C Comments on the programs

The lines of the Word file `programs.docx` are numbered. This document refers to these numbers. The Word file has an *alter ego* in text format (file `programs.txt`). The programs and comments are sorted by their order of appearance in the paper and the appendices.

C.1 ARFIMA$(0, d, 0)$ specifications reach level S in the strict sense

Lines 5 to 47: This `R` program is related to Section 3.5. I verify that the ARFIMA$(0, d, 0)$ specifications reach level S in the strict sense (i.e., credibilities per period are positive regardless of the risk exposure). Precision matrices are derived first, then the related correlation matrices. The program verifies that all the off-diagonal generalized partial autocorrelation coefficients are positive (see Section 2.4). From stationarity, the variance–covariance matrices of the random effects are Toeplitz. The Toeplitz property is lost in the inversion, which justifies the adjective "generalized."
Figure 8 gives the minimum value of the off-diagonal generalized partial autocorrelation coefficients as a function of the partial differencing coefficient d. The length of the history is $T = 100$. From hereditarity, level S is reached for histories shorter than a century. The mesh size for d equals 0.01.

C.2 Compatibility between autocovariance functions with positivity properties and exponentials of stationary Gaussian vectors

The programs are related to Section 4.2.

Let γ be an autocovariance function. If $\rho = \log(1+\gamma)/\log(1+\gamma_0)$ is an autocorrelation function, then γ is followed by the exponential of a stationary Gaussian vector. To verify (or to prove) the admissibility of ρ as an autocorrelation function, the natural approach is the Herglotz-Bochner theorem, linking admissibility with the nonnegativity of the Fourier transform.

The paper does not retain this approach. Fourier transforms ususally do not have a closed form. If a numerical approach is retained (discrete Fourier transform, fast Fourier transform), then an approximation of the spectral density is obtained on a grid defined in the frequency domain. To reach substantial results, an approximation framework has to be built, and this is taxing.

In my paper, the verification strategy remains in the time domain. The Levinson–Durbin recursion takes an autocorrelation function as an input, and generates partial autocorrelation coefficients as an output. Other outputs are the other filtering coefficients, and the accuracy of the prediction. I focus on the first output. The Levinson–Durbin recursion can be used backwards: as long as the entries of the output range in]-1, 1[, the truncated input is a positive definite autocorrelation function. A verification is less than a proof. A proof could stem from the definition of]-1, 1[as an attraction basin with a Lyapunov function. However, the partial autocorrelation coefficients depend on all the previous values. Using a dynamical system framework to solve the problem is not obvious.

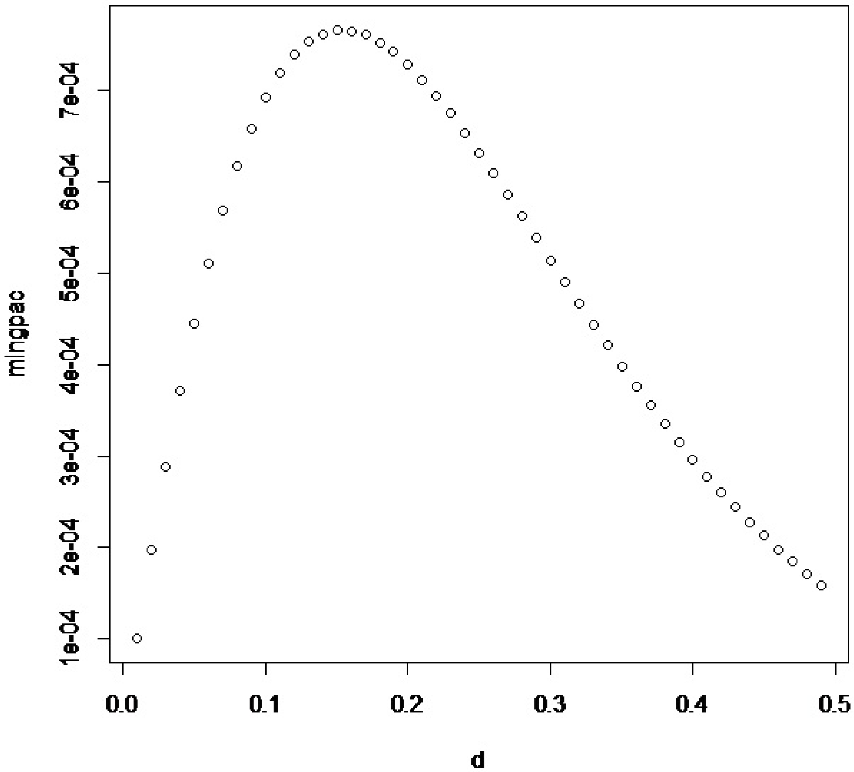

Figure 8: Minimum of the off-diagonal generalized partial autocorrelation coefficients (as a function of d, for $T = 100$).

