POISSON MODELS WITH DYNAMIC RANDOM EFFECTS AND NONNEGATIVE CREDIBILITIES PER PERIOD

BY

JEAN PINQUET

ABSTRACT

This paper provides a toolbox for the credibility analysis of frequency risks, with allowance for the seniority of claims and of risk exposure. We use Poisson models with dynamic and second-order stationary random effects that ensure nonnegative credibilities per period. We specify classes of autocovariance functions that are compatible with positive random effects and that entail nonnegative credibilities regardless of the risk exposure. Random effects with nonnegative generalized partial autocorrelations are shown to imply nonnegative credibilities. This holds for ARFIMA$(0, d, 0)$ models. The AR(p) time series that ensure nonnegative credibilities are specified from their precision matrices. The compatibility of these semiparametric models with log-Gaussian random effects is verified. Gaussian sequences with ARFIMA$(0, d, 0)$ specifications, which are then exponentiated entrywise, provide positive random effects that also imply nonnegative credibilities. Dynamic random effects applied to Poisson distributions are retained as products of two uncorrelated and positive components: the first is time-invariant, whereas the autocovariance function of the second vanishes at infinity and ensures nonnegative credibilities. The limit credibility is related to the three levels for the length of the memory in the random effects. The limit credibility is less than one in the short memory case, and a formula is provided.

KEYWORDS

Credibility, linear filtering, precision matrices, Perron–Frobenius, generalized partial autocorrelation coefficients, Levinson–Durbin recursion, hereditarity, Wold decomposition, spectral measure, ergodicity, generalized method of moments, limit credibility.

JEL codes: C14, C18, C23.

Astin Bulletin 50(2), 585-618. doi:10.1017/asb.2020.4

1. Introduction

This paper provides a toolbox for the credibility analysis of frequency risks, with allowance for the seniority of claims and of risk exposure. We use Poisson models with dynamic and stationary random effects that ensure nonnegative credibilities per period. We specify classes of autocovariance functions that are compatible with positive random effects and that entail nonnegative credibilities regardless of the risk exposure. To achieve this goal, the main tools from the theory of stationary time series (Cramér and Leadbetter, 1967; Ash and Gardner, 1975) are used in both the time and frequency domains.

Experience rating models on nonlife insurance data are derived from a two-step analysis. An analysis of risk distributions is first performed on a mixture model. Then, a prediction is derived at the individual level (see Frees (2004) for a survey of longitudinal data analysis, Lemaire (2013) for a survey of actuarial models in automobile insurance, and Angers *et al.* (2018) for a parametric model on longitudinal and exogenously stratified count data). This time series step uses the posterior distributions of the random effects in a parametric setting, whereas linear predictors are obtained from a second-order semiparametric approach.

The stationary random effects used in Poisson mixtures are linked with a plausible invariance assumption: the predictive ability of events on frequency risk depends on their seniority but not on calendar time. For nonlife insurance risks, this predictive ability is usually positive and it decreases with the seniority of events. The autocovariance function of the random effects is then expected to be positive and decreasing for positive lags. These results are confirmed using real-life data (Pinquet *et al.*, 2001; Bolancé *et al.*, 2003). Dynamic random effects enable the reduction of discrepancies between the actuarial and the real-world experience rating structures. The main results on this issue are recalled in Section 7.1.

The credibilities per period must be nonnegative for a credibility analysis of frequency risks, to prevent negative values for the linear predictors of the frequency. However, negative credibilities can be obtained for some periods from a positive autocovariance function, as described in more detail in Section 2. Thus, while linear prediction has obvious advantages in terms of readability and simplicity of derivations, it also has shortcomings. In summary, linear credibility constrains the shape but relaxes the support of the predictor.

To simplify the derivations, we first assume that the frequency risk before distribution mixing is constant for each individual. We study the vector of stacked credibilities per period as a function of frequency risk, and the conditions that force this vector to stay in the nonnegative orthant. Two polar cases are investigated, which correspond to the limits at the endpoints of the interval of frequencies. The main weakness of the linear credibility approach (credibilities per period can be negative) is related to the limit at infinity. This limit is related to a filtering equation of the random effects. Then, we obtain a sufficient condition for nonnegative credibilities that is valid regardless of the frequency risk level and the length of the individual history. Consider the inverse variance–covariance matrices of the random effects, which are also referred to as precision matrices: the sufficient condition is the nonpositivity of their off-diagonal entries. A statistical interpretation is that the generalized partial autocorrelation coefficients of the random effects are nonnegative. The preceding results also hold in a strict sense, and the proof uses the Perron–Frobenius theorem. This sufficient condition for nonnegative credibilities also holds with time-varying frequencies and hence is valid regardless of the risk exposure.

The credibilities per period are defined as filtering coefficients with an invariance assumption on frequency risks. The Levinson–Durbin recursion on the filtering coefficients is recalled because it is instrumental in this paper.

Semiparametric specifications of random effects that fulfill the sufficient condition for nonnegative credibilities are presented in Section 3. This condition holds for ARFIMA$(0, d, 0)$ models. The AR(p) processes that fulfill the condition are specified from their precision matrices. Gaussian sequences with ARFIMA$(0, d, 0)$ specifications, which are then exponentiated entrywise, provide positive random effects that also ensure nonnegative credibilities.

Semiparametric specifications apply to positive random effects in Poisson mixtures, which raises a compatibility issue between the aforementioned autocovariance functions and positive random effects. AR(1) models with positive autocorrelations entail positive credibilities. They are followed by the autoregressive gamma process, which is a discrete time version of the Cox–Ingersoll–Ross process and does not cross the zero barrier (see Lu (2018) for an application to experience rating). Dynamic random effects are treated as discrete time stochastic interest rates with a positivity constraint. With the Levinson–Durbin recursion, we verify that ARFIMA$(0, d, 0)$ models and AR(p) specifications that ensure nonnegative credibilities ($p = 1, 2, 3$) are compatible with log-Gaussian random effects. This result provides a further two autocovariance specifications for credibility analysis. Thus, three semiparametric specifications for random effects are obtained in Section 4, which have the expected positivity properties and either long or short memory.

The autocovariance functions retained in Section 4 vanish at infinity. However, this ergodicity property is not a desirable assumption (see Table 1, Section 7.1). Placing side by side a time-invariant random effect and a dynamic random effect with a vanishing autocovariance function provides an answer to this issue. The autocovariance specifications presented in Section 5 are in line with the Wold additive decomposition of a stationary random sequence, although a multiplicative specification is needed in the present setting. The sufficient condition for nonnegative credibilities still holds if a time-invariant random effect is placed side by side with a dynamic random effect from any of the specifications obtained in Section 4. The three levels for the length of the memory in the random effects are reached by the specifications of Section 5, which include the time-invariant case.

https://doi.org/10.1017/asb.2020.4 Published online by Cambridge University Press

Section 6 relates the limit credibility to the three levels for the length of the memory in the random effects, with a time-invariant assumption on frequency risks. The limit credibility is less than one in the short memory case. A formula is obtained from the definition of credibilities per period as filtering coefficients and from derivations of spectral densities. The limit credibility equals one otherwise, whether the variance of the time-invariant component is greater than or equal to zero. A case study is presented in Section 7 from the estimated autocovariances of the random effects given in Pinquet *et al.* (2001). The autocovariance specifications of Section 5 are fitted to the estimated autocovariances of random effects with a generalized method of moments (GMMs). The estimated variance of the time-invariant component strongly depends on the length of the memory in the dynamic component of the random effect. Between–within derivations are then applied to the random effects and the ergodicity assumption is discussed. Concluding remarks are given in Section 8. Technicalities are relegated to an appendix, which is shared between the printed version of the paper (Appendix A) and a document (Appendix B) available in an online folder. This folder contains programs, which are commented on in Appendix C.

2. Necessary and Sufficient Conditions for Nonnegative Credibilities at the Period Level

2.1. Credibilities and autocovariance functions of random effects

This section links the credibility coefficients with the autocovariance structure of the random effects in Poisson mixtures. We consider a longitudinal dataset. The number of events per year (e.g., claims of a given type) is the dependent variable and we write

$$N_{i,t} \sim P\left(\lambda_{i,t}\, u_{i,t}\right); i = 1, \ldots, m; t = 1, \ldots, T_i. \tag{2.1}$$

Equation (2.1) specifies a Poisson distribution on the count variable, which is observed on an unbalanced panel dataset. Frequency risk based on the observable information is denoted by λ and can be estimated *ad libitum* (e.g., log-linear specification for the expectation, neural network, regression tree, or random forest). The regression components cannot include the individual history in a mixture model, in order to restrict experience rating to the second component of the expectation. The Poisson distributions in (2.1) are conditional on $U_{i,t} = u_{i,t}$, where $U_{i,t}$ is a positive random effect.

By default, a second-order (weak) stationary framework is retained in this paper for the random effects. However, there is an exception that is related to the issue of the compatibility between autocovariance functions and positive random effects. Strictly stationary log-Gaussian sequences are retained in Section 4.1.

The autocovariance function of U is denoted by γ_U, with the following assumption:

$$\gamma_U(h) > 0 \; \forall h \in \mathbb{Z}. \tag{2.2}$$

The random effects are positive, with an expectation equal to one. This is assumed without loss of generality because of the intercept. The individual index i is removed in the prediction. Linear credibility is related to an affine probabilistic regression of U_{T+1} with respect to $N_1, \ldots, N_T$ in the model with random effects (Bühlmann (1967), Bühlmann and Gisler (2006)). The results are derived with the exact moments, which should be replaced by their estimated counterparts. The predictor is

$$L\left(U_{T+1} \mid N_1, \ldots, N_T\right) = \left(1 - \sum_{h=1}^{T} \alpha_{T,h}\right) + \left(\sum_{h=1}^{T} \alpha_{T,h} \frac{N_{T+1-h}}{\lambda_{T+1-h}}\right), \tag{2.3}$$

where the index h refers to the lag between the forecast period and the risk exposure period. The credibility for period $T+1-h$ is denoted by $\alpha_{T,h}$. The experience-rated frequency premium for period $T+1$ is the product of this predictor and of λ_{T+1}, with estimated parameters. If V_U^T is the variance–covariance matrix of $U_1, \ldots, U_T$, then the vector of stacked credibilities per period is (see Pinquet *et al.* (2001)):

$$v_\alpha^T \stackrel{\text{def}}{=} \underset{1 \le h \le T}{vec}\left(\alpha_{T,h}\right) = \left[I_T + \left(\Lambda_T\, V_U^T\right)\right]^{-1} \Lambda_T\, v_{\gamma_U}^T = \left[\Lambda_T^{-1} + V_U^T\right]^{-1} v_{\gamma_U}^T, \tag{2.4}$$
$$\text{with: } \Lambda_T = \underset{1 \le h \le T}{diag}\left(\lambda_{T+1-h}\right); \; v_{\gamma_U}^T = \underset{1 \le h \le T}{vec}\left(\gamma_U(h)\right).$$

In this paper, V_U^T is assumed to be positive definite, regardless of the length T of the history. In this case, the autocovariance function γ_U is termed positive definite. This assumption allows the identification of the filtering coefficients derived later in the paper. In this framework, we exclude time-invariant random effects and some—but not all—of the deterministic random effects (which refers to the Wold decomposition, see Equation (5.28)).

2.2. Necessary conditions for nonnegative credibilities: nonnegative autocovariances and filtering coefficients

In the following, we suppose that the frequency risks $(\lambda_t)_{t=1,\ldots,T}$ depend on the latent individual index but not on time. This restriction is retained to keep the derivations tractable. If λ is the time-invariant frequency risk, then Equation (2.4) becomes

$$v_\alpha^T(\lambda) = \lambda \left[I_T + \left(\lambda\, V_U^T\right)\right]^{-1} v_{\gamma_U}^T = \left[\frac{I_T}{\lambda} + V_U^T\right]^{-1} v_{\gamma_U}^T \text{ if } \lambda > 0. \tag{2.5}$$

https://doi.org/10.1017/asb.2020.4 Published online by Cambridge University Press

The vector of stacked credibilities is expressed as a function of the annual frequency risk. These two formulations entail two polar results

$$\begin{cases} v_{\alpha}^{T}(0)=0 \\ \left(v_{\alpha}^{T}\right)'(0)=v_{\gamma_U}^{T} \end{cases} ; \lim_{\lambda\to+\infty} v_{\alpha}^{T}(\lambda)=\left[V_U^T\right]^{-1} v_{\gamma_U}^{T}. \tag{2.6}$$

Geometrically, the first part of (2.6) means that the vector of stacked credibilities starts from the vertex of the nonnegative orthant, with a derivative belonging to the interior of this set. This results from the assumption made in (2.2). The vector of stacked credibilities must remain in the nonnegative orthant to prevent negative values for linear predictors of the frequency. In the neighborhood of zero, the behavior of v_{α}^{T} is satisfactory because of the positivity assumption on the autocovariances. In addition, a second-order semiparametric analysis of Poisson mixtures is better justified in a low-frequency setting.

