All claims occur in the middle of their corresponding policy year, i.e., we set $\tau = t - 0.5$ for all claims reported in the t-th year, where $t = 1, 2, \ldots$.

In the remaining of this subsection, we discuss how the parameters involved in our dynamic credibility model (2) are estimated. Notice that there are four parameters η, α, β, and γ in our model, once Assumption 1 is in place. α and γ capture the exponential decay effects of a claim and a priori classification factor, respectively, and β measures the self-excitation level upon the arrival of a claim, while η is a vector of parameters and determines ν_t^* given covariates $\mathbf{x}_t$ (recall $\nu_t^* = \exp(\eta \mathbf{x}_t)$). To estimate these parameters, we follow a two-step approach that is used in Pechon et al. (2019).

Step 1. We estimate the coefficients vector η, related with a priori classification factor $\nu_t^* = \exp(\eta\, \mathbf{x}_t)$, using the maximum likelihood estimation (MLE) method and setting the claim counts as in the standard Poisson GLM (1). The estimated values of η are summarized in Table 2.

Table 2: Estimated values of η. Below, C.IM and D.IM stand for CoverageIM and InDeductIM, respectively, which are continuous variables from Table 1, and S.E. denotes standard error of an estimate.

Variable	Intercept	City	County	Misc	School	Town	C.IM	D.IM
Estimate	-4.08068	0.95112	1.79571	-2.72573	-0.91916	-0.39509	0.11426	0.07088
S.E.	0.32399	0.19067	0.19756	1.01477	0.27753	0.27715	0.04590	0.00720

Step 2. With η estimated, we apply the MLE to calibrate the remaining parameters α, β, and γ, where the log likelihood is given by

$$\ell(\alpha, \beta, \gamma) = \sum_{i=1}^{I} \sum_{t=1}^{T_i} \log f(n_{it} | \lambda_{i,t-1}), \quad \text{subject to} \quad \beta < \alpha.$$

In the above, I is the total number of observed policies in the training dataset, T_i is the number of years with observations for the i-th policy, and $f(\cdot|\lambda)$ denotes the Poisson probability function with mean λ. The imposed constraint $\beta < \alpha$ comes from the stationarity condition (3). The estimated values of α, β, and γ, denoted by $\hat{\alpha}, \hat{\beta}$, and $\hat{\gamma}$, are given by

$$\hat{\alpha} = 0.3176, \quad \hat{\beta} = 0.2132, \quad \text{and} \quad \hat{\gamma} = 0.1307.$$

3.2 Comparison of Different Credibility Models

In this subsection, we compare the following three credibility models:

1. the naïve credibility model (Poisson GLM (1));
2. the static credibility model (see Frangos and Vrontos (2001));
3. the proposed dynamic credibility model (2).

As commented previously, the naïve model is a special limit case of the proposed dynamic model (2) with $\alpha, \beta \downarrow 0$, and $\gamma = 0$. With the naïve credibility model in (1), we easily obtain

$$\mathbb{E}[N_{T+1}|\mathscr{F}_T] = \nu^*_{T+1} = \exp(\eta\, \mathbf{x}_{T+1}). \tag{4}$$

As a consequence, there is no consideration of unobservable heterogeneity or potential dependence among observed claims under the naïve model.

Under the static credibility model, Frangos and Vrontos (2001) show that

$$\mathbb{E}[N_{T+1}|\mathscr{F}_T] = \frac{r + \sum_{t=1}^{T} N_t}{r + \sum_{t=1}^{T} \nu^*_t}\, \nu^*_{T+1}, \tag{5}$$

where r captures the responsiveness of credibility premium formula to past claims, with a bigger r corresponding to less responsiveness. Note that $\frac{r+\sum_{t=1}^{T} N_t}{r+\sum_{t=1}^{T} \nu^*_t}$ is the posterior expectation of the static random effects, which does not evolve over time, given $\mathscr{F}_T$. As seen from (5), the predictive premium of N_{T+1} depends on the past claim records only through their summation $\sum_{t=1}^{T} N_t$, which does not comply with the common practice that recent claim experiences are given more weights (Rule (2)).