Lines 53 to 87: an example from an AR(1) specification with $\gamma = 2$ and $\rho_1 = 0.5$ provides the ten first entries of the output of the Levinson–Durbin recursion (variable `pacub`). They all range in]-1, 1[, hence the ten first entries of the AR(1) specification are compatible with a stationary log-Gaussian sequence. The partial autocorrelation coefficients oscillate around zero. Results suggest that the whole sequence vanishes. It is worth noting that the sequence $pac(\rho)$ vanishes at infinity if ρ is an autocorrelation function with a positive innovation. Indeed, the accuracy of linear filtering is updated by the Levinson–Durbin recursion with

$$\sin^2(\psi_{T+1}) = \sin^2(\psi_T) \times (1 - pac_{T+1}^2).$$

Hence, $\sin^2(\psi_\infty) = \prod_{h=1}^{+\infty}(1 - pac_h^2)$, where $\sin^2(\psi_\infty)$ is the ratio between the variance of the innovation and the variance of the sequence. If $\sin^2(\psi_\infty) > 0$ (i.e. if the sequence is non deterministic in Wold's parlance), then the partial autocorrelations are square summable and hence vanish at infinity.

Lines 89 to 149: loop on the filtering coefficient φ_1 and on the variance of the random effect in the verification for AR(1) specifications with positive autocorrelations. The variance γ_0 is spanned in the interval $]0,\ \gamma_0^{\max} = 5]$ with a mesh size termed `gammesh` in the program. The filtering coefficient φ_1 is spanned in the interval $]0, 1[$ with a mesh size termed `phimesh` in the program. The maximum length of the history is the variable `length`. The output of the program is that the maximum absolute value of $pac(\rho)$ is always reached for $h = 1$. Hence the admissibility condition is always fulfilled, as $[pac(\rho)]_1 = \rho_1 = \frac{\log(1+\gamma_0\varphi_1)}{\log(1+\gamma_0)}$. A better understanding of the link between ρ and $pac(\rho)$ would make these verifications useless.

The verifications are also achieved for AR(p) specifications, with $p = 2, 3$, and for specifications reaching the N2 level (nonnegative linear filtering). For $p = 2$, the verification includes Region II in Figure 3 (Section A.1) that corresponds to the counterexample given in the paper. In this region, experience rating can be thought of (the autocovariances are nonnegative, and level N1 is reached), but linear credibility is not recommended.

Lines 152 to 228: verification program for AR(2) specifications with nonnegative autocovariances. The notations are similar to those for the AR(1) specifications.

Lines 230 to 295: verification program for AR(3) specifications with nonnegative filtering coefficients. The notations are different (older version). The filtering coefficients related to stationary AR(3) specifications are described in Section B.14.

For the AR(p) specifications with nonnegative filtering coefficients ($p = 1, 2, 3$) and for the AR(p) specifications that reach level N1 (with $p = 2$), the maximum absolute value of $pac(\rho)$ is always reached for $h = 1, \ldots, p$ and is less than one.

The Yule-Walker equations are used to generate the autocorrelations from the filtering coefficients. For $p = 2$, we have:

$$\begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} \Leftrightarrow$$
$$\begin{pmatrix} 1-\varphi_2 & 0 \\ -\varphi_1 & 1 \end{pmatrix} \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} = \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix},$$

hence

$$\rho_1 = \frac{\varphi_1}{1-\varphi_2}.$$

Then the autocorrelation sequence is extended with Yule-Walker.
The result is actually obvious with $\rho_1 = \varphi_1\rho_0 + \varphi_2\rho_{-1}$, and $\rho_0 = 1$; $\rho_{-1} = \rho_1$. The matrix equations can be used for higher orders. For the AR(3) specifications, we have

$$\begin{pmatrix} 1-\varphi_2 & -\varphi_3 & 0 \\ -\varphi_1-\varphi_3 & 1 & 0 \\ -\varphi_2 & -\varphi_1 & 1 \end{pmatrix} \begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{pmatrix} = \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \end{pmatrix}.$$

This equation provides ρ_1, ρ_2, ρ_3 as a function of φ_1, φ_2, φ_3. Then the autocorrelation sequence is extended with Yule-Walker.