The weak point of the linear credibility approach is related to the limit result given in (2.6). The limit $\left[V_U^T\right]^{-1} v_{\gamma_U}^{T}$, which is denoted by $v_{\varphi_U}^{T}$, is derived from the affine probabilistic regression of U_t with respect to $U_{t-1},\ldots,U_{t-T}$, $\forall t>T$. This limit relates to linear filtering of the random effects. Some entries of $v_{\varphi_U}^{T}$ can be negative even if the autocovariances are positive. To see this, consider a stationary AR(2) specification of $U^c=U-1$:

$$\begin{aligned} &(I-x_1L)(I-x_2L)\,U_t^c=\varepsilon_t \Leftrightarrow \\ &U_t^c=(x_1+x_2)U_{t-1}^c-x_1x_2U_{t-2}^c+\varepsilon_t, \end{aligned} \tag{2.7}$$

where L is the lag operator and ε is a white noise process. We assume that $1>x_2\geq x_1>0$. If ρ is the autocorrelation function of U, then we obtain

$$\rho_1=\frac{x_1+x_2}{1+x_1x_2};\ \frac{\rho_{h+2}}{\rho_{h+1}}=f\left(\frac{\rho_{h+1}}{\rho_h}\right)\ \forall h\in\mathbb{N},\ \text{with}\, f(r)=(x_1+x_2)-\frac{x_1x_2}{r}$$

from the Yule–Walker equations. From $f(x_2)=x_2, f'(x_2)\leq 1$ and the concavity of f, we obtain $f(r)\leq r\ \forall r\geq x_2$. As $\rho_1/\rho_0=\rho_1\geq x_2$, the sequence $(\rho_{h+1}/\rho_h)_{h\in\mathbb{N}}$ decreases from ρ_1 to x_2. The correlation coefficients are positive and decreasing for positive lags but

$$\lim_{\lambda\to+\infty} v_{\alpha}^{2}(\lambda)=v_{\varphi_U}^{2}=\begin{pmatrix} x_1+x_2 \\ -x_1x_2 \end{pmatrix}\notin\left(\mathbb{R}^+\right)^2.$$

This example shows the weak point of the linear credibility approach. The location of these specifications in the stationarity triangle related to AR(2) specifications is given in Appendix A.1. Two arguments follow in defense of linear credibility.

- Frequency risks per period are low at the individual level in nonlife insurance, and we are closer to the first polar case.

https://doi.org/10.1017/asb.2020.4 Published online by Cambridge University Press

- There are many stationary time series with nonnegative filtering coefficients, such as the AR(1) specification with positive autocorrelations. The autoregressive gamma process is AR(1) and does not cross the zero barrier. This paper addresses the issue of compatibility between autocovariance functions and positive random effects. For instance, the specifications of the counterexample are compatible with log-Gaussian sequences (see Section 4 and Appendix A.1).

The two polar results given in (2.6) entail two necessary conditions for nonnegative credibilities. Nonnegative autocovariances result from the first polar case and nonnegative filtering coefficients from the limit at infinity. The second necessary condition for nonnegative credibilities is

$$\left[V_U^T\right]^{-1} \; v_{\gamma U}^T = v_{\varphi U}^T \in \left(\mathbb{R}^+\right)^T . \tag{2.8}$$

2.3. Sufficient conditions for nonnegative credibilities and the nonpositivity of the off-diagonal entries of precision matrices

The first part of Equation (2.5) implies

$$v_\alpha^T(\lambda) = \lambda \left[\lambda\, I_T + \left[V_U^T\right]^{-1}\right]^{-1} \; v_{\varphi U}^T. \tag{2.9}$$

If all the entries of $\left[\lambda\, I_T + \left[V_U^T\right]^{-1}\right]^{-1}$ are nonnegative regardless of the frequency risk λ, then sufficient conditions for nonnegative credibilities per period are obtained. This remark motivates the following proposition.

Proposition 1. *Together with (2.8), the condition*

$$P_{h\tau} \leq 0 \; \forall h,\; \tau = 1, \ldots, T,\; h \neq \tau \; \left(\text{with } P \overset{\text{def}}{=} \left[V_U^T\right]^{-1}\right), \tag{2.10}$$

implies $v_\alpha^T(\lambda) \in \left(\mathbb{R}^+\right)^T \;\; \forall \lambda \geq 0$. *The matrix P is called a precision matrix, and the nonpositive off-diagonal entries of P correspond to nonnegative generalized partial autocorrelation coefficients of the random effects (see Section 2.4).*

The strict positivity result: $\lambda > 0 \Rightarrow v_\alpha^T(\lambda) \in \left(\mathbb{R}^{+*}\right)^T$ *is obtained with the supplementary condition*

$$|h - \tau| = 1 \Rightarrow P_{h\tau} < 0. \tag{2.11}$$

The proof of this proposition uses a class of matrices that is detailed in Berman and Plemmons (1994). A square matrix P of order T is called an M-matrix if

$$P = s \times (I_T - B)\,, \text{ with: } s > 0\,; B \geq 0\,; \rho(B) \leq 1. \tag{2.12}$$

More precisely, (i) s is a scalar; (ii) $B \geq 0$ holds entrywise; (iii) ρ is the spectral radius function (i.e., the maximum modulus of the eigenvalues). If the M-matrix P is nonsingular, then $\rho(B) < 1$ as a consequence of the Perron–Frobenius theorem (see Appendix B.1). Conversely, $\rho(B) < 1 \Rightarrow P$ is nonsingular, with $P^{-1} = (1/s) \times \sum_{n \in \mathbb{N}} B^n$. As $B \geq 0$ entrywise, we have that

Lemma 2. *The inverse of a nonsingular M-matrix is nonnegative entrywise.*

The precision matrix $P = [V_U^T]^{-1}$ is an M-matrix if condition (2.10) is fulfilled because a positive definite matrix with nonpositive off-diagonal entries is an M-matrix (see Appendix B.1). The same property holds for $\lambda\, I_T + [V_U^T]^{-1}$ if $\lambda \geq 0$. Indeed, $\lambda\, I_T + [V_U^T]^{-1}$ is positive definite and the off-diagonal elements are those of $[V_U^T]^{-1}$. From Lemma 2, all the entries of $[\lambda\, I_T + [V_U^T]^{-1}]^{-1}$ are nonnegative. Therefore, Proposition 1 is proved in the broad sense by (2.8) and (2.9).

There is a strict version of Lemma 2. If the entries of the M-matrix are negative on the subdiagonal and the superdiagonal, all the entries of the inverse are positive. Applying this new version to the M-matrix $\left[V_U^T\right]^{-1}$ proves Proposition 1 in a strict sense because $v_{\varphi U}^T \neq 0$ (see Appendix B.1). ■

Proposition 1 is illustrated with an AR(1) sequence, with $\rho_1 \in]0,\ 1[$. From stationarity, V_U^T can be assumed to be a correlation matrix without loss of generality. The precision matrix has a tridiagonal structure. The diagonal elements of P are $P_{11} = P_{TT} = \frac{1}{1-\rho_1^2}$; $P_{hh} = \frac{1+\rho_1^2}{1-\rho_1^2}$ if $1 < h < T$. Besides, $\left(v_{\varphi U}^T\right)_1 = \rho_1$; $P_{h\tau} = \frac{-\rho_1}{1-\rho_1^2}$ if $|h - \tau| = 1$. All the other entries of $v_{\varphi U}^T$ and P are zero. If $\rho_1 \in]0,\ 1[$, then the assumptions of Proposition 1 are fulfilled in the strict sense and the credibilities are positive if $\lambda > 0$.

An interpretation of these results with a varying length of the history stems from the Levinson–Durbin recursion (see Section 3.2). Condition (2.10) implies that the credibilities per period are nonnegative, regardless of the annual frequency risk λ and the length τ of the history if $\tau < T$ (see Proposition 5).

2.4. Generalized partial autocorrelation coefficients and correlations derived from precision matrices

The off-diagonal entries of a precision matrix have a sign at the opposite of generalized partial autocorrelation coefficients. To see this, let $r(U_h,\ U_\tau \mid U_s, \ldots, U_t)$ (with $s \leq h < \tau \leq t$) denote the correlation coefficient between the residuals of the affine probabilistic regression of U_h and U_τ with respect to the variables $\{U_s, \ldots, U_t\} - \{U_h,\ U_\tau\}$. The partial autocorrelation coefficient of U_h and U_τ derived in Section 3.2 by the Levinson–Durbin recursion corresponds to $r(U_h,\ U_\tau \mid U_h, \ldots, U_\tau)$. Then $r(U_h,\ U_\tau \mid U_s, \ldots, U_t)$ is called a generalized partial autocorrelation coefficient, which is linked to the precision matrix $P = \left[V_U^T\right]^{-1}$ by

https://doi.org/10.1017/asb.2020.4 Published online by Cambridge University Press

$$r(U_h,\ U_\tau \mid U_1, \ldots, U_T) = -P_{h\tau} / \left(\sqrt{P_{hh}}\,\sqrt{P_{\tau\tau}}\right) \overset{\text{def}}{=} -R_{h\tau}\ (h \neq \tau) \tag{2.13}$$

(see Appendix B.2). Equation (2.13) provides the interpretation given in Proposition 1 for condition (2.10). If P is seen as a variance–covariance matrix, then R is a correlation matrix.

2.5. Sufficient conditions extended to time-varying frequency risks

Proposition 1 can be extended to time-varying frequency risks $\lambda_1, \ldots, \lambda_T$ and hence to time-varying regression components.

Proposition 3. *The positivity result on credibilities obtained in Proposition 1 is valid in the wide and in the strict sense for time-varying frequency risks $\lambda_1, \ldots, \lambda_T$. With the notations of equation (2.4), we have that*

$$\forall \lambda_1, \ldots, \lambda_T > 0:\ \left[\Lambda_T^{-1} + V_U^T\right]^{-1} v_{\gamma_U}^T = v_\alpha^T \in \left(\mathbb{R}^+\right)^T,$$

if conditions (2.8) and (2.10) are fulfilled on the random effects. The credibilities per period are positive if the supplementary condition (2.11) holds.

The proof is given in Appendix A.2. We use a geometrical interpretation of linear filtering coefficients and of credibilities per period. These coefficients are interpreted from barycentric coordinates in different affine bases included in the simplex. The proof uses the vocabulary of affine and projective geometry.

3. Autocovariance Functions for Random Effects that Ensure Nonnegative Credibilities per Period Regardless of the Risk Exposure

3.1. Credibilities per period as linear filtering coefficients

If frequency risks are time-invariant, then the credibility vector given in (2.5) is obtained from the linear filtering of an auxiliary sequence. For $\lambda > 0$, we write

$$X_t \overset{\text{def}}{=} A_t + (U_t - 1) \overset{\text{def}}{=} A_t + U_t^c,\ t \in \mathbb{N}^*. \tag{3.14}$$

The sequence A is white noise, with a variance equal to $1/\lambda$. The sequence A accounts for risk exposure and is uncorrelated with U. The time series A, $U^c = U - 1$, and X are centered, with $v_{\gamma_X}^T = v_{\gamma_U}^T\ \forall T \in \mathbb{N}^*$. From the definition of X and equation (2.5), we obtain

$$V_X^T = \frac{I_T}{\lambda} + V_U^T;\ v_\alpha^T(\lambda) = \left[V_X^T\right]^{-1} v_{\gamma_X}^T = v_{\varphi_X}^T. \tag{3.15}$$

The entries of $v_{\varphi_X}^T = \left[V_X^T\right]^{-1} v_{\gamma_X}^T$ are obtained from the linear probabilistic regression of X_t with respect to $X_{t-1}, \ldots, X_{t-T}\ \forall t > T$ and hence from linear filtering of X. We write

$$v_{\varphi_X}^T = \underset{1 \le h \le T}{vec}\left(\varphi_{T,h}^X\right) \Rightarrow X_t = \sum_{h=1}^{T} \varphi_{T,h}^X \; X_{t-h} + E_t^T \;\; \forall t > T. \tag{3.16}$$

From (3.15), the vector $v_\alpha^T(\lambda)$ of stacked credibilities per period equals $v_{\varphi_X}^T$. Hence, credibilities per period are obtained from linear filtering of X. In equation (3.16), t refers to contract time, T is the length of the history, and h is a lag.

As is well-known, $\varphi_{T,T}^X$ (the last filtering coefficient in (3.16)) is the correlation coefficient between the residuals of the linear probabilistic regression of X_t and X_{t-T} with respect to the intermediate variables. This result justifies calling $\varphi_{T,T}^X$ a partial autocorrelation coefficient (see Section 2.4). We write $\varphi_{T,T}^X = pac_T^X$.