Last, the predictive premium under the proposed dynamic credibility model (2) is given by

$$\mathbb{E}[N_{T+1}|\mathscr{F}_T] = \mathrm{e}^{-\gamma T}\nu^*_{T+1} + \sum_{\tau_j < T} \beta\, \mathrm{e}^{-\alpha(T-\tau_j)}. \tag{6}$$

With all the parameters estimated in Section 3.1, we first assess in-sample goodness-of-fit of each model. A model is preferred if it produces a larger value of loglikelihood or a smaller values of Akaike information criterion (AIC) or Bayesian information criterion (BIC). From Table 3, we observe that the proposed dynamic credibility model (2) is the most favored one among the three models considered, under all three goodness-of-fit measures during the 2006-2010 in-sample period.

Table 3: In-sample goodness-of-fit of the three credibility models.

	Naïve	Static	Dynamic
Loglikelihood	-923.54	-883.04	-866.23
AIC	1863.08	1784.07	1754.45
BIC	1915.60	1843.15	1826.65

Having verified the superior performance of in-sample fitting of the proposed dynamic credibility model, we proceed to investigate its prediction performance on the out-of-sample validation set. (Recall that the training dataset is from 2006 to 2010, while the out-of-sample validation set contains observed IM claims and associated policy characteristics in 2011.) To compare model performance, we use root-mean-square error (RMSE) and mean absolute error (MAE) as metrics, which are defined as follows:

$$\text{RMSE}: \sqrt{\frac{1}{I}\sum_{i=1}^{I}\left(N_{i,T_i+1}-\hat{N}_{i,T_i+1}\right)^2} \quad \text{and} \quad \text{MAE}: \frac{1}{I}\sum_{i=1}^{I}\left|N_{i,T_i+1}-\hat{N}_{i,T_i+1}\right|,$$

where I is the number of policies in the out-of-sample validation set and N_{i,T_i+1} (resp. $\hat{N}_{i,T_i+1}$) denotes the observed (resp. predicted) claim frequency of the i-th policy in the (T_i+1)-th year. A lower RMSE or MAE is preferred.

We present the out-of-sample prediction performance of all three models in Table 4. As expected, the naïve model is the worst among the three models in terms of prediction performance. Therefore, to gauge the relative improvement (in percentage) of a model over the naïve model, we also report the results under the columns of "Relative Improvement", which are computed by

$$\text{Relative Improvement}_{\text{Method}} = \frac{\text{Measured Value}_{\text{Naïve}} - \text{Measured Value}_{\text{Method}}}{\text{Measured Value}_{\text{Naïve}}}.$$

From Table 4, we observe that the proposed dynamic credibility model outperforms the static credibility model in terms of both RMSE and MAE. As such, we conclude that the dynamic credibility model (2) is an appealing improvement to the static model in insurance ratemaking.

Table 4: Out-of-sample model performance in prediction. Note that the prediction formula is (4) for the naïve model (standard Poisson GLM), (5) for the static model, and (6) for the dynamic model.

	Measured Values			Relative Improvement		
	Naïve	Static	Dynamic	Naïve	Static	Dynamic
RMSE	0.6620	0.5192	0.5149	0%	21.57%	22.22%
MAE	0.1224	0.1125	0.1050	0%	8.11%	14.25%

3.3 Impact of α and β on Ratemaking

The proposed dynamic credibility model (2) has two key parameters α and β, where α captures the exponential decay effect and β measures the self-excitation effect of a claim. Both effects are novel to the model (2), and are not captured (or captured in a different manner) in the static credibility model. Therefore, we devote this subsection to the investigation of the impact of α and β on ratemaking. In the analysis, we set $\gamma = \alpha - 0.15$ and remark that the key results hold for a wide range of γ, including the case of $\gamma = 0$.