For the AR(2) specifications, with `phimesh`=0.01, a maximum value of γ_0 equal to 5 and `length`=100, the maximum absolute of value of $pac(\rho)$ equals 0.9953. This is reached for the two following values:

$$\varphi_1 = 0.99;\ \varphi_2 = 0;\ \gamma_0 = \gamma_0^{\max} = 5;\ h = 1;\ [pac(\rho)]_1 = 0.9953.$$

$$\varphi_1 = 0;\ \varphi_2 = 0.99;\ \gamma_0 = \gamma_0^{\max} = 5;\ h = 2;\ [pac(\rho)]_2 = -0.9953.$$

Lines 297 to 356: verification program for the ARFIMA$(0, d, 0)$ specifications. The partial autocorrelation coefficients of $\rho = \log(1 + \gamma)/\log(1 + \gamma_0)$ are positive, decreasing, and vanish at infinity for every value of (d, γ_0) from the grid. The partial autocorrelation coefficients related to the ARFIMA$(0, d, 0)$ specifications equal $d/(h - d)$ ($h \in \mathbb{N}^*$) and have the same properties. These results show the strong positivity properties of the ARFIMA$(0, d, 0)$ specifications.

C.3 The entrywise exponentiation of stationary Gaussian vectors of type ARFIMA$(0, d, 0)$ reach level S in the strict sense

This result is quoted at the end of Section 4.2 in the paper. Compatibility is not an issue, as the autocovariance specifications are followed by log-Gaussian sequences. The specifications $\exp(\gamma) - 1$, where γ is of the ARFIMA$(0, d, 0)$ type, are weakly ergodic and have a long memory.

Lines 365 to 406: This `R` program verifies the positivity of such specifications. The derivations are obtained as in Section C.1. The generalized partial autocorrelations are always positive, and the minimum is reached for the lowest value of d and the largest value of γ_0 retained in the verification.

C.4 Level S is not (but nearly) maintained by entrywise exponentiation of stationary Gaussian vectors of type AR(p)

Lines 417 to 456: This `R` program verifies the positivity for an example (the result is quoted at the end of Section 4.2). Let γ denote an AR(1) autocovariance function ($\gamma(h) = 0.5^{|h|}\ \forall h \in \mathbb{Z}$). I derive the minimum of the generalized partial autocorrelations of the sequence $\exp(\gamma) - 1$, depending on the length T of the history (varying from 2 to 100): see Figure 9. The minimum is negative, and equals -7.8×10^{-17}, for $T = 87$. The minimum is negative but very close to zero. Besides, level S is a sufficient condition for nonnegative credibilities, but the paper does not prove that it is necessary. These specifications are used in Pinquet et al. (2001) and Bolancé et al. (2003).

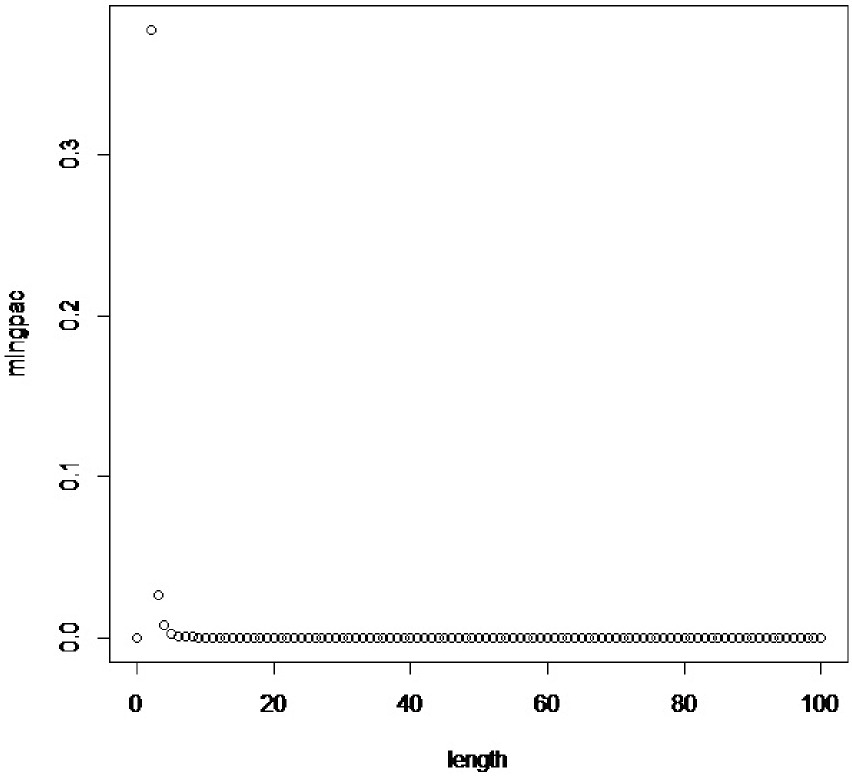

Figure 9: Minimum of the generalized partial autocorrelations of the sequence $\exp(\gamma) - 1$, depending on the length T of the history.