3.2. The Levinson–Durbin recursion

From (3.15), the credibilities per period are filtering coefficients. Consequently, they follow the Levinson–Durbin recursion with respect to the length of the history (Levinson, 1946; Durbin, 1960). The Levinson–Durbin recursion derives the filtering coefficients and can be used to obtain both numerical and theoretical results about the credibilities.

As recursions will be applied later to either X or U, the underlying variable is latent in the notation. Autocorrelations are used in the recursion and are denoted by $\rho_h = \gamma_h/\gamma_0$. The autocorrelation function is assumed to be positive definite (i.e., all the correlation matrices are positive definite). This is the case for X, from (3.15). The correlation matrix of T consecutive values is denoted by R_T and the vector of stacked autocorrelations $vec_{1 \le h \le T}(\rho_h)$ by v_ρ^T. The corresponding vector of stacked filtering coefficients is denoted by v_φ^T, with $v_\varphi^T = R_T^{-1} v_\rho^T$.

The Levinson–Durbin algorithm simultaneously provides a recursion formula for the filtering coefficients and for the accuracy of the linear prediction. The relative prediction error from a T-period history (see (3.16) for X) reflects the accuracy of the prediction and is denoted by

$$\frac{\gamma_{E^T}(0)}{\gamma_0} = 1 - \left\| v_\rho^T \right\|_{R_T^{-1}}^2 \overset{\text{def}}{=} \sin^2(\psi_T). \tag{3.17}$$

Proposition 4. *The scalar product of two vectors $x,\ y \in \mathbb{R}^T$ is denoted by $(\underset{\sim}{x} \mid y)$. The central symmetry applied to matrices is denoted by $A \to \widetilde{A}$ (i.e., $\widetilde{A}_{i,j} = A_{m+1-i,\,n+1-j}$ if A has m rows and n columns). We write $aetl(x)$ $(x \in \mathbb{R}^T,\ T \ge 2)$ for the vector with all of the components of x except the last. The vector of stacked filtering coefficients v_φ^T and $\sin^2(\psi_T)$ is obtained by the following recursion with respect to the length T.*
Initial values:

$$v_\varphi^1 = pac_1 = \rho_1\,;\ \sin^2(\psi_1) = 1 - pac_1^2.$$

https://doi.org/10.1017/asb.2020.4 Published online by Cambridge University Press

Recursion $T \longrightarrow T+1\ (T \in \mathbb{N}^*)$:

$$\varphi_{T+1,T+1} = pac_{T+1} = \frac{\rho_{T+1} - (\widetilde{v_\varphi^T} \mid v_\rho^T)}{\sin^2(\psi_T)}; \tag{3.18}$$

$$aetl(v_\varphi^{T+1}) = v_\varphi^T - \left(pac_{T+1}\, \widetilde{v_\varphi^T}\right); \tag{3.19}$$

$$\sin^2(\psi_{T+1}) = \sin^2(\psi_T)\,(1 - pac_{T+1}^2) \tag{3.20}$$

The proof is recalled here because its intermediate results are used throughout the present paper. The block-matrix identity for positive definite matrices

$$\begin{pmatrix} A & b \\ b' & c \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + \gamma d d' & -\gamma d \\ -\gamma d' & \gamma \end{pmatrix}, \text{ with } \gamma = \frac{1}{c - b' A^{-1} b}; d = A^{-1} b \tag{3.21}$$

implies $c > b' A^{-1} b$. Equation (3.21) and $v_\varphi^T = R_T^{-1} v_\rho^T$ lead to

$$R_{T+1} = \begin{pmatrix} R_T & \widetilde{v_\rho^T} \\ \widetilde{v_\rho^T}' & 1 \end{pmatrix} \Rightarrow R_{T+1}^{-1} = \begin{pmatrix} R_T^{-1} + \frac{\widetilde{v_\varphi^T}\, \widetilde{v_\varphi^T}'}{\sin^2(\psi_T)} & -\frac{\widetilde{v_\varphi^T}}{\sin^2(\psi_T)} \\ -\frac{\widetilde{v_\varphi^T}'}{\sin^2(\psi_T)} & \frac{1}{\sin^2(\psi_T)} \end{pmatrix}. \tag{3.22}$$

The proof that $R_T^{-1}\, \widetilde{v_\rho^T} = \widetilde{v_\varphi^T}$ results from symmetry arguments. Besides the usual symmetry (with respect to the main diagonal), the correlation matrix is a Toeplitz matrix, which means that its entries are constant along the direction of the main diagonal. This results from stationarity. These two invariance properties imply central symmetry, namely $\widetilde{R_T} = R_T$. The identity $\widetilde{AB} = \widetilde{A}\,\widetilde{B}$ holds for every pair of matrices A and B that can be multiplied together. Given that identity matrices are centrally symmetric, $A = \widetilde{A}$ and A invertible implies $A^{-1} = \widetilde{A^{-1}}$. Then $R_T^{-1}\, \widetilde{v_\rho^T} = \widetilde{R_T^{-1}}\, \widetilde{v_\rho^T} = \widetilde{R_T^{-1}\, v_\rho^T} = \widetilde{v_\varphi^T}$. The recursion is obtained from these preliminary results. The first two steps of the recursion arise from applying the block-matrix definition of R_{T+1}^{-1} to $R_{T+1}^{-1}\, v_\rho^{T+1} = v_\varphi^{T+1}$. The relative prediction error equals $\sin^2(\psi_T) = 1 - \left\| v_\rho^T \right\|_{R_T^{-1}}^2 = 1 - (v_\varphi^T \mid v_\rho^T)$. The last step of the recursion is obtained with block-matrices of order $T+1$. ■

3.3. Positivity levels for autocovariance functions and their hereditary properties

Propositions 1 and 3 are now extended to a varying length of the history from the block-matrix identities given in (3.22).

Proposition 5. *Consider a dynamic random effect U. If condition (2.10) (resp. conditions (2.10) and (2.11)) holds for $T+1$, then the credibilities are nonnegative (resp. positive) regardless of the length τ of the history and the positive* *frequency risks per period $\lambda_1, \ldots, \lambda_\tau$ if $\tau \leq T$. For a mathematical object, a hereditary property inherits to all the subobjects. The property defined by (2.10) (resp. by (2.10) and (2.11)) is hereditary.*

Due to the stationarity of the random effects, condition (2.10) can also be expressed in terms of the correlation matrices. We write $V_U^T = \gamma_U(0)\, R_U^T$. Consider (3.22) and the northeast block of $\left[R_U^{T+1}\right]^{-1}$. If all of the off-diagonal entries of $\left[R_U^{T+1}\right]^{-1}$ are nonpositive, then the vector of filtering coefficients $v_{\varphi_U}^T$ belongs to the nonnegative orthant. The northwest block of $\left[R_U^{T+1}\right]^{-1}$ in (3.22) is greater than or equal to $\left[R_U^T\right]^{-1}$ (with entrywise inequalities) if $v_{\varphi_U}^T \in \left(\mathbb{R}^+\right)^T$. Then, the off-diagonal elements of $\left[R_U^T\right]^{-1}$ are also nonpositive, a property that allows backward induction. From Proposition 3, the credibilities of a T-period history are nonnegative regardless of the frequency risks. Using backward induction, this nonnegativity property actually holds for every length τ of the history, provided $\tau \leq T$. The hereditarity of this property is obtained. In a strict sense, condition (2.11) applies to the entries associated with a lag equal to ± 1. The lag does not vary in the backward induction and the result also holds in a strict sense. ∎

The sufficient condition for nonnegative credibilities is obtained from the study of three levels of positivity on the autocovariance functions.

- Level N1 specifications refer to nonnegative autocovariance functions. Reaching level N1 is a necessary but not sufficient condition for nonnegative credibilities (see the first polar case in (2.6), and the counterexample that follows).
- Level N2 specifications refer to nonnegative linear filtering. Reaching level N2 is a necessary condition for nonnegative credibilities (see the second polar case in (2.6)). From the Levinson–Durbin recursion used forward, it is easily seen that level N2 specifications reach level N1. The hereditarity of level N2 is established from the Levinson–Durbin recursion used backwards (see the proof of Proposition 6). In addition, level N2 specifications have a nonnegative partial autocorrelation function.
- Level S specifications refer to precision matrices for which the off-diagonal entries are nonpositive. This is a sufficient condition for nonnegative credibilities regardless of the risk exposure. Level S specifications have a hereditary property. The off-diagonal entries of a precision matrix have a sign at the opposite of generalized partial autocorrelation coefficients. Hence, the generalized partial autocorrelation coefficients of level S specifications are nonnegative. Links between the positivity properties of autocovariance functions and the dependence properties of the count variables in Poisson mixtures can be found in Purcaru and Denuit (2003).

3.4. AR(p) specifications with positivity properties

Autoregressive models that reach levels N2 and S are specified in this section. To avoid an intercept, the autoregressive equations are applied to centered sequences. The positive random effect U with an expectation equal to one is replaced by $U - 1 = U^c$. The sequences U and U^c share the same filtering coefficients, except for the intercept. An AR(p) sequence U is defined by $U_t^c = \sum_{1\le h\le p} \varphi_h\, U_{t-h}^c + \varepsilon_t\, \forall t > p$, with $\varphi_p \neq 0$. The sequence ε is white noise and is called the innovation of U.

Proposition 6. *A stationary AR(p) sequence U reaches level N2 if and only if*

$$\varphi_h \geq 0\, \forall h = 1, \ldots, p;\ \sum_{1\le h\le p} \varphi_h < 1. \tag{3.23}$$

The Levinson–Durbin recursion is used backwards in the proof and the hereditarity of level N2 is established (see Appendix B.4).

The next proposition characterizes AR(p) specifications that reach level S.

Proposition 7. *A stationary AR(p) sequence U (with $p \geq 2$) reaches level S if and only if the conditions*

$$\varphi_h \geq \sum_{1\le \tau\le p-h} \varphi_\tau\, \varphi_{\tau+h}\, \forall h = 1, \ldots, p-1 \tag{3.24}$$

supplement those given in (3.23). If the inequality is strict for $h = 1$, then the credibilities are positive for any sequence of positive frequency risks. Levels N2 and S are equivalent for $p = 2$. Furthermore, condition (3.24) follows from (3.23) if the $(\varphi_h)_{h=1,\ldots,p}$ are decreasing.

Specifications that reach level S refer to the nonpositivity of the off-diagonal entries of the precision matrices. Proposition 7 results from the derivation of the precision matrices of AR(p) sequences (see Appendix B.5) and from the hereditarity of level S specifications.

3.5. ARFIMA(0, d, 0) specifications on random effects imply positive credibilities

ARFIMA$(0, d, 0)$ specifications are defined by $(I - L)^d\, U_t^c = \varepsilon_t$, where ε is white noise, L is the lag operator, and d ranges in $]0,\ 1/2[$. The autocorrelation function ρ_d is obtained from the recursion $\rho_d(h) = \frac{h-1+d}{h-d}\, \rho_d(h-1)\, \forall h \in \mathbb{N}^*$. This function is not summable and ARFIMA(0, d, 0) specifications have a long memory. For all d in $]0,\ 1/2[$, we have that

$$U_t^c = \sum_{h=1}^{+\infty} \varphi_h U_{t-h}^c + \varepsilon_t\,;\, \varphi_h \geq 0\, \forall h \in \mathbb{N}^*;\ \sum_{h=1}^{+\infty} \varphi_h = 1; \varphi \text{ is decreasing on } \mathbb{N}^*. \tag{3.25}$$

https://doi.org/10.1017/asb.2020.4 Published online by Cambridge University Press

The long memory property implies $\sum_{h=1}^{+\infty} \varphi_h = 1$. Hence, the intercept is eliminated in (3.25) and U^c can be replaced by U in this equation. Because $\varphi_1 = d$ and $\varphi_{h+1}/\varphi_h = (h-d)/(h+1)\ \forall h \in \mathbb{N}^*$ if γ_U is of the ARFIMA$(0, d, 0)$ type, φ is positive and decreasing. Therefore, $\varphi_h > \sum_{\tau=1}^{+\infty} \varphi_\tau\, \varphi_{\tau+h}\ \forall h \in \mathbb{N}^*$. Propositions 6 and 7 suggest that the ARFIMA$(0, d, 0)$ specifications reach level S in the strict sense. We verify this result numerically from the hereditarity of level S. The correlation matrix of order T and associated with an ARFIMA$(0, d, 0)$ specification is denoted by R_d^T. The off-diagonal entries of $\left[R_d^T\right]^{-1}$ are negative for $T = 100$ and for $d = i/100$, $(i = 1, \ldots, 49)$. From Proposition 5, level S is reached in the strict sense by these ARFIMA$(0, d, 0)$ specifications for individual histories that are shorter than a century.