We consider a hypothetical policyholder whose $\nu_t^* = \nu$ for $t = 1, \dots, 5$ (recall there are a total of 5 years in the training dataset), and who reported one claim over the five years from 2006 to 2010. For this policyholder, we then have $\sum_{t=1}^{5} N_t = 1$, where N_t is the number of claims in the t-th year. In the subsequent studies, we consider three scenarios of prior risk factors (low risk with $\nu = 0.05$; medium risk with $\nu = 0.1$; high risk with $\nu = 0.2$) and compute credibility premiums by varying the values of either α or β, which are displayed in Figure 1. In all plots of Figure 1, the five blue curves depict the credibility premium as a function of α (in the left panel) or β (in the right panel) obtained under the dynamic credibility model (i.e., by (6)), with each one corresponding to a possible realization scenario of the claim history (i.e., there is one claim in one of the five years and zero in the remaining four years). As a comparison, we also plot the credibility premium predicted by the static credibility model (i.e., by (5)), which is always shown by the red horizontal line (note that only the summation $\sum_{t=1}^{5} N_t$, *not* the individual values N_t, is used in the prediction formula (5)). We explain the key findings of Figure 1 as follows:

- Recall that there is only one claim from the considered policy over the five-year training period from 2006 to 2010, implying five possible cases $N_t = 1$, where $t = 1,2,3,4,5$. Here, if $N_1 = 1$, then there was one claim in the first year (i.e., 2006) and zero claim in the remaining four years (i.e., from 2007 to 2010) during the training period. (The other four cases can be understood in a similar way.) As such, in the case of $N_1 = 1$ (resp. $N_5 = 1$), the reported claim is the furthest (resp. nearest) to the prediction year 2011. With that in mind, we easily understand why in all the plots of Figure 1 the case of $N_1 = 1$ (resp. $N_5 = 1$) always yields the lowest (resp. highest) credibility premium, among the five blue curves. In fact, there is a strict increasing relation between the credibility premium and the year of the claim (a larger value of "year" here means the claim is more recent to the current time). This result is consistent with the intuition (also the practice) that a recent claim should have a bigger impact on the credibility premium than an outdated claim (see Rule (2) in Section 1). However, this is violated under the static credibility model, since only the summation $\sum_{t=1}^{5} N_t$ enters the credibility premium formula (5) and all the five cases of $N_t = 1$ lead to exactly the same prediction. Put differently, a claim in 2006 would have the *same* impact as a claim in 2010 under the static credibility model, when we predict the claim count in 2011. Therefore, we conclude that the proposed dynamic credibility model (2) is more realistic than the static credibility model, by taking into account the arrival time of a claim in the process of ratemaking.
- Both α (exponential decay parameter) and β (self-excitation parameter) have a major impact on the credibility premium. As an example, in the upper left plot, when α increases from 0 to 0.35, the credibility premium reduces by more than 50%. The credibility premium is a decreasing function of α. This is because, as α increases, the decay effect becomes stronger and thus the contribution of a past claim to the prediction of future claim count decreases. The credibility premium is an increasing function of β. Recall that β is the *instantaneous* shock (increase) on the future intensity from a claim. In consequence, when β increases, each past claim will result in a bigger instantaneous, and thus long-term, impact on the prediction of future claim count.

4 CONCLUSIONS

We propose a dynamic credibility model on claim count, by incorporating two essential features–self-excitation and exponential decay of Hawkes processes into the standard Poisson GLM. To fit for actuarial applications, the claim intensity consists of a baseline level (from the Poisson GLM) and an accumulated self-excitation level, both adjusted by a separate exponential term. The proposed model is consistent with the two rules frequently used in the insurance practice (i.e., past claims increase premium and more recent claims are assigned with bigger weights in premium). The empirical investigations (using the LGPIF dataset) confirm that the proposed dynamic credibility model outperforms the standard Poisson GLM (also called the naïve model) and the static credibility model in all three metrics–loglikelihood, AIC, and BIC (see Table 3) using the in-sample data. In the out-of-sample validation, the dynamic credibility model still outperforms both the standard Poisson GLM and the static credibility model in both RMSE and MAE. We also conduct a sensitivity analysis to investigate how exponential decay (captured by α) and self-excitation (captured by β) affect the credibility premium. The results show that both play a vital role in the credibility premium, with α negatively and β positively correlated with the credibility premium. To summarize, the proposed dynamic credibility model has better in-sample and out-of-sample performance, comparing to the standard Poisson GLM and the static credibility model in insurance ratemaking.