C.5 Proposition 8 for ARFIMA$(0, d, 0)$ specifications and their entrywise exponentiation

The programs refer to the end of Section 5 in the paper. The goal is to place side by side a time-invariant random effect and a dynamic and ergodic random effect that reaches level S, and to maintain this positivity level for the product of these random effects. Proposition 8 gives a sufficient condition. This condition (i.e., the sums of the lines of the precision matrices are nonnegative) is proved in Appendix B.9 for the AR(p) specifications that reach level S. The hereditarity of this condition is proved in Appendix B.9. Hereditarity is used to verify that ARFIMA$(0, d, 0)$ specifications and their entrywise exponentiation fulfil this condition. If γ is an ARFIMA$(0, d, 0)$ autocovariance function, the program verifies that γ and $\exp(\gamma) - 1$ fulfill the condition. The set of ARFIMA$(0, d, 0)$ autocovariance functions is a cone, which is not the case for their transforms by $\gamma \to \exp(\gamma) - 1$. The verification is then achieved on the fractional differencing parameter d only for γ, and for both d and γ_0 for $\exp(\gamma) - 1$.

Lines 469 to 517: This `R` program verifies the positivity condition for the ARFIMA$(0, d, 0)$ specifications. Let R_d^T be the correlation matrix of an ARFIMA$(0, d, 0)$ specification applied to a history with a length equal to T. The function $d \to min(\left[R_d^T\right]^{-1} 1_T)$ decreases on $[0,\ 0.5[$ from unity to a positive limit, equal to 0.0064 in 0.5^- for $T = 100$. From hereditarity, the condition given in Proposition 8 is valid for histories shorter than a century. The vectors $\left[R_d^T\right]^{-1} 1_T$ are centrally symmetric, and the minimum is reached at the middle.

Lines 521 to 561: This `R` program verifies the positivity condition for the ARFIMA$(0, d, 0)$ specifications exponentiated entrywise. The positivity condition is fulfilled, and the minimum values are obtained at the middle of the vectors. They decrease with d and increase with γ_0.

C.6 Verification of the formula for the limit credibility in the short memory case

A formula for the limit credibility in the short memory case is given in Proposition 10, Section 6. The proof (see Appendix B.11) uses the dominated convergence theorem, and a bounding condition is needed (see (35) in Proposition 10). This condition is not easy to prove, and the program commented on here verifies that the limit holds for AR(1) specifications. As an example, let us use: $\lambda = 1/15$; $U = Q$, $\gamma_Q(h) = 0.8^{|h|}\ \forall h \in \mathbb{Z} \Rightarrow \|\gamma_Q\|_1 = 9$. The credibilities per period are obtained from linear filtering on $X = U + A = Q + A$, with A a weak white noise uncorrelated with U and that accounts for risk exposure. The variance of A equals $1/\lambda = 15$. Let t_α^T be the total credibility for T periods, and E^T be the sequence of residuals in the linear filtering of X with the last T values. Using spectral densities, it is easily seen that (see Appendix B.11)

$$\left(1 - t_\alpha^T\right)^2 s_X(0) \ = \ s_{E^T}(0).$$

In the example, $s_X(0) = \|\gamma_X\|_1 = \|\gamma_A\|_1 + \|\gamma_Q\|_1 = 15 + 9 = 24$. The accuracy of the linear prediction is derived in the Levinson–Durbin recursion by $\sin^2(\psi_T) = \gamma_{E^T}(0)/\gamma_X(0)$.

The innovation I^X is the limit of E^T as T goes to infinity. Hence, the limit of $\gamma_{E^T}(0)$ as T goes to infinity equals $\gamma_{I^X}(0)$ from the definition of the innovation as the limit residual. This result holds entrywise on the autocovariance function, which means that $\lim_{T \to +\infty} \gamma_{E^T} = \gamma_{I^X} = \gamma_{I^X}(0)\, \delta_0$ (I^X is white noise) entrywise on $\mathbb{Z}$.

To obtain the formula for the limit credibility, we need the result $\lim_{T \to +\infty} s_{E^T}(0) = s_{I^X}(0)$ $(= \gamma_{I^X}(0))$. This result holds if the limit and the sums can be interchanged. This is true by the dominated convergence theorem, provided a bounding condition is satisfied. The monotone convergence theorem could also be thought of, the only non-obvious assumption being the entrywise decrease of γ_{E^T} with T. From the example, we obtain the following table.

TABLE 7

TOTAL CREDIBILITY, SQUARED SINUS, VARIANCE AND SUM OF AUTOCOVARIANCES OF THE RESIDUAL AS A FUNCTION OF THE LENGTH OF THE HISTORY

T	t_α^T	$\sin^2(\psi_T)$	$\gamma_{E^T}(0)$	$s_{E^T}(0)$
1	0.1281	0.9939	15.90	18.24
5	0.1445	0.9946	15.91	17.56
10	0.1760	0.9943	15.91	16.30
20	0.1853	0.9942	15.91	15.93
40	0.1859	0.9942	15.91	15.91

The results to four digits are unchanged for higher values of T. The variance of the residual $\gamma_{E^T}(0)$ reaches its limit $\gamma_{I^X}(0)$ very quickly. The formula for the limit credibility is valid if $s_{E^T}(0)$ reaches the same limit. The table shows that this is the case, but the convergence is not as fast as that of $\gamma_{E^T}(0)$. Hence, the 'whitening' of E^T is not reached as quickly as the limit of the variance. As expected, we have that $\gamma_{I^X}(0) = 15.91 > 15 = 1/\lambda$.