4. Three Types of Autocovariance Specifications that Ensure Nonnegative Credibilities and that are Compatible with Log-Gaussian Random Effects

4.1. The compatibility issue: A verification strategy with the Levinson–Durbin recursion

Second-order stationary specifications for random effects, which ensure nonnegative credibilities in Poisson mixtures, are given in Sections 3.4 and 3.5. These specifications apply to positive random effects. The compatibility of autocovariance functions with positive random effects is then another positivity issue in credibility analysis. Log-Gaussian sequences are retained in this paper to address this compatibility issue.

The distributions of the stationary Gaussian sequences considered in this section are fully specified from the expectation and the autocovariance function. Given a positive definite autocovariance function γ, Cholesky decompositions on nested variance–covariance matrices that are related to γ provide a sequence of lower matrices. These matrices are then applied to a sequence of independent standard normal variables. A stationary Gaussian sequence is then obtained with the given autocovariance function. Furthermore, entrywise exponentiation provides log-Gaussian sequences that can be used as dynamic random effects. All of these sequences are strictly stationary.

Given a stationary Gaussian sequence W with an autocovariance function γ_W, we have that

$$U \overset{\text{def}}{=} \frac{\exp(W)}{E[\exp(W)]} \Rightarrow \gamma_U = \exp(\gamma_W) - 1 \Rightarrow \gamma_W = \log(1 + \gamma_U). \tag{4.26}$$

These equations must be understood entrywise. Every finite linear combination of the entries of W is Gaussian in this fully specified framework. This result is used in the proof of (4.26), which is detailed in Appendix B.6. The log-Gaussian sequence U that is defined from (4.26) can be used as a positive random effect, with an expectation equal to one. The interplay between semiparametric and fully specified models on second-order stationary random effects is described in Appendix B.7.

In Section 4.2, AR(p) models with positivity properties and the ARFIMA$(0, d, 0)$ specifications are applied to U in (4.26). The compatibility of γ_U with exponentials of Gaussian time series is verified in Section 4.2. From (4.26), this depends on whether or not $\gamma \overset{\text{def}}{=} \log(1+\gamma_U)$ is an autocovariance function. If this is the case, then there exists a Gaussian sequence W such as $\gamma_W = \log(1+\gamma_U)$. The compatibility of γ_U with exponentials of Gaussian time series is then obtained from (4.26). With the Levinson–Durbin recursion, we can verify that $\rho \overset{\text{def}}{=} \gamma/\gamma(0) = \log(1+\gamma_U)/\log(1+\gamma_U(0))$ is an autocorrelation function and conclude to the admissibility of γ as an autocovariance function.

The Levinson–Durbin recursion implies a duality relation between autocorrelation and partial autocorrelation coefficients. Given a sequence ρ of autocorrelation coefficients, the sequence of partial autocorrelation coefficients is denoted by $pac(\rho)$. If ρ is a positive definite autocorrelation function, then all the partial autocorrelation coefficients belong to $]-1, 1[$. Conversely, as long as the entries of $pac(\rho)$ belong to $]-1, 1[$, the corresponding truncations of ρ are positive definite. To see this, consider equation (3.21). If A is positive definite, then so is the block-matrix if $c > b' A^{-1} b = \|b\|^2_{A^{-1}}$. Suppose now that the correlation matrix R_T is positive definite. From equations (3.22) and (3.20) (with the latter written from $T-1$ to T), R_{T+1} is positive definite if and only if $|pac_T| < 1$. If $|pac_T| = 1$ (resp. if $|pac_T| > 1$), then the signature of R_{T+1} is $(T, 0)$ (resp. $(T, 1)$). For example, we have from (3.18): $pac_2 \in]-1, 1[\Leftrightarrow 2\rho_1^2 - 1 < \rho_2 < 1$. Together with $pac_1 = \rho_1 \in]-1, 1[$, we obtain the well-known conditions for R_3 to be positive definite.

4.2. Three types of autocovariance specifications for the credibility analysis of frequency risks

The autoregressive models specified in Section 3.4 are considered first. We apply the verification strategy presented in Section 4.1. Positive definiteness is verified for truncations of $\log(1+\gamma_U)$ on a grid of parameter values, with $T = 100$ (a value larger than any length of an individual history). The positive definiteness of $\log(1+\gamma_U)$ is verified for AR(p) specifications ($p = 1, 2, 3$) on U that reach level N2 and even level N1 for $p = 2$ (see Appendix A.1). These autocovariance specifications are then compatible with log-Gaussian random effects.

The grid of parameter values retained for the numerical verification is described hereafter. From Proposition 6, the AR(p) specifications that reach level N2 are related to a polytope in the space of filtering coefficients. This polytope is the convex hull of the vertices related to the affine canonical basis of $\mathbb{R}^p$. The set of filtering coefficients is equal to this polytope minus the simplex.

https://doi.org/10.1017/asb.2020.4 Published online by Cambridge University Press

Meanwhile, the level S specifications are related to a subset because of the supplementary restrictions defined in Proposition 7 if $p \geq 2$. The polytopes corresponding to $p = 1, 2, 3$ are spanned by grids including the origin, with a mesh size equal to 0.01. We also span an interval related to the variance of the random effect U. This variance is comparable to one and we span the interval $[0.01,\ 5]$ with a mesh size equal to 0.01. With the verification strategy, the truncations of $\log(1+\gamma_U)$ at the horizon of a century are found to be positive definite on the grid.

The parameter grid that is retained for the verifications on ARFIMA$(0, d, 0)$ specifications is defined as follows: the fractional differencing parameter d ranges in $\{i/100\}_{i=1,\ldots,49}$, and the variance of the random effect is spanned as in the autoregressive case. These specifications are compatible with log-Gaussian sequences at the horizon of a century.

A simpler way to obtain log-Gaussian random effects is to apply a given autocovariance specification to a Gaussian sequence W and then to exponentiate entrywise (see (4.26)). We obtain a log-Gaussian sequence U, with an autocovariance function $\gamma_U = \exp(\gamma_W) - 1$. A credibility analysis with U is possible if γ_U reaches level S. If the ARFIMA$(0, d, 0)$ specifications described in the previous paragraph are applied on W, then the random effect U reaches level S at the horizon of a century. The verification is carried out as in Section 3.5. Although the specification of U is no longer of the ARFIMA$(0, d, 0)$ type, the long memory property is maintained.

In this section, three types of autocovariance specifications for random effects applied on Poisson distributions are obtained, with the expected positivity properties. If AR(p) specifications that reach level S are applied to Gaussian sequences W in (4.26), then the resulting sequences U do not reach level S. Therefore, this fourth type of specification is not retained in the paper.

5. A Multiplicative Decomposition of the Random Effects, in Line with the Wold Theorem

The autocovariance functions retained in Section 4 vanish at infinity. However, this ergodicity property is not a desirable assumption (see Table 1, Section 7.1). Placing side by side a time-invariant random effect and a dynamic random effect with a vanishing autocovariance function provides an answer to this issue. The autocovariance specifications presented in this section are in line with the Wold additive decomposition of a stationary random sequence, although a multiplicative specification is needed in the present setting.

Applying Equation (3.16) to an infinite length history and to a centered and stationary sequence X leads to the preliminary equation

$$X_t = \sum_{h=1}^{+\infty} \varphi_h\, X_{t-h} + I_t^X, \text{ with } I_t^X \stackrel{\text{def}}{=} \lim_{T \to +\infty} E_t^T. \tag{5.27}$$

https://doi.org/10.1017/asb.2020.4 Published online by Cambridge University Press

The process I^X is called the innovation of X. As I_t^X is uncorrelated with the past values of X, the sequence I^X is white noise. Iterating (5.27) leads to the Wold decomposition

$$X_t = D_t^X + \sum_{h=0}^{+\infty} c_h \, I_{t-h}^X \stackrel{\text{def}}{=} D_t^X + PND_t^X \,, \tag{5.28}$$

with $c_0 = 1$ and $c \in L^2(\mathbb{N})$. The time series D^X and PND^X are called the deterministic and purely nondeterministic components of X, respectively, which means that $I^{D^X} = 0\,; D^{PND^X} = 0$. The AR($p$) and ARFIMA($0, d, 0$) time series are purely nondeterministic. Meanwhile, deterministic time series are completely predictable. Time-invariant random variables ($D_t = D$) are deterministic. The other deterministic time series are characterized in the frequency domain (see Appendix B.8).

The Wold decomposition is useful to classify stationary time series but it cannot be applied as such to positive random effects. The dynamic random effects $(U_t)_{t\in\mathbb{N}^*}$ applied to Poisson distributions in the credibility analysis are decomposed multiplicatively:

$$U_t \stackrel{\text{def}}{=} P\, Q_t \; \forall t \in \mathbb{N}^*. \tag{5.29}$$

The variables P and $(Q_t)_{t\in\mathbb{N}^*}$ are assumed to be positive and uncorrelated, with $E(P) = E(Q_t) = 1 \; \forall t \in \mathbb{N}^*$. The time-invariant random effect P (unrelated to the symbol P used for precision matrices) is the equivalent of the deterministic component in the Wold decomposition. The other deterministic time series cannot be used as random effects in the credibility analysis because their autocovariance functions oscillate around zero (see Appendix B.8). In addition, the dynamic random effect Q is assumed to be purely nondeterministic. Therefore, the autocovariance function γ_Q vanishes, from (5.28). The dynamic random effect Q is also assumed to reach level S (i.e., the off-diagonal entries of the precision matrix $\left[V_Q^T\right]^{-1}$ are nonpositive for every length T). This is the case if Q follows one of the specifications obtained in Section 4.2. If σ_P^2 is the variance of P, then

$$\gamma_U(h) = \sigma_P^2 + \left[(1 + \sigma_P^2)\, \gamma_Q(h)\right] \; \forall h \in \mathbb{Z}, \tag{5.30}$$

if P^2 and $Q_t Q_{t+h}$ are assumed to be uncorrelated $\forall t, h$. Indeed, $E(U_t U_{t+h}) = E(P^2)\, E(Q_t Q_{t+h}) = (1 + \sigma_P^2)\,(1 + \gamma_Q(h))$, which leads to equation (5.30). Then, $\lim_{h\to+\infty} \gamma_Q(h) = 0 \Rightarrow \lim_{h\to+\infty} \gamma_U(h) = \sigma_P^2$. Owing to the time-invariant component, it is not assumed that γ_U vanishes and this is desirable (see Table 1, Section 7.1).

For a policyholder i, the dynamic component $Q_{i,t}$ of $U_{i,t} = P_i\, Q_{i,t}$ oscillates around $E(Q) = 1$. From the ergodicity assumption on γ_Q, the time averages of Q converge toward one at the individual level. The ergodicity properties of stationary time series are detailed in Appendix B.9.

In Sections 3 and 4, we discuss the positivity properties of the dynamic random effects, which correspond to Q in (5.29). Suppose that Q reaches level S. We want this property to be preserved if a time-invariant random effect is included. The random effect U that is defined in (5.29) also ensures nonnegative credibilities, provided that the following condition holds.

Proposition 8. *From (5.29) and (5.30), the random effect U reaches level S if Q reaches level S and if*

$$\left[V_Q^T\right]^{-1} 1_T \in \left(\mathbb{R}^+\right)^T \ \forall T \in \mathbb{N}^*. \tag{5.31}$$

The intercept is denoted by 1_T in equation (5.31). This positivity result is valid in a wide and in a strict sense. The condition given in (5.31) is fulfilled if Q is $AR(p)$. Moreover, the positivity condition defined in (5.31) is hereditary, as defined in Proposition 5.

The proof is given in Appendix B.10. Equation (5.31) means that the sum of the entries for each line of a precision matrix related to Q is nonnegative. Using hereditarity, condition (5.31) is verified numerically for the semiparametric specifications related to ARFIMA$(0, d, 0)$ models at the horizon of a century and on the parameter grids that were discussed in Section 4.

6. Limit Credibility and the Length of the Memory in the Random Effects

In this section, U is defined from (5.29) and (5.30), and we assume time-invariant frequency risks (i.e., $\lambda_t = \lambda > 0 \ \forall t \in \mathbb{N}^*$). From (3.15), the credibilities per period are obtained from linear filtering of X, with

$$X_t = A_t + (U_t - 1) = A_t + P\,Q_t - 1, \ t \in \mathbb{N}^*. \tag{6.32}$$

The white noise A accounts for risk exposure, with $V(A_t) = 1/\lambda \ \forall t \in \mathbb{N}^*$.

We first give a result on total credibility, denoted by $t_\alpha^T = \sum_{h=1}^T \alpha_{T,h}$ for T periods. It is also assumed that the dynamic random component Q reaches level S, and that condition (5.31) is fulfilled. From Proposition 8, the random effect U implies nonnegative credibilities.