ACKNOWLEDGMENTS

Himchan Jeong was supported by New Faculty Start-up Grant (NFSG) from Simon Fraser University. Bin Zou acknowledges a start-up grant from the University of Connecticut.

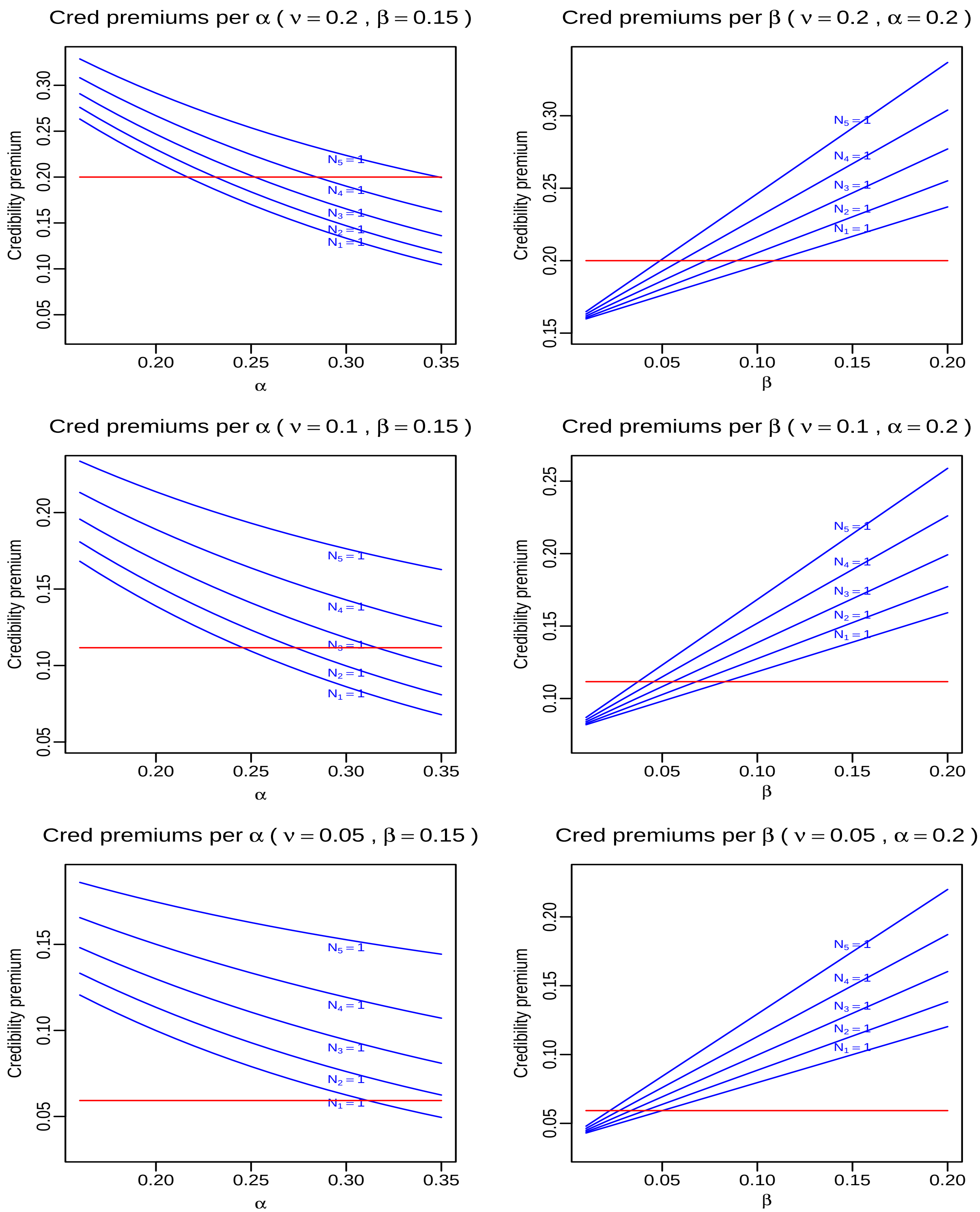

Figure 1: Credibility premiums with varying values of α and β. We set $\nu_t^* = \nu$, $\gamma = \alpha - 0.15$, and $\sum_{t=1}^{5} N_t = 1$ in all plots. The red horizontal line is obtained by (5) under the static credibility model, while the five blue curves are computed by (6) under the proposed dynamic credibility model, with each corresponding to one scenario of $N_t = 1$, where $t = 1, \ldots, 5$.