Lines 573 to 609: This `R` program generates the above table. The variables $\gamma_Q(0)$ and $\|\gamma_Q\|_1$ are defined by lines 579 and 580. Lines 581 and 582 provide $\gamma_X(0)$ and $\|\gamma_X\|_1 = s_X(0)$. Then the Levinson–Durbin recursion is applied to ρ_X, with $\gamma_X = \gamma_Q + \frac{\delta_0}{\lambda} \Rightarrow$ $\rho_X = \frac{\lambda\gamma_Q(0)}{1+\lambda\gamma_Q(0)}\,\rho_Q$ on $\mathbb{N}^*$.

C.7 Estimated autocovariances of random effects from three variables: number of events, frequency premiums, and period index

The unconstrained estimations of the autocovariances given in Table 1 are obtained from a regression on individual data that is not disclosed. See Section A.3 and Pinquet et al. (2001) for the formulas on the autocovariances.

The program **(lines 625 to 718)** is written with the `IML` module for matrix calculus included in `SAS`. The individuals are the pairs (policyholder, risk exposure period), and three variables are used: the number of events (variable `n`), a frequency premium (i.e. an estimated expectation, conditional on covariates: variable `prfreq`), and the period index (variable `period`). The individuals are sorted by increasing periods. The variable `prfreq` is derived from the Poisson regression carried out in Pinquet et al. (2001), with a log-linear specification for the expectation. Any other type of regression (neural network, regression tree, or random forest) can be used to obtain the frequency premium. The only constraint to impose is that the regression components do not use the individual history. The variable `period` allows handling the two first ones in an unbalanced panel dataset framework.

C.8 Rough GMM on AR(p) and ARFIMA$(0, d, 0)$ models for the ergodic component of the random effect

The AR(p) specifications applied to Q that reach level S for $p \leq 3$ are estimated in Table 2, Section 7.2. The specifications are spanned and the autocovariances are generated as indicated in Section C.2. The `R` program (**lines 729 to 980**) derives the estimates of the unconstrained autocovariance function $\widehat{\gamma_U}$ given in Table 1.

- The variance of a time-invariant random effect is the mean of the estimated autocovariances. Uniform weights are not optimal because risk exposure in the estimation strongly decreases with the lag (see Section A.3). This shortcoming is corrected in Section C.10 by means of an asymptotically efficient estimation based on individual frequency premiums.

- The autocovariance function of a random effect with a white noise specification for Q is $\gamma_{U,\alpha} = \sigma_P^2 + ((1+\sigma_P^2)\,\sigma_Q^2\,\delta_0)$, with $\alpha = (\sigma_P^2, \sigma_Q^2)$. The adjustment error $\gamma_{U,\alpha} - \widehat{\gamma_U}$ nullifies at the optimum for $h = 0$, with

$$\widehat{\sigma_Q^2} = \frac{\widehat{\gamma_U(0)}\ -\ \widehat{\sigma_P^2}}{1\ +\ \widehat{\sigma_P^2}}.$$

  The estimator of σ_P^2 minimizes the sum of squared errors for positive lags, hence $\widehat{\sigma_P^2} = \overline{\widehat{\gamma_U}}^{h>0}$.

- The estimation of the other specifications is derived from the minimization of the function $\alpha \longrightarrow \|\gamma_{U,\alpha} - \widehat{\gamma_U}\|^2$. This function is not necessarily convex, and we span the parameter space to find the solution. With three parameters, as is the case with AR(1) and ARFIMA$(0, d, 0)$ specifications, the derivations are easy with a mesh size equal to 0.01. For AR(2) and AR(3) specifications, local derivations are recommended. The program iterates a local search initialized by the optimum reached at the previous order of the autoregressive process, the supplementary filtering coefficient being equal to zero. The optimum is reached in less than 10 iterations with a mesh size equal to 0.01. The optimum might be local, but global derivations derived outside the program show that we have reached the global optimum. Positivity levels N2 and S are equivalent for AR(2) specifications (see Proposition 7). Hence, the triangle defined by $\varphi_1 \geq 0$; $\varphi_2 \geq 0$; $\varphi_1 + \varphi_2 < 1$ must be spanned (regions Ia and Ib in the stationarity triangle, Appendix A.1). For the AR(3) specifications, the positivity level N2 is defined by $\varphi_h \geq 0$ $(h = 1, 2, 3)$; $\varphi_1 + \varphi_2 + \varphi_3 < 1$. Level S is reached with the supplementary conditions $\varphi_1 \geq \varphi_1\varphi_2 + \varphi_2\varphi_3$; $\varphi_2 \geq \varphi_1\varphi_3$ (see Appendix B.14 for geometrical representations).