Proposition 9. *If U reaches level S, then the total credibility is less than one and it increases with the length of the history for every annual frequency risk level λ ($\lambda > 0$).*

The Levinson–Durbin recursion applied to X (see (6.32) and (3.19)) implies: $1 - t_\alpha^{T+1} = (1 - pac_{T+1})\,(1 - t_\alpha^T)$, and

$$t_\alpha^T = 1 - \prod_{1 \le h \le T} (1 - pac_h). \tag{6.33}$$

https://doi.org/10.1017/asb.2020.4 Published online by Cambridge University Press

Proposition 9 is obvious from (6.33) if the partial autocorrelation coefficients range within [0, 1[. These coefficients are credibilities related to the last periods of individual histories and they are nonnegative as U reaches level S. The autocovariance function γ_X is positive definite from (3.15). Hence, the partial autocorrelation coefficients are less than one. ■

From Proposition 9, there is a limit credibility $\lim_{T\to+\infty} t_\alpha^T \stackrel{\text{def}}{=} t_\alpha^\infty \leq 1$. Whether the limit credibility is less than or equal to one depends on the length of the memory in the stationary random effect U. The three levels for the length of the memory are described hereafter.

- The short memory case corresponds to a summable autocovariance function. We have: $\gamma_U \in L^1(\mathbb{Z}) \Leftrightarrow \sigma_P^2 = 0$ (i.e., $U = Q$) and $\gamma_Q \in L^1(\mathbb{Z})$. This is the case if Q follows an $AR(p)$ specification that reaches level S.
- The intermediate level for the length of the memory corresponds to an autocovariance function that is not summable (i.e., with a long memory) but vanishes at infinity, hence is weakly ergodic. From (5.30), the limit of γ_U at infinity is equal to σ_P^2. The intermediate level is related to $\sigma_P^2 = 0$ and to $\gamma_Q \notin L^1(\mathbb{Z})$ (e.g., $Q = U$ follows an $ARFIMA(0, d, 0)$ specification).
- The highest level for the length of the memory corresponds to the non-ergodic case, hence to $\sigma_P^2 > 0$. This level includes the basic credibility model, with $U = P,\ Q = 1$.

Proposition 10 determines the limit credibility in the short memory case. The sum of the autocovariances of Q (that are nonnegative) is denoted by $\|\gamma_Q\|_1$. The spectral density of X related to the counting measure on the time domain $\mathbb{Z}$ is denoted by $s_X(\theta) = \sum_{h\in\mathbb{Z}} \gamma_X(h)\, e^{-i\theta h}$ (see Appendix B.8). Hence, $s_X(0) = \sum_{h\in\mathbb{Z}} \gamma_X(h) = (1/\lambda) + \|\gamma_Q\|_1$ as $\gamma_X = (\delta_0/\lambda) + \gamma_U$ and $U = Q$.

Proposition 10. *Suppose that the dynamic random effect Q satisfies $\gamma_Q \in L^1(\mathbb{Z})$, and that $U = Q$ (i.e., $\sigma_P^2 = 0$ in (5.29)). Then, the limit credibility t_α^∞ is less than one.*

If λ ($\lambda > 0$) is the annual frequency risk, then the formula

$$\left(1 - t_\alpha^\infty\right)^2 = \frac{s_{I^X}(0)}{s_X(0)} = \frac{\gamma_{I^X}(0)}{(1/\lambda) + \|\gamma_Q\|_1} \tag{6.34}$$

$$\textit{holds if } |\gamma_{E^T}(h)| \leq C(h)\, \forall T \in \mathbb{N}^*,\ \forall h \in \mathbb{Z},\ \textit{(see (3.16)) with } C \in L^1(\mathbb{Z}). \tag{6.35}$$

The proof is obtained from the definition of credibilities per period as filtering coefficients and from the transformation of spectral densities by linear filtering (see Appendix B.11). The variance of the innovation of X, $\gamma_{I^X}(0)$, satisfies $\gamma_{I^X}(0) \geq \gamma_A(0) = 1/\lambda$. Proposition 10 also provides an upper bound for the limit credibility.

Considering the intermediate level of the memory (i.e., $U = Q$, and Q has a long memory), equation (6.34) suggests that $t_\alpha^\infty = 1$. An example supporting this intuition is given in Section 7.2.

The limit credibility is equal to one for a time-invariant random effect. Not surprisingly, this result is obtained for the highest level of the memory in the random effects.

Proposition 11. *Suppose that $V(P) = \sigma_P^2 > 0$. The limit credibility is then equal to one for any positive value of λ.*

The proof is given in Appendix B.12, using Cesàro sums. Appendix B.13 relates the length of the memory to the spectral measure of U and to the regularity level at 0 (in the frequency domain) of the related distribution function.

7. A Case Study

7.1. Why should we use dynamic random effects?

In this case study, we use the estimated autocovariances of random effects reported by Pinquet *et al.* (2001). To avoid a selection bias, the working sample is an unbalanced panel dataset. The lengths of the histories range between 1 and 7 years. The estimators are derived with a method of moments in a Poisson mixture model with regression components. The estimated autocovariances are given in Table 1.

This unconstrained estimation decreases with the lag, which is a significant empirical result. The credibilities per period have a globally decreasing shape (see Table 2, Pinquet *et al.* (2001)). This property enables the reduction of discrepancies between the actuarial and the real-world experience rating structures. Examples are detailed hereafter.

First is the issue of the duration without claims that is necessary to offset the increase in premium after a claim. With the golden standard credibility model (i.e., with time-invariant random effects) the increase in frequency risk exposure necessary to offset the claim is equal to one. This exposure includes the occurrence of the claim. If the annual frequency risk λ is equal to $1/15$, then 14 years without claims are necessary to offset the increase in premium. This duration ranges between 3 and 7 years in real-world rating structures. Credibilities that decrease with the lag reduce the offset duration. Indeed, a year without claim increases not only risk exposure but also the seniority of the past claims. The double discount effect of time-varying credibilities decreases the offset duration. Meanwhile, policyholders without claims have lower no-claim discounts with time-varying credibilities than with the golden standard credibility model. This results from lower total credibilities if they are derived with dynamic random effects.

The second example relates to total and limit credibility. From Equation (2.3), total credibility is the no-claim discount in the prediction of

TABLE 1

ESTIMATED AUTOCOVARIANCES OF THE MULTIPLICATIVE RANDOM EFFECT U.

h (lag)	0	1	2	3	4	5	6
$\widehat{\gamma_U(h)}$	1.269	0.802	0.615	0.586	0.553	0.457	0.442

TABLE 2

ESTIMATIONS FOR NESTED AUTOREGRESSIVE SPECIFICATIONS OF Q THAT REACH LEVEL S.

Specification of Q	$\widehat{\sigma_P^2}$	$\widehat{\gamma_Q(0)}$	$\widehat{\varphi_1}$	$\widehat{\varphi_2}$	$\widehat{\varphi_3}$	SSE
Absent	0.67	0	0	0	0	0.4973
AR(0)	0.58	0.44	0	0	0	0.0855
AR(1)	0.47	0.54	0.45	0	0	0.0073
AR(2)	0.45	0.56	0.4	0.07	0	0.0055
AR(3)	0.36	0.67	0.42	0.05	0.11	0.0018

frequency risks. Limit credibility is equal to one in the golden standard credibility model, but no-claim discounts in real-world rating structures are far from reaching 100%. This discrepancy is partly explained by the long-term horizon in the derivation of commercial premiums and by information asymmetry between the incumbent insurer and its competitors. Hence, senior policyholders subsidize newcomers and the no-claim discounts of senior policyholders are reduced by commercial premiums. However, limit credibilities are less than one if the random effects have a short memory (see Proposition 10), which brings the actuarial premiums closer to those of the real world.

A last argument to motivate the use of dynamic random effects follows. Premiums can be derived by statistical models with a long-term horizon, if the premium is seen as a control variable in a dynamic programming on the value of the policyholder (Taylor (1986), Emms and Haberman (2005)). Therefore, using the seniority of claims in the prediction enhances the derivations.

7.2. GMMs estimations from estimated autocovariances of the random effects

The AR(p) specifications that reach level S (see Proposition 7) for $p \leq 3$ and applied to Q in (5.29) are estimated in Table 2. The AR(p) models are considered in a wide sense (i.e., $\varphi_p \neq 0$ is not an assumption). Hence, these families are nested. The models are fitted to the unconstrained estimations of autocovariances given in Table 1 with a least squares approach (see Appendix A.3). The sum of the squared errors is denoted by SSE. A white noise process is denoted by AR(0).

The estimations are derived from the grid defined in Section 4.2. The short memory specifications for the dynamic component of U estimated in Table 2

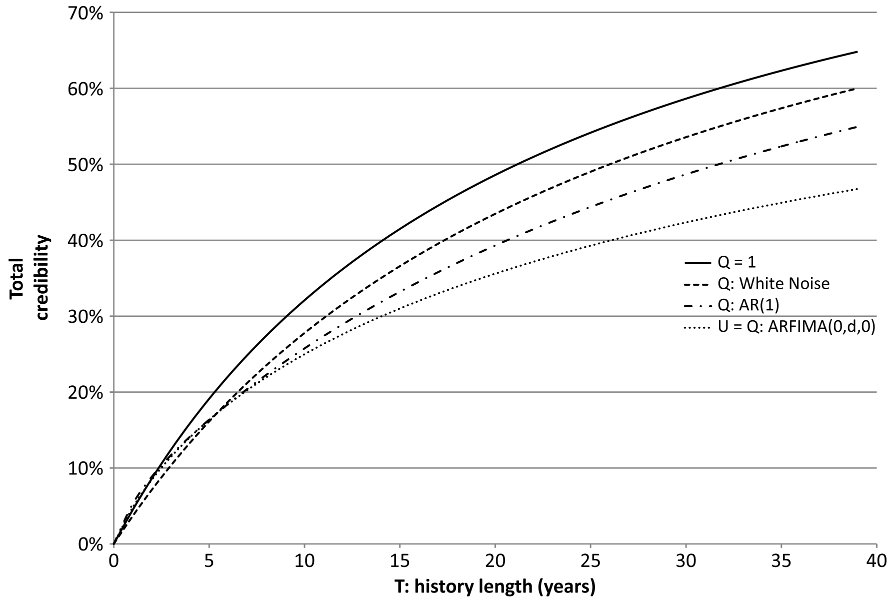


FIGURE 1: Total credibility as a function of the history length and the specification of the dynamic component Q, with $\lambda = 0.07$.

always lead to a positive variance for a time-invariant component P. The fitted specifications of the random effects reach the highest level of memory, which implies a limit credibility equal to one from Proposition 11. The stationary AR(3) specifications are defined from their filtering coefficients in Appendix B.14.

We then estimate the semiparametric specifications of Q that have a long memory. The fitted distribution from the ARFIMA$(0, d, 0)$ specification applied to Q is obtained at the frontier of the parameter set, with $\widehat{\sigma_P^2} = 0\,; \widehat{\gamma_Q(0)} = 1.28\,; \widehat{d} = 0.37; SSE = 0.0067$.

A similar result is obtained if the ARFIMA$(0, d, 0)$ specification is applied to a Gaussian sequence W, with $Q = \exp(W)/E[\exp(W)]$. Then $\widehat{\sigma_P^2} = 0\,; \widehat{\gamma_W(0)} = 0.83\,; \widehat{d} = 0.4; SSE = 0.0087$.

In both cases, the time-invariant component of the random effect U is removed in the fitted specification, and $U = Q$. There is a long memory in the random effects but not at the highest level. In Section 6, we argue that (6.34) suggests a limit credibility equal to one, without providing a proof. If the fitted ARFIMA$(0, d, 0)$ distribution is applied to Q and if $\lambda = 0.07$, then $1 - t_\alpha^T$ and $2.4 \times T^{-0.4}$ are equivalent at infinity. The limit credibility is equal to one for all of the estimations derived in the case study.

Figure 1 compares the evolution of the total credibility ($T \to t_\alpha^T$), with a length T ranging from 1 to 40 years. The annual frequency risk is set to $\lambda = 0.07$. We retain four specifications out of the seven that are estimated in the case study. The dynamic component of the random effect is (a) absent; (b) white noise; (c) AR(1); or (d) ARFIMA$(0, d, 0)$. The first three specifications reach the highest level of the memory in the random effect. For $T \geq 10$, the total credibility derived from the ARFIMA$(0, d, 0)$ specification is the lowest, thanks to the lower level of memory in the random effect.