REFERENCES

Ahn, J. Y., H. Jeong, and Y. Lu. 2021. "On the Ordering of Credibility Factors". *Insurance: Mathematics and Economics* 101:626–638.

Albrecher, H., J. C. Araujo-Acuna, and J. Beirlant. 2021. "Fitting Nonstationary Cox Processes: An Application to Fire Insurance Data". *North American Actuarial Journal* 25(2):135–162.

Bacry, E., I. Mastromatteo, and J.-F. Muzy. 2015. "Hawkes Processes in Finance". *Market Microstructure and Liquidity* 1(01):1550005.

Bolancé, C., M. Denuit, M. Guillén, and P. Lambert. 2007. "Greatest Accuracy Credibility with Dynamic Heterogeneity: The Harvey-Fernandes model". *Belgian Actuarial Bulletin* 7(1):14–18.

Bolancé, C., M. Guillén, and J. Pinquet. 2003. "Time-varying Credibility for Frequency Risk Models: Estimation and Tests for Autoregressive Specifications on the Random Effects". *Insurance: Mathematics and Economics* 33(2):273–282.

Brouhns, N., M. Guillén, M. Denuit, and J. Pinquet. 2003. "Bonus-Malus Scales in Segmented Tariffs with Stochastic Migration Between Segments". *Journal of Risk and Insurance* 70(4):577–599.

Cheung, E. C., W. Ni, R. Oh, and J.-K. Woo. 2021. "Bayesian Credibility under a Bivariate Prior on the Frequency and the Severity of Claims". *Insurance: Mathematics and Economics* 100:274–295.

Denuit, M., and Y. Lu. 2021. "Wishart-gamma Random Effects Models with Applications to Nonlife Insurance". *Journal of Risk and Insurance* 88(2):443–481.

Dionne, G., and C. Vanasse. 1989. "A Generalization of Automobile Insurance Rating Models: the Negative Binomial Distribution with a Regression Component". *ASTIN Bulletin* 19(2):199–212.

Frangos, N. E., and S. D. Vrontos. 2001. "Design of Optimal Bonus-Malus Systems with a Frequency and a Severity Component on an Individual Basis in Automobile Insurance". *ASTIN Bulletin* 31(1):1–22.

Frees, E. W., G. Lee, and L. Yang. 2016. "Multivariate Frequency-Severity Regression Models in Insurance". *Risks* 4(1):4.

Hawkes, A. G. 1971. "Spectra of some Self-exciting and Mutually Exciting Point Processes". *Biometrika* 58(1):83–90.

Jang, J., and R. Oh. 2021. "A Review on Poisson, Cox, Hawkes, Shot-noise Poisson and Dynamic Contagion Process and their Compound Processes". *Annals of Actuarial Science* 15(3):623–644.

Jeong, H. 2020. "Testing for Random Effects in Compound Risk Models via Bregman Divergence". *ASTIN Bulletin* 50:777–798.

Jeong, H., and E. A. Valdez. 2020. "Predictive Compound Risk Models with Dependence". *Insurance: Mathematics and Economics* 94:182–195.

Li, H., Y. Lu, and W. Zhu. 2020. "Dynamic Bayesian Ratemaking: A Markov Chain Approximation Approach". *North American Actuarial Journal* 25(2):186–205.

Oh, R., Y. Lee, D. Zhu, and J. Y. Ahn. 2021. "Predictive Risk Analysis using a Collective Risk Model: Choosing Between Past Frequency and Aggregate Severity Information". *Insurance: Mathematics and Economics* 96:127–139.

Pechon, F., M. Denuit, and J. Trufin. 2019. "Multivariate Modelling of Multiple Guarantees in Motor Insurance of a Household". *European Actuarial Journal* 9(2):575–602.

Pinquet, J. 2020. "Poisson Models with Dynamic Random Effects and Nonnegative Credibilities per Period". *ASTIN Bulletin* 50(2):585–618.