In Section 7, I argue that ergodicity should not be an assumption for the random effects, especially if the ergodic specification has a short memory. From the autocovariances given in Table 1, Figure 10 exhibits the adjustment errors for three specifications:
1) U: AR(1) ; 2) U: ARFIMA$(0, d, 0)$; 3) U: time-invariant component$\times$AR(1).
The fit is poor for the short memory specification (SSE=0.0700, vs. SSE=0.0067 and 0.0073 for the two other specifications, as seen in Section 7).

C.9 Limit credibility in the ARFIMA$(0, d, 0)$ case: an example

The formula related to the short memory case

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

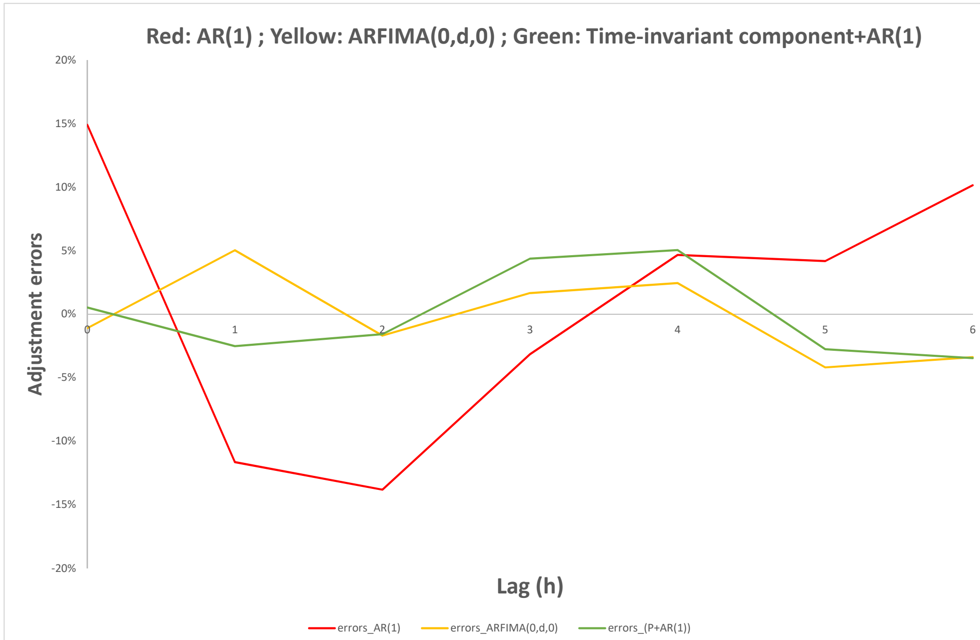

Figure 10: Goodness of fit for specifications with short, long and very long memory.

suggests that the limit credibility equals one at the intermediate level of the memory (ergodicity, and non-summability of the autocovariance function of the random effects).

This paper does not prove this result, but we study the behavior of the total credibility for an example. With the ARFIMA$(0, d, 0)$ specification fitted with the rough GMM estimation and $\lambda = 0.07$, then $1 - t_\alpha^T$ and $2.4 \times T^{-0.4}$ are equivalent at infinity. This result is obtained from the graph of $f : \log(T) \longrightarrow \log(1 - t_\alpha^T)$, derived by the `R` program **(lines 989 to 1032)**. Figure 11 shows evidence of a slant asymptote for the function f.

C.10 Derivations for total credibilities in Figure 1

Lines 1041 to 1078: This `R` program derives the total credibilities for the ARFIMA$(0, d, 0)$ specification estimated in Section 7. Because they are filtering coefficients of X if the frequency risk is time-invariant, the credibilities are obtained by a Levinson–Durbin recursion. The recursion on the total credibility is given at the beginning of Section 6. We derive ρ_X, with $\gamma_X = \gamma_U + \frac{\delta_0}{\lambda} \Rightarrow \rho_X = \frac{\lambda\gamma_U(0)}{1+\lambda\gamma_U(0)}\rho_U$ on $\mathbb{N}^*$. In the example, we have $Q = U$ and $\lambda = 0.07$.