In the case study, the estimations are obtained in an overidentified framework. Indeed, there are more estimating functions than parameters to estimate: seven estimating functions related to the autocovariances and one to five parameters. Including regression components in a just identified setting with the likelihood equations of a Poisson model maintains an overidentified framework for the Poisson mixtures. In this setting, the GMMs (Hansen (1982)) use the sample mean of the estimating functions, which maps the parameter set onto a differentiable manifold. The parameters are estimated by the GMM together with a metric that is used to determine the minimum distance between the manifold and the origin of the space related to the estimating functions (see Appendix A.3). An efficient GMM estimation would be a refinement of the rough estimation strategy retained in this case study, which avoids derivations at the individual level and is a partial GMM estimation. An analysis of longitudinal count data is presented in Appendix A.3 from different GMM estimations of Poisson mixtures.

GMM estimations of Poisson mixtures combine two advantages: readable moment-based statistics and a constrained estimation approach.

7.3. Between–within derivations and the ergodicity assumption on the random effects

In Section 6, the annual frequency risk based on the observable information is assumed not to vary within the individual histories. In what follows, between–within derivations on longitudinal data are applied to the frequency premiums and the random effects. The total variability of a variable y defined on a panel dataset indexed by $i = 1, \ldots, m$; $t = 1, \ldots, T_i$ splits into a within and a between component, with

$$\sum_{i,t} (y_{i,t} - y_{\bullet\bullet})^2 = \sum_i \sum_t (y_{i,t} - y_{i,\bullet})^2 + \sum_i T_i (y_{i,\bullet} - y_{\bullet\bullet})^2. \tag{7.36}$$

The symbols $\bullet$ are related to averages. Equation (2.1) derived in the model with random effects leads to $E(N_{i,t} \mid U_{i,t}) = \lambda_{i,t}\, U_{i,t}\ \forall i, t$. Frequency premiums derived from a Poisson regression with covariates are denoted by $\widehat{\lambda}_{i,t}$. From the regression that leads to Table 1, the within variability accounts for 24% of total variability if $y = \widehat{\lambda}$ in (7.36).

Between–within decompositions can also be derived on unobservable information. If y is the random effect defined in (5.29), then we have $y_{i,t} = U_{i,t} = P_i\, Q_{i,t}\ \forall i, t$. A time-invariant random effect ($Q_{i,t} = 1\ \forall i, t$) nullifies the within

https://doi.org/10.1017/asb.2020.4 Published online by Cambridge University Press

TABLE 3

LIMITS OF THE BETWEEN AND WITHIN SAMPLE VARIANCES OF THE RANDOM EFFECTS. FOR TWO SPECIFICATIONS DRAWN FROM TABLE 2.

Specification of Q	Q: ARFIMA(0,d,0) $\sigma_P^2 = 0; \gamma_Q(0) = 1.28; d = 0.37$		Q: AR(1) $\sigma_P^2 = 0.47; \gamma_Q(0) = 0.54; \varphi_1 = 0.45$	
Limit of the sample variance	Between	Within	Between	Within
$T = 5$	0.782	0.498	0.796	0.468
$T = 10$	0.649	0.631	0.656	0.608
$T = 20$	0.541	0.739	0.569	0.695
$T = 40$	0.452	0.828	0.521	0.743
$T = +\infty$	$\sigma_P^2 = 0$	$(1 + \sigma_P^2)\,\gamma_Q(0) = 1.28$	$\sigma_P^2 = 0.47$	$(1 + \sigma_P^2)\,\gamma_Q(0) = 0.794$

variability of U, as does the assumption that has been retained in Section 6 for λ.

We consider a virtual and balanced ($T_i = T\ \forall i = 1, \ldots, m$) panel dataset. The limits of the between and within sample variances of U when m goes to infinity are given in Table 3 for various values of T and for two specifications of the random effects drawn from Table 2. These two specifications relate to a) Q : ARFIMA$(0, d, 0)$; b) Q : AR(1).

The limits of the between and within variances of U when T goes to infinity are equal to σ_P^2 and $(1 + \sigma_P^2)\,\gamma_Q(0)$, respectively (see Appendix B.15). The ergodic assumption retained in papers on Poisson models with dynamic random effects (Pinquet *et al.*, 2001; Bolancé *et al.*, 2003; Lu, 2018) is related to $\sigma_P^2 = 0$ and hence implies a null limit for the between variance of U. Consider dynamic random effects as an alternative to the golden standard (time-invariant random effects). If the baby is thrown out with the bath water (i.e., if the time-invariant component is removed in the dynamic specification of the random effect), then the "good guy - bad guy" stratification on unobservable information is erased in the long run by time averages. This is not a satisfactory assumption and the between–within result provides a supplementary motivation to maintain a time-invariant component in the specification of the random effect. However, the between variances of the random effects are far from reaching their null limit in the ergodic specification of Table 3 (Q : ARFIMA$(0, d, 0)$) at the horizon of the life of a policyholder in the portfolio.

7.4. Dispersion and efficiency of the predictors on the portfolio: a comparative study

In this section, we study three predictors associated with a) $Q = 1$; b) Q : AR(1); and c) Q : ARFIMA$(0, d, 0)$, where Q is defined as in (5.29). The working sample is the set of policyholders who reach the maximum length of risk exposure ($T = 7$ years). The number of claims reported in the last year by these 80994 policyholders (see Table A.1, Appendix A.3) is predicted from the first 6

TABLE 4

ELEMENTARY STATISTICS FOR THE LINEAR CREDIBILITY PREDICTORS

	Mean	Std. Dev.	Min.	Max.
$Q = 1$	0.968	0.394	0.595	4.67
Q: AR(1)	0.978	0.360	0.656	5.41
Q: ARFIMA(0,d,0)	0.978	0.352	0.657	5.06

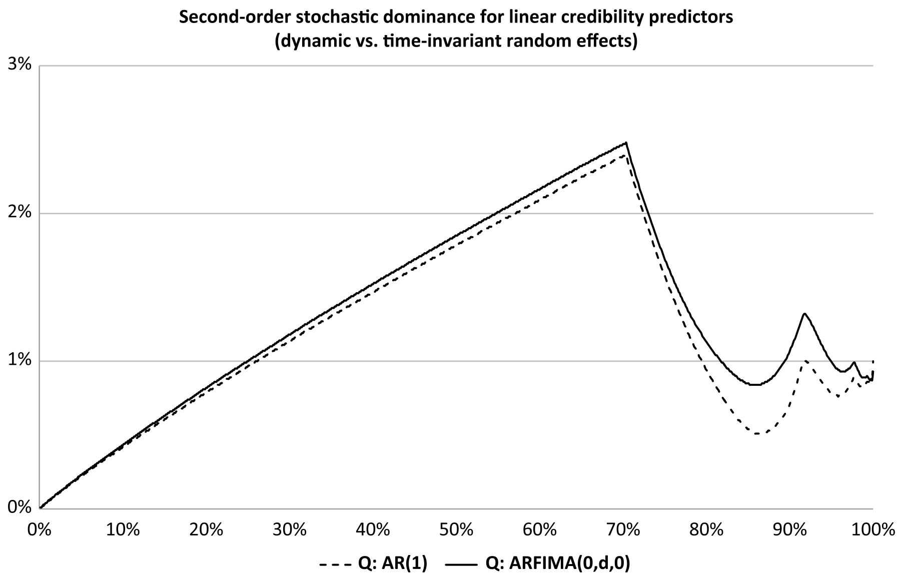


FIGURE 2: Graphs of $p \to \int_0^p [q_{AR}(\alpha) - q_{CST}(\alpha)]\ d\alpha$, $p \to \int_0^p [q_{ARF}(\alpha) - q_{CST}(\alpha)]\ d\alpha$.

years, with the frequency premiums that lead to Table 1. Estimation has been rerun from histories restricted to six periods. Elementary statistics for the linear credibility predictors are given in Table 4.

There is less global dispersion in the predictors related to dynamic random effects. This is anticipated from the lower total credibilities derived from dynamic random effects (see Figure 1). However, the maximum values are higher for the last two specifications. They correspond to policyholders with numerous and recent claims.

The predictors related to dynamic random effects are actually less dispersed than the first predictor (given that their mean is greater) in terms of a second-order stochastic dominance criterion. Let q_{CST}, q_{AR}, and q_{ARF} be the empirical quantile functions of the three predictors. The aforementioned result is obtained if the following conditions hold:

$$\forall p \in [0, 1] : \int_0^p [q_{AR}(\alpha) - q_{CST}(\alpha)]\ d\alpha \geq 0; \int_0^p [q_{ARF}(\alpha) - q_{CST}(\alpha)]\ d\alpha \geq 0.$$

From Figure 2, these conditions are fulfilled.

The two functions for which positivity is verified are almost affine from 0% to 70%, which corresponds to the frequency of policyholders without claims during the first 6 years. The slope corresponds to the difference between quantiles related to time-invariant and dynamic random effects. The sharp downturn that appears afterwards corresponds to policyholders who reported one claim with an important seniority. These claims generate a greater increase in premium with the first rating model than with those based on dynamic random effects. The seniority of the claims decreases when p varies from 70 to 90%. The full picture is actually more intricate, as a given quantile of the three predictors is related to different policyholders. A mean is an integral of quantiles; hence, the values of the functions for $p = 1$ correspond to differences between the means given in Table 4.

The efficiency of a prediction is usually assessed from the receiver operating characteristics (ROC) curves. Experience-rated frequency premiums are obtained as products of a linear credibility predictor of the random effect, and of a prior frequency premium based on covariates. Policyholders are sorted by descending experience-rated premiums. The weights of the policyholders are uniform along the horizontal axis, whereas they are proportional to the risk variable (the number of claims) along the vertical axis. The higher the ROC curve, the more efficient is the prediction. The Gini index summarizes the ROC curve and is defined as for Lorenz curves. The Gini indices corresponding to (a) $Q = 1$; (b) Q : AR(1); and (c) Q : ARFIMA$(0, d, 0)$ are 0.179, 0.182, and 0.183, respectively.

The log-likelihood is another measure of the efficiency of the rating structure. Using the experience-rated frequency premiums as parameters in the Poisson distributions, the log-likelihood increases by 13.84 (resp. by 15.53) if the specification of the random effect goes from $Q = 1$ to Q : AR(1) (resp. from $Q = 1$ to Q : ARFIMA$(0, d, 0)$). Although not applicable here, the likelihood ratio statistic would correspond to a rejection of the null assumption of time-invariance for the random effects.

Addressing the seniority issue with dynamic random effects allows distinguishing between short-term and long-term effects of claims and risk exposure and extrapolating beyond the length of the panel dataset. The shorter durations to offset the increase in premium after a claim that are obtained with dynamic random effects (see Section 7.1) correspond to real-life practices and have good incentive properties (Henriet and Rochet (1986)). The toolbox provided by this study for the credibility analysis of frequency risks, with allowance for the seniority of claims and of risk exposure, is easy to implement. Programs with comments are available in the online folder.

8. Concluding Remarks

For Poisson mixtures, the estimated autocovariances of random effects give the relevant information for the credibility analysis of frequency risks in the time dimension. Random effects with nonnegative generalized partial autocorrelations are shown to imply nonnegative credibilities regardless of the risk exposure. Three types of semiparametric specifications that meet this condition and that are compatible with positive random effects are discussed, with either long or short memory. The related autocovariance functions vanish at infinity, and a time-invariant random effect can be multiplied with these ergodic random effects without loss of the positivity properties. The three levels for the length of the memory in the random effects are reached by the last specifications, which are fitted to estimated autocovariances of random effects with GMMs. The paper also discusses the limit credibility, depending on the length of the memory in the random effects. Although several results are verified numerically but are not proved, the accuracy of the verifications makes these results usable for practitioners.

Acknowledgments

I thank the editors, the referees, participants at the ETH Zürich seminar, Philippe Soulier, and Marie-Cécile Fagart for useful comments. I thank Alain Trognon for his positive appreciation of this paper.

Supplementary Material

Section A is an appendix included in the printed version of the paper. Section B is a mathematical appendix accessible in the online folder. The online folder also contains programs (R, SAS and Mathematica code). They are commented on in a companion file (Section C).

To view supplementary material for this article, please visit https://doi.org/10.1017/asb.2020.4

References

ANGERS, J.-F., DESJARDINS, D., DIONNE, G. and GUERTIN, F. (2018) Modelling and estimating individual and firm effects with count panel data. *ASTIN Bulletin,* **48**(3), 1049–1078.

ASH, R.B. and GARDNER, M.F. (1975) *Topics in Stochastic Processes*, Vol. 27, New York: Academic Press.

BERMAN, A. and PLEMMONS, R.J. (1994) *Nonnegative Matrices in the Mathematical Sciences*, Philadelphia: SIAM.

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

BÜHLMANN, H. (1967) Experience rating and credibility. *ASTIN Bulletin: The Journal of the IAA,* **4**(3), 199–207.

BÜHLMANN, H. and GISLER, A. (2006) *A Course in Credibility Theory and Its Applications*, Heidelberg: Springer.