Sundt, B. 1988. "Credibility Estimators with Geometric Weights". *Insurance: Mathematics and Economics* 7(2):113–122.

A UNIFORM DISTRIBUTION OF CLAIM ARRIVAL TIMES

In this section, we drop Assumption 1 but assume that the arrival time of a claim follows a uniform distribution (see Assumption 2 below). The main message is that all the results under Assumption 2 are close to those under Assumption 1 in Section 3.

Assumption 2 The arrival times of all claims follow independent uniform distributions.

We simulate the arrival times for all claims, denoted by $\{\tau_j^{(m)}\}$, where $m = 1,\dots,M$ (with M denoting the total number of runs), during the training period and apply the same MLE method to estimate α and β. Under Assumption 2, we obtain $\hat{\alpha} = \text{mean}(\alpha) = 0.17333$, $\text{s.d.}(\alpha) = 0.04933$; $\hat{\beta} = \text{mean}(\beta) = 0.16683$, $\text{s.d.}(\beta) = 0.04509$; and $\hat{\gamma} = \text{mean}(\gamma) = 0.14149$, $\text{s.d.}(\gamma) = 0.05561$; and The three goodness-of-fit metrics under Assumption 2 are computed as follows: Loglikelihood = -869.14 (2.74), AIC = 1760.28 (5.49), and BIC = 1832.48 (5.49), where the numbers inside parentheses are the corresponding standard deviations. We observe that the means of all three goodness-of-fit matrices are very close to their counterparts in Table 3.

We next compare the prediction performance on the out-of-sample validation set of the same three models as in Section 3.2, where the proposed dynamic model is calibrated under Assumption 2. The results reported in Table 5 show that the proposed dynamic credibility model outperforms the benchmark Poisson GLM (the naïve model) in both RMSE and MAE, and is superior to the static credibility model in terms of MAE, but not in RMSE, which is similar to the conclusion drawn under Assumption 1 (see Table 4 for details).

Table 5: Out-of-sample model performance in prediction under Assumption 2.

	Measured Values			Relative Improvement		
	Naïve	Static	Dynamic	Naïve	Static	Dynamic
RMSE	0.6620	0.5192	0.5302	0%	21.57%	19.91%
MAE	0.1224	0.1125	0.1094	0%	8.11 %	10.63%

B A NUMERICAL EXPERIMENT

In this section, we carry out a numerical experiment based on a synthetic dataset to further illustrate the usefulness of our proposed model (2). For this purpose, consider a dataset consisting of 1,000 policies over six years, denoted by $\{N_{i,t}, \mathbf{x}_i | t = 1,\dots,6,\, i = 1,\dots,1000\}$, and suppose that $N_{i,t}$ is given by

$$N_{i,t} \sim \text{Poisson}((1-\rho)\lambda_i + \rho N_{i,t-1}), \quad \rho = \mathrm{e}^{-0.5}, \lambda_i = \mathrm{e}^{-1+2\mathbf{x}_i}, \mathbf{x}_i \sim \mathcal{N}(0, 0.3^2). \tag{7}$$

We will use the data from the first five years ($t = 1,\dots,5$) to calibrate four models: (i) the naïve model, (ii) the static model, (iii) the proposed dynamic model, and (iv) the true model as in Equation (7), and use the sixth year ($t = 6$) to test their performance in prediction.

In the numerical experiment, we repeat the generation of dataset and model fitting for 100 times to minimize the possible impact from variations due to random seeds. We then compute the average of the RMSEs and MAEs for each model and report the results in Table 6. We observe that the proposed dynamic model outperforms both the naïve and static models significantly, and is even close to the true model. This further justifies the usefulness of our proposed model.

Table 6: Out-of-sample model performance in prediction with the synthetic data generated from (7).

	Measured Values				Relative Improvement			
	Naïve	Static	Dynamic	True	Naïve	Static	Dynamic	True
RMSE	0.8375	0.7881	0.7336	0.6655	0%	5.90%	12.41%	20.55%
MAE	0.5667	0.5341	0.4494	0.4162	0%	5.75%	20.70%	26.56%