C.11 Second-order stochastic dominance comparisons between linear credibility predictors, ROC curves (Section 7.4)

Lines 1093 to 1368: This SAS program derives linear credibility predictors for the last risk exposure period of policyholders observed over seven years (the maximum length in the sample). Estimation has been rerun from histories restricted to six periods. Elementary statistics are presented in Table 4, then integrated differences of quantiles are derived in order to assess the second-order stochastic dominance between predictors. Given that their mean is greater on the sample, the predictors related to dynamic random effects are less dispersed than the predictor related to a time-invariant random effect. Then the ROC curves and related Gini coefficients are derived, as well as log-likelihoods derived from experience rated frequency premiums.

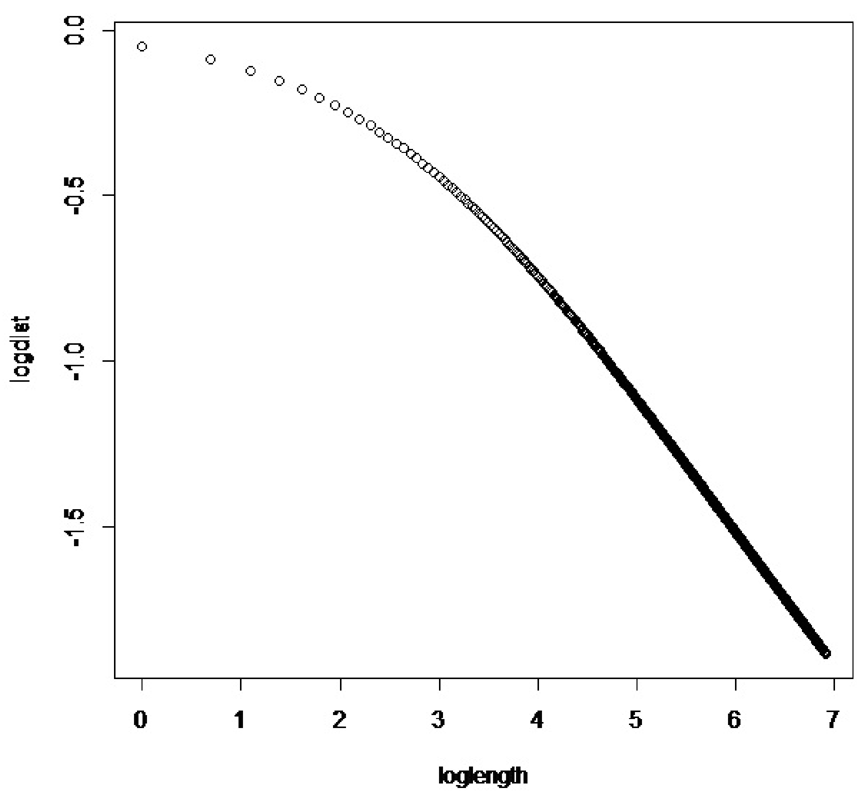

Figure 11: graph of $f:\ \log(T) \longrightarrow \log(1 - t_\alpha^T)$

C.12 Asymptotically efficient GMM on Poisson mixtures from given frequency premiums, and an example with re-estimation of the regression parameters

Lines 1405 to 1715: This SAS program implements the GMM strategy mentioned at the end of Section A.3. Efficient GMM estimators are derived from given frequency premiums derived at the individual level (individual: pair (policyholder, period)). Estimations are performed for time-invariant and dynamic random effects with an AR(1) or ARFIMA(0,d,0) ergodic component. The input is described in Section C.7. The estimating functions are those related to the second-order moments equations ((38), (39) and (40) in Section A.3). We obtain the following results, to be compared with those of Table 2.

$$Q = 1 : \ \widehat{\sigma_P^2} = 0.72\,.$$

$$Q : \ \text{AR}(1).\ \widehat{\sigma_P^2} = 0.51\,;\ \widehat{\gamma_Q(0)} = 0.50\,;\ \widehat{\varphi_1} = 0.39.$$

$$Q : \ \text{ARFIMA}(0, d, 0).\ \widehat{\sigma_P^2} = 0.01\,;\ \widehat{\gamma_Q(0)} = 1.27\,;\ \widehat{d} = 0.37.$$

In what follows, I show an example of comprehensive GMM estimation that uses the $k = 19$ covariates retained in Pinquet et al. (2001). They include the intercept and are all binary. The parameters of the Poisson regression are re-estimated. If the parameters of the mixing distribution are taken on the grid, minimizing the distance to the manifold defined in (42) leads to estimators of α that are given in Table 8 and compared with those of Table 2. The values of the two objective functions are given for the two estimations.

TABLE 8

ESTIMATIONS OF THE AR(2) SPECIFICATION FOR Q (PARTIAL AND EFFICIENT GMM).