CRAMÉR, H. and LEADBETTER, M.R. (1967) *Stationary and Related Stochastic Processes: Sample Function Properties and Their Applications*, New York: Courier Corporation.

DENUIT, M., GUILLEN, M. and TRUFIN, J. (2019) Multivariate credibility modelling for usage-based motor insurance pricing with behavioural data. *Annals of Actuarial Science,* **13**(2), 378–399.
DURBIN, J. (1960) The fitting of time-series models. *Revue de l'Institut International de Statistique,* **28**(3), 233–244.
EMMS, P. and HABERMAN, S. (2005) Pricing general insurance using optimal control theory. *ASTIN Bulletin: The Journal of the IAA,* **35**(2), 427–453.
FREES, E.W. (2004) *Longitudinal and Panel Data: Analysis and Applications in the Social Sciences*, Cambridge, UK: Cambridge University Press.
GAO, G., MENG, S. and WÜTHRICH, M.V. (2019) Claims frequency modeling using telematics car driving data. *Scandinavian Actuarial Journal,* **2019**(2), 143–162.
HANSEN, L.P. (1982) Large sample properties of generalized method of moments estimators. *Econometrica,* **50**(4), 1029–1054.
HENRIET, D. and ROCHET, J.-C. (1986) La logique des systèmes bonus-malus en assurance automobile: une approche théorique. *Annales d'Économie et de Statistique*, **1**, 133–152.
LEMAIRE, J. (2013). *Automobile Insurance: Actuarial Models*, Vol. 4, New York: Springer.
LEVINSON, N. (1946) The Wiener (root mean square) error criterion in filter design and prediction. *Studies in Applied Mathematics,* **25**(1–4), 261–278.
LIANG, K.-Y. and ZEGER, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika,* **73**(1), 13–22.
LU, Y. (2018). Dynamic frailty count process in insurance: A unified framework for estimation, pricing, and forecasting. *Journal of Risk and Insurance,* **85**(4), 1083–1102.
NELDER, J.A. and WEDDERBURN, R.W. (1972) Generalized linear models. *Journal of the Royal Statistical Society: Series A (General),* **135**(3), 370–384.
PINQUET, J., GUILLÉN, M. and BOLANCÉ, C. (2001) Allowance for the age of claims in bonus-malus systems. *ASTIN Bulletin,* **31**(2), 337–348.
PURCARU, O. and DENUIT, M. (2003) Dependence in dynamic claim frequency credibility models. *ASTIN Bulletin: The Journal of the IAA,* **33**(1), 23–40.
TAYLOR, G.C. (1986) Underwriting strategy in a competitive insurance environment. *Insurance: Mathematics and Economics,* **5**(1), 59–77.
WÜTHRICH, M.V. (2017) Covariate selection from telematics car driving data. *European Actuarial Journal,* **7**(1), 89–108.

JEAN PINQUET
Department of Mathematics and Economics
University of Paris Nanterre
200, avenue de la République
92001 Nanterre Cedex, France
E-mails: pinquet@parisnanterre.fr; jpinquet@free.fr

A. APPENDIX

A.1. Specifications of type AR(2)

Nota Bene: this appendix uses results presented in Sections 2–4.

Specifications of type AR(2) are represented in Figure A.1 by the stationarity triangle defined by the constraints $|\varphi_2| < 1;\ |\varphi_1| < |1-\varphi_2|$. Positive definiteness is equivalent to the conditions $pac_1,\ pac_2 \in\ ]-1,\ 1[$ on the partial autocorrelation coefficients. The stationarity triangle is obtained from $pac_2 = \varphi_2,\ pac_1 = \varphi_1/(1-\varphi_2)$. The latter result is obtained

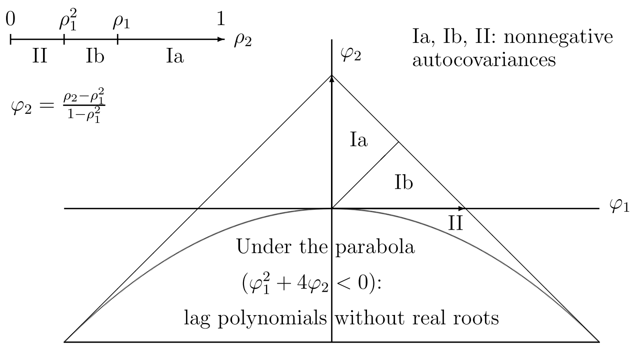


FIGURE A.1: Stationarity triangle for AR(2) time series.

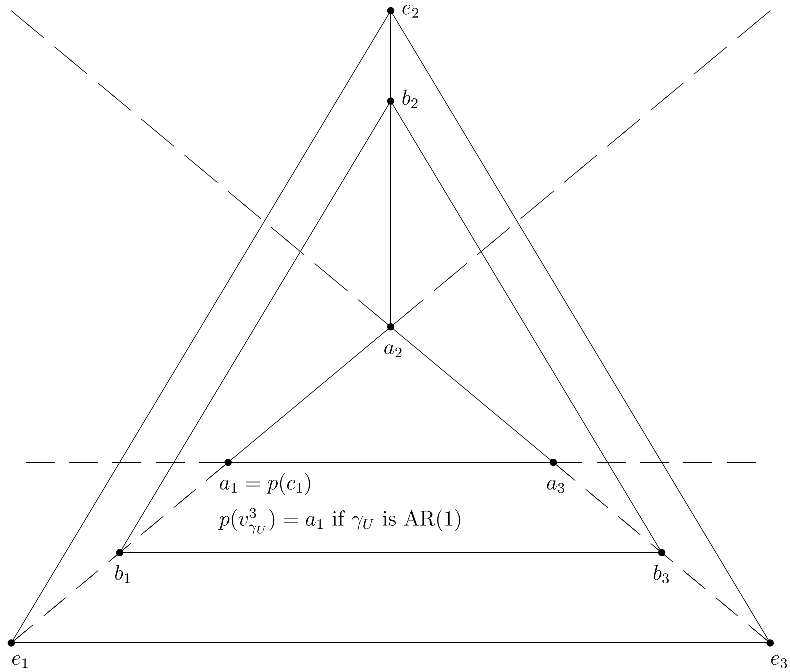
from the Levinson–Durbin recursion used backwards. Region I contains specifications that ensure nonnegative credibilities by Propositions 1, 3, 6, and 7. It is split into Regions Ia and Ib. In Region Ia, the autocovariance function is not decreasing for positive lags, whereas this is the case in Region Ib. Region II is linked with the example given in Section 2 to stress the weak point of the linear credibility approach. The autocovariance functions are similar to those of Region Ib (they are positive and decreasing for positive lags), but the credibility for the second lag of a 2-year history is negative if the frequency risk is large enough. We verify that the autocovariance functions are compatible with log-Gaussian random effects in the three regions. In the verification, the grid for (φ_1, φ_2) defined in Section 4.2 is extended to Region II. Specifications in Region II may be thought of for experience rating from posterior distributions of log-Gaussian random effects, but linear credibility is not recommended.

A.2. Proof of Proposition 3

A formal proof of Proposition 3 is given in Appendix B.3. This section comments an example, in order to explain the geometrical approach.

Comments on Figure A.2: in this example, $T=3$. We use the vocabulary of affine and projective geometry. Consider the simplex S, i.e., the convex hull of the canonical vector basis $\{e_1, e_2, e_3\}$ of $\mathbb{R}^3$. All the vectors $x \neq 0$ in the nonnegative orthant are projected onto the simplex with $p(x) = \mathbb{R}x \cap S = x/s(x)$, where $s(x)$ is the sum of the components of x. In Figure A.2, we have $a_h = p(c_h)$ $(h = 1, 2, 3)$, where c_h is the h^{th} column of the variance–covariance matrix of the random effects. The vector of the covariances of random effects between the prediction period and the history is denoted by $v^3_{\gamma U}$. The convex hull $CH(a_1, a_2, a_3)$ is the set of points with nonnegative barycentric coordinates in the affine basis $\{a_1, a_2, a_3\}$. Nonnegative linear filtering means that $p(v^3_{\gamma U}) \in CH(a_1, a_2, a_3)$. The nonpositivity of the off-diagonal entries of the precision matrix means that e_h is at the opposite of a_h with respect to the $\{a_\tau\}_{\tau \neq h}$ $\forall h = 1, 2, 3$. In Figure A.2, this means that e_h belongs to the cone delimited with dashed lines and with a vertex equal to a_h $(h = 1, 2, 3)$. The same property holds if e_h is replaced by b_h, with $b_h \in]e_h, a_h[$ $\forall h = 1, 2, 3$. In that case, $CH(a_1, a_2, a_3) \subset CH(b_1, b_2, b_3)$. This result is obtained with the tools of matrix calculus used in the proof of Proposition 1, and the proof is actually simpler in this case. The signs of the credibilities are the same entrywise as those of the barycentric coordinates of $p(v^3_{\gamma U})$ in an affine basis $\{b_1, b_2, b_3\}$, with $b_h \in]e_h, a_h[$ $\forall h = 1, 2, 3$. The weights that define b_h as a barycenter of e_h and a_h $(h = 1, 2, 3)$ are given in Figure A.2, and they include the

https://doi.org/10.1017/asb.2020.4 Published online by Cambridge University Press



$T = 3$; (e_1, e_2, e_3) : canonical basis of $\mathbb{R}^3$; $V^3_U = (c_1\, c_2\, c_3)$; $a_h = p(c_h)$ $(h = 1, 2, 3)$;
$p(x) = x/s(x)$, where $s(x)$ is the sum of the entries of x; $v^3_{\gamma_U} = \underset{1\le h\le 3}{vec}(\gamma_U(h))$;
$b_h = \pi_h e_h + (1 - \pi_h)a_h$, with $1/\pi_h = 1 + (\lambda_{T+1-h}\, s(c_h))$ $(h = 1, 2, 3)$.

FIGURE A.2: A geometrical proof of Proposition 3.

time-varying frequency risks. As $p(v^3_{\gamma_U}) \in CH(a_1,\ a_2,\ a_3) \subset CH(b_1,\ b_2,\ b_3)$, the credibilities are nonnegative.

The example described in Figure A.2 is further detailed in Appendix B.3.

A.3. Estimations using the generalized method of moments (GMM) of Poisson mixtures for longitudinal count data

In this section, we detail the implementation of the GMM on Poisson mixtures for longitudinal count data. The estimations of the case study are extended to models that take into account risk exposure, with or without reestimation of the Poisson model.

The parameter of the Poisson model has, for instance, a log-linear specification, with

$$N_{i,t} \sim P(\lambda_{i,t}),\ \lambda_{i,t} = d_{i,t}\ \exp(\beta\, x_{i,t})\, i = 1, \ldots, m;\, t = 1, \ldots, T_i.$$

The regression components are represented by a column vector and the parameters by the line vector $\beta,\ \beta \in \mathbb{R}^k$. A duration $d_{i,t}$ of risk exposure is included (the default duration is one year). The maximum length of the time series in the portfolio is denoted by $T_{\max} = \max_{1\le i\le m} T_i$ ($T_{\max} = 7$ in the case study). The autocovariance function of U is parameterized by α. Then, the vector of parameters in the Poisson model with random effects is denoted by $\theta = (\beta, \alpha)$.

There is no relation with the symbol θ used for the analysis in the frequency domain, and α is unrelated to the credibility weights. Let p be the dimension of α (e.g., $p=3$ if Q follows an ARFIMA(0, d, 0) specification with $\alpha=(\sigma_P^2, \gamma_Q(0), d)$). The corresponding parameter set A has a boundary in $\mathbb{R}^p$. The global parameter set $\Theta=\mathbb{R}^k \times A$ also has a boundary in $\mathbb{R}^{k+p}$. If $\gamma_{U,\alpha}$ is the fitted autocovariance function, some useful moments in the model with random effects follow:

$$E(N_{i,t}\mid\theta)=\lambda_{i,t}(\beta);\ V(N_{i,t}\mid\theta)=\lambda_{i,t}(\beta)+\left[\lambda_{i,t}^2(\beta)\,\gamma_{U,\alpha}(0)\right];$$

$$Cov(N_{i,t}\ ,\ N_{i,t-h}|\,\theta)=\lambda_{i,t}(\beta)\,\lambda_{i,t-h}(\beta)\,\gamma_{U,\alpha}(h)\,(1\le h<t).$$

We denote $N_{i,t}-\lambda_{i,t}(\beta)$ by $res_{i,t}(\beta)$. A matrix $F(\theta)$ of $k+T_{\max}$ estimating functions is defined on the panel dataset as follows (the superscript denotes the column index, and the $n=\sum_{i=1}^{i=m} T_i$ lines of F are indexed by i, t):

$$F_{i,t}^{j}(\theta)=res_{i,t}(\beta)\,(x_{i,t})_j\ (j=1,\ldots,k); \tag{A.1}$$

$$F_{i,t}^{k+1}(\theta)=(res_{i,t}(\beta))^2-N_{i,t}-\lambda_{i,t}^2(\beta)\,\gamma_{U,\alpha}(0); \tag{A.2}$$

$$F_{i,t}^{k+1+h}(\theta)=(res_{i,t}(\beta)\,res_{i,t-h}(\beta))-(\lambda_{i,t}(\beta)\lambda_{i,t-h}(\beta)\,\gamma_{U,\alpha}(h))\,(1\le h<t); \tag{A.3}$$

$$F_{i,t}^{k+1+h}(\theta)=0\,(t\le h<T_{\max}=7). \tag{A.4}$$

The first block corresponds to the contribution of the pair (i, t) to the likelihood equations in the Poisson model. The second block is related to the unconstrained estimation of the variance of a time-independent random effect. The third and fourth blocks are related to moment estimators for autocovariances with positive lags. The sample mean of the estimating functions $\frac{1}{n}\sum_{i,t} F_{i,t}(\theta)=\overline{F(\theta)}$ nullifies in $\mathbb{R}^{k+T_{\max}}$ if β is replaced by the maximum likelihood estimation in the Poisson model and if the $\gamma_{U,\alpha}(h)$ $(h=0,1,\ldots,T_{\max}-1)$ are replaced by the unconstrained estimators of the autocovariances that lead to Table 1.