	$\widehat{\sigma_U^2}$	$\widehat{\gamma_Q(0)}$	$\widehat{\varphi_1}$	$\widehat{\varphi_2}$	$g(\theta) = \left\Vert \overline{F(\theta)} \right\Vert^2_{\left[\overline{{}^tF(\theta)\,F(\theta)}\right]^{-1}}$	SSE
Table 2 (rough GMM)	0.45	0.56	0.40	0.07	762×10^{-8}	55×10^{-4}
Efficient GMM	0.49	0.52	0.39	0.03	636×10^{-8}	75×10^{-4}

The estimators of β almost do not vary between the two estimations, as: $\max_{1 \le j \le 19} \left| \widehat{\beta}_j - \widehat{\beta}_j^0 \right| = 1.7 \times 10^{-3}$.

Each estimator minimizes a criterion (i.e. the sum of squared errors SSE for $\widehat{\alpha}^0$, the function g for $\widehat{\theta} = (\widehat{\beta}, \widehat{\alpha})$). The goodness of fit of the two GMM estimations on the different lags is given in Table 9.

TABLE 9

GOODNESS OF FIT FOR THE PARTIAL AND EFFICIENT GMM ESTIMATIONS.

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
$\gamma_{U,\widehat{\alpha}^0}(h) - \widehat{\gamma_U(h)}$	−0.007	−0.003	0.031	−0.033	−0.048	0.022	0.024
$\gamma_{U,\widehat{\alpha}}(h) - \widehat{\gamma_U(h)}$	−0.004	0	0.020	−0.030	−0.033	0.047	0.054

The efficient GMM estimation takes into account the risk exposure and gives lower weights to the last lags than the rough derivations of the case study, as expected by the results given in Table 6. This explains why the efficient estimation outperforms the rough strategy for the low values of the lags and is beaten for the last two values.

We provide details on the relative weights of the two estimations. The sample average of the estimating functions is split into two blocks: $M = \overline{F} = \begin{pmatrix} M_\beta & M_\alpha \end{pmatrix}$. The first block corresponds to the likelihood equations in the Poisson model. Then $M_\beta(\widehat{\beta}^0) = 0$, which implies $g(\widehat{\theta}^0) = \left\|M_\alpha(\widehat{\theta}^0)\right\|^2_{W_{\alpha\alpha}(\widehat{\theta}^0)}$ from (42). Equation (41) provides a matrix $W^0_{\alpha\alpha}$ that expresses $\widehat{\theta}^0$ as an estimator of the GMM type, but $W^0_{\alpha\alpha} \neq W_{\alpha\alpha}(\widehat{\theta}^0)$.

At the optimum, the objective function g minimized by the efficient GMM actually concentrates on the diagonal of $W_{\alpha\alpha}$. As $\widehat{\beta}$ is close to $\widehat{\beta}^0$, $M_\beta(\widehat{\beta})$ is close to $M_\beta(\widehat{\beta}^0) = 0$ and the entries of $g(\widehat{\theta}) = \left\|M(\widehat{\theta})\right\|^2_{W(\widehat{\theta})}$ that are located outside the $\alpha - \alpha$ block have a negligible contribution. Indeed, we have $g(\widehat{\theta}) = 636.3 \times 10^{-8}$, with

$$\underbrace{\left\|M(\widehat{\theta})\right\|^2_{W(\widehat{\theta})}}_{636.3\times10^{-8}=} = \underbrace{\left\|M_\alpha(\widehat{\theta})\right\|^2_{W_{\alpha\alpha}(\widehat{\theta})}}_{636\times10^{-8}} + \underbrace{\left\|M_\beta(\widehat{\theta})\right\|^2_{W_{\beta\beta}(\widehat{\theta})}}_{+3.1\times10^{-8}} + \underbrace{2\left(M_\alpha(\widehat{\theta}) \mid M_\beta(\widehat{\theta})\right)_{W_{\alpha\beta}(\widehat{\theta})}}_{-2\times1.4\times10^{-8}}.$$

The contribution of the diagonal of $W_{\alpha\alpha}(\widehat{\theta})$ is equal to $\left\|M_\alpha(\widehat{\theta})\right\|^2_{diag(W_{\alpha\alpha}(\widehat{\theta}))} = 608.5 \times 10^{-8}$. The weights associated to the squared errors reported in Table 9 decrease with the lag for the efficient GMM estimation.

C.13 Derivations in the projective plane

Lines 1717 to 1741: This `R` program derives the matrices related to the projective approach in the affine plane that contains the simplex (see Section B.3).

C.14 Mathematica programs for Figures 5 and 6

Lines 1743 to 1760: This Mathematica program generates a 3D plot of the stationarity domain for AR(3) time series, and a 3D plot of the set associated to stationary times series of the AR(3) type that reach level S.

C.15 Between–within derivations for Table 3

Lines 1769 to 1814: This `R` program provides a between–within analysis of the random effects (Table 3). See Appendix B.15 for the formulas.