Cars could be connected to the internet of things in the near future. Then, the incumbent insurer could have access to the driver's behavior in continuous time (see Wüthrich (2017), Denuit *et al.* (2019) and Gao *et al.* (2019) for an analysis of this type of data). Data will then be organized as risk exposure spells (car trips). In that case, the duration of risk exposure will be advantageously replaced by mileage as an offset variable in the Poisson regression. The previous equations can be adapted to such data as follows. The risk exposure spells replace the periods. The calendar date $d_{i,t}^{cal}$ is related to the beginning of the spell indexed by (i, t) and increases with t, $\forall i=1,\ldots, m$. Let D denote a duration unit equal to 1 year or less. Two spells indexed by (i, t_1) and (i, t_2) (with $t_1\le t_2\le T_i$) generate a lag h, equal to the integer part of $(d_{i,t_2}^{cal}-d_{i,t_1}^{cal})/D$. Then, the estimating functions described previously can be used.

The model is overidentified if there are more estimating functions than parameters to estimate ($k+T_{\max}>k+p \Leftrightarrow T_{\max}>p$). This is the case for all the models estimated in Section 7.2, with $T_{\max}=7$ and $p\le 5$. Just identified models ($p=T_{\max}$) are also feasible. The expectation of $F_{i,t}(\theta)$ nullifies for a Poisson mixture associated with $\theta=(\beta,\alpha)$. The sample mean of the estimating functions $\theta\rightarrow\overline{F(\theta)}$ maps the parameter set Θ onto a manifold of dimension $k+p$ with a boundary and embedded in the space $\mathbb{R}^{k+T_{\max}}$ related to the estimating functions. Let W be a positive definite matrix of order $k+T_{\max}$. The generalized method of moments estimator $\widehat{\theta}^W$ minimizes the distance defined by W between the origin of $\mathbb{R}^{k+T_{\max}}$ and the manifold. Hence

$$\widehat{\theta}^W=\arg\min_{\theta\in\Theta}\ \|M(\theta)\|_W^2\ ,\ \text{with } M(\theta)=\overline{F(\theta)}.$$

https://doi.org/10.1017/asb.2020.4 Published online by Cambridge University Press

For example, the model is just identified for the simplest Poisson mixture (with a time-invariant random effect) if the estimating functions related to (A.3) and (A.4) are removed. If W is block diagonal with respect to β and α (with $\alpha = \sigma_P^2$), we have

$$\widehat{\sigma_P^2}^W = \left[\frac{\sum_i \sum_{1 \le t \le T_i} (N_{i,t} - \lambda_{i,t}(\widehat{\beta}^0))^2 - N_{i,t}}{\sum_i \sum_{1 \le t \le T_i} \lambda_{i,t}^2(\widehat{\beta}^0)} \right]^+,$$

where $\widehat{\beta}^0$ is the maximum likelihood estimator of β in the Poisson model. The GMM estimator is the positive part of the usual unconstrained moment-based estimator of σ_P^2. The GMM estimator nullifies in the case of underdispersion.

Statistical inference in a GMM framework deals with the choice of the estimating functions and of the metric defined by W. From the estimating functions, we write W as a block matrix

$$W = \begin{pmatrix} W_{\beta\beta} & W_{\beta\alpha} \\ W_{\alpha\beta} & W_{\alpha\alpha} \end{pmatrix},$$

where α is related to the last three blocks of columns of $F(\theta)$. The estimation retained in the case study is of the GMM type. If $\widehat{\beta}^0$ maximizes the likelihood in the Poisson model, then consider a block diagonal matrix W^0 (i.e., $W^0_{\beta\alpha} = {}^t W^0_{\alpha\beta} = 0$), with

$$W^0_{\alpha\alpha} = \underset{0 \le h < T_{\max}}{diag} \left(\frac{1}{\left[\sum_{i \mid T_i > h} \sum_{h < t \le T_i} \lambda_{i,t}(\widehat{\beta}^0) \lambda_{i,t-h}(\widehat{\beta}^0) \right]^2} \right). \tag{A.5}$$

Then $\widehat{\theta}^{W^0} = (\widehat{\beta}^0, \widehat{\alpha}^0)$, where $\widehat{\alpha}^0$ is the rough estimation retained in the case study—which avoids individual data. As the risk exposure decreases with the lag, the weights used by the rough estimation increase with the lag.

If $E[F(\theta)] = 0$ (the random variables are latent), the efficient metrics W belong to $\mathbb{R}^{+*} \left(E[{}^t F(\theta) F(\theta)] \right)^{-1}$. Denoting

$$W(\theta) = \left[\frac{1}{n} \sum_{i,t} {}^t F_{i,t}(\theta)\, F_{i,t}(\theta) \right]^{-1} = \left[\overline{{}^t F(\theta)\, F(\theta)} \right]^{-1}, \; n = \sum_{i=1}^{i=m} T_i,$$

an asymptotically efficient GMM estimation is obtained from the minimization of

$$g(\theta) = \| M(\theta) \|^2_{W(\theta)} = \left\| \overline{F(\theta)} \right\|^2_{\left[\overline{{}^t F(\theta)\, F(\theta)} \right]^{-1}}. \tag{A.6}$$

The GMM approach is widely implemented in statistical software. A popular alternative approach to estimate mixture models is the generalized estimating equations (GEE) method (Liang and Zeger (1986)). The estimation of α and β is obtained from a round trip between these two vectors. For a given value of α, the vector β is estimated in a generalized linear model framework (Nelder and Wedderburn (1972)). The round trip is closed by nullifying estimating functions in a just identified setting. The vector α is then a function of the observations and of β. An advantage of the GEE with respect to the GMM is that optimization is avoided. The function to minimize in (A.6) is highly nonlinear and the GMM requires cumbersome numerical derivations. However, the GMM approach outperforms the GEE in several ways.

TABLE A.1

DISTRIBUTION OF THE LENGTH OF THE POLICYHOLDERS HISTORIES.

T	1	2	3	4	5	6	7
m_T	42880	33750	29497	29005	28720	24542	80994

TABLE A.2

NUMBER OF INDIVIDUALS THAT CONTRIBUTE TO THE ESTIMATION OF THE AUTOCOVARIANCES.

h (lag)	0	1	2	3	4	5	6
$n_h = \sum_{h<T\leq 7} (T-h)\, m_T$	1172701	903313	676805	484047	320786	186530	80994

- All of the estimating functions at hand can be used in an overidentified framework, and no information is lost by a GMM estimation. Conversely, a choice must be made with a GEE estimation. The estimation of α with the GEE if Q is of the ARFIMA$(0, d, 0)$ type would entail a choice of three estimating equations out of the seven at hand, and a subsequent loss of information.
- The estimator of α with the GEE is unconstrained, but the parameter space has a boundary. For instance, the constrained estimated variance of the time-invariant random effect nullifies in the case study if an ARFIMA$(0, d, 0)$ specification is retained for the dynamic random effect. With the GEE, the estimator of the variance would be negative. The estimation would have to be rerun without this parameter.
- With the generalized linear model setting in which β is estimated for a given value of α, it is implicitly assumed that α parameterizes a variance–covariance matrix. The GMM is not restricted to second-order semiparametric analysis.
- A last advantage of the GMM is its simple geometric interpretation.

An efficient GMM estimation uses data at the individual level and takes into account the risk exposure related to each lag. The link between the lag and risk exposure is described on the database used in Pinquet *et al.* (2001) in what follows. The number of policyholders with a T period history ($T = 1, \ldots, T_{\max} = 7$) is denoted by m_T. From Table A.1, the number of policyholders is $m = \sum_{T=1}^{T=7} m_T = 269388$. The individuals in the statistical analysis are pairs policyholder-period, indexed by (i, t). The number of individuals is $n = \sum_{T=1}^{T=7} T\, m_T = 1172701$.

From this table, we count the individuals that contribute to the estimation of the autocovariances, depending on the lag h. A policyholder i contributes to the estimation of $\gamma_U(h)$ or $\gamma_{U,\alpha}(h)$ in the matrix $F(\theta)$ if $T_i > h$, and the number of contributing lines is $T_i - h$. The total number n_h of lines of the matrix that contribute to this estimation is $\sum_{h<T\leq 7} (T-h)\, m_T$. The results are reported in Table A.2.

The risk exposure strongly decreases with the lag. Efficient GMM procedures take this fact into account.

An estimation strategy that ranges between the rough and the comprehensive GMM estimation consists in keeping unchanged the parameters related to the first-order moments. The frequency premiums $\widehat{\lambda}_{i,t} = \widehat{E}\left(N_{i,t} \mid x_{i,t}\right)$ that are derived from a Poisson regression or from an alternative strategy (neural network, regression tree, or random forest) are treated as constants. The estimating functions given in (A.2), (A.3), and (A.4) are then used with the frequency premiums and the residuals $n_{i,t} - \widehat{\lambda}_{i,t}$. Implementations are commented on in Appendix C.12. We obtain, for instance:

- Time-invariant random effect ($Q = 1$): $\widehat{\sigma_U^2} = 0.72$.
- If Q is AR(1) : $\widehat{\sigma_P^2} = 0.51; \widehat{\sigma_Q^2} = 0.5; \widehat{\varphi_1} = 0.39$.
- If Q is ARFIMA$(0, d, 0)$: $\widehat{\sigma_P^2} = 0.01; \widehat{\sigma_Q^2} = 1.27; \widehat{d} = 0.37$.

The estimated variance of a time-invariant random effect is greater than 0.67, that is, the rough estimation obtained in Table 2. This result is not surprising from the comment following Table A.2. An efficient estimation overweights the adjustment errors related to the first lags. As the estimated autocovariance function decreases, the efficient GMM estimation of the variance is greater than the corresponding rough estimation, which relies on uniform weights and equals the mean of the autocovariances.

In this paper and in Pinquet *et al.* (2001), we have obtained, from a given regression, no less than four estimations of the variance of a time-invariant random effect. These are:

$$\widehat{\sigma_U^2}^a = \frac{\sum_{i,t} (n_{i,t} - \widehat{\lambda}_{i,t})^2 - n_{i,t}}{\sum_{i,t} \widehat{\lambda}_{i,t}^2} = 1.269; \widehat{\sigma_U^2}^b = \frac{\sum_i (n_{i,\circ} - \widehat{\lambda}_{i,\circ})^2 - n_{i,\circ}}{\sum_i \widehat{\lambda}_{i,\circ}^2} = 0.779,$$

and the two GMM estimators (rough and efficient, respectively, equal to 0.67 and 0.72).

The estimator $\widehat{\sigma_U^2}^b$ is based on numbers of claims and frequency premiums summed at the policyholder level ($n_{i,\circ} = \sum_t n_{i,t}\,; \widehat{\lambda}_{i,\circ} = \sum_t \widehat{\lambda}_{i,t}$). Disentangling the numerator and denominator with respect to the lag expresses $\widehat{\sigma_U^2}^b$ as a weighted average of the estimated autocovariances. This estimator is greater than the rough GMM estimator because the weights related to $\widehat{\sigma_U^2}^b$ decrease with the lag (see the arguments following Table A.2), whereas the rough GMM estimator uses uniform weights.

The first estimation markedly exceeds the other ones because it reflects a short-term effect. Distinguishing short-term from long-term effects requires dynamic random effects in the mixture model.

https://doi.org/10.1017/asb.2020.4 Published online by Cambridge University Press