B Appendix

B.1 Proof of Proposition 1

1) Perron–Frobenius. If $B \geq 0$ entrywise, then B has a nonnegative eigenvalue λ (termed the Perron root), with $\lambda = \rho(B)$. If $I_T - B$ is an M-matrix, we have $\lambda \in [0, 1]$. If $\lambda = 1$, then $I_T - B$ is singular. ■

2) A positive definite matrix whose off-diagonal entries are nonpositive is an M-matrix. Let P be such a matrix, expressible as $P = s \times (I_T - B)$ where T is the order of P and $s = \max_{1 \leq h \leq T} P_{hh} > 0$. The matrix B is symmetric, with $B \geq 0$ entrywise. We are left to prove $\rho(B) < 1$, as P is nonsingular. Now P is positive definite, hence $\frac{x^{'} B x}{x^{'} x} < 1 \; \forall x \in \mathbb{R}^T - \{0\}$. Hence, the eigenvalues of B are less than one. For $x \neq 0$, the following inequalities are obtained:

$$x^{'} x = |x|^{'} |x| \underset{(a)}{>} |x|^{'} B \, |x| \underset{(b)}{\geq} \left| x^{'} B \, x \right| \geq - x^{'} B \, x. \tag{43}$$

The absolute value function is applied entrywise to x in (43). The matrix P is positive definite, which implies inequality (a). Besides, inequality (b) stems from $B \geq 0$ and the triangular inequality. Then $x^{'} x > - x^{'} B \, x \; \forall x \neq 0$. Hence, the eigenvalues of B are greater than -1. ■

3) Proof of the proposition in the strict sense.

As $P = s \times (I_T - B)$, condition (11) applied to $P = \left[V_U^T \right]^{-1}$ is equivalent to $|\tau - h| = 1 \Rightarrow B_{h\tau} > 0$. We define a square matrix $\underline{B}$ of order T, by

$$|\tau - h| = 1 \Rightarrow \underline{B}_{h\tau} = B_{h\tau} > 0, \; |\tau - h| \neq 1 \Rightarrow \underline{B}_{h\tau} = 0.$$

All the entries of B and $\underline{B}$ are nonnegative, and we have $B \geq \underline{B}$, where the inequality is defined entrywise. We write $\underline{B} = B_L + B_U$, which is an additive lower–upper decomposition of $\underline{B}$. Then

$$0 \leq n < T \Rightarrow \left\{ (B_U^n)_{h\tau} > 0 \Leftrightarrow \tau - h = n \right\}, \; \text{ and } n \geq T \; \Rightarrow B_U^n = 0. \tag{44}$$

A similar property holds for B_L. Equation (44) and the nonnegativity of all the entries lead to the following entrywise results

$$\left(\frac{P}{s} \right)^{-1} = \sum_{n \in \mathbb{N}} B^n \geq \sum_{n \in \mathbb{N}} (B_L + B_U)^n \geq \sum_{0 \leq n < T} \left[(B_L)^n + (B_U)^n \right] \, > \, 0.$$

In addition, some of the filtering coefficients are positive (i.e., $v_{\varphi_U}^T \neq 0$ as $v_{\varphi_U}^T \in (\mathbb{R}^+)^T$ from condition (8)), which provides the proof in the strict sense. To see this, we start from the probabilistic regression $U_t^c = \sum_{h=1}^T \varphi_{T,h} U_{t-h}^c + E_t^T$, with $U^c = U - 1$ and $Cov(E_t^T, U_{t-h}) = 0 \; \forall t \in \mathbb{N}^*, \forall h = 1, \ldots, T$. If $v_{\varphi_U}^T = 0$, then $Cov(U_t, U_{t-h}) = 0 \; \forall t \in \mathbb{N}^*, \forall h = 1, \ldots, T$. The matrices V_U^T and $P = \left[V_U^T \right]^{-1}$ would be diagonal, which is a contradiction to condition (11). ■

B.2 Precision matrices and generalized partial autocorrelation coefficients

We denote the residuals of the affine probabilistic regression of U_h and U_τ with respect to $(U_t)_{t=1,\ldots,T; \, t \neq h,\tau}$ as $U_h^{*(h,\tau)}$ and $U_\tau^{*(h,\tau)}$. Then $r(U_h, \, U_\tau \, | \, U_1, \ldots, U_T) \overset{\text{def}}{=} r(U_h^{*(h,\tau)}, U_\tau^{*(h,\tau)})$.

The vector space spanned by the centered variables $(U_t^c = U_t - 1)_{t=1,\dots,T}$ is endowed with the inner product $\langle X, Y\rangle = E(X\,Y)$. In this vector space, the dual basis of $(U_t^c)_{t=1,\dots,T}$ is denoted by $(U_t^*)_{t=1,\dots,T}$. Then

$$E(U_t^*\,U_s^c) = 1_{[t=s]}\ \forall t, s = 1,\dots,T \Rightarrow E(U_t^*\,U_s^*) = P_{ts}\,\forall t, s\left(P = \left[V_U^T\right]^{-1}\right). \tag{45}$$

The precision matrix is the variance–covariance matrix of the variables of the dual basis. From the residual definition of $U_h^{*(h,\tau)}$ and $U_\tau^{*(h,\tau)}$, we have $U_s^{*(h,\tau)} \in \mathbb{R}U_h^* + \mathbb{R}U_\tau^*$ for $s = h, \tau$. We write

$$U_s^{*(h,\tau)} = a_{sh}U_h^* + a_{s\tau}U_\tau^* \text{ for } s = h, \tau.$$

The identities $\left\langle U_s - U_s^{*(h,\tau)},\, U_t^*\right\rangle = 0\ \forall s, t \in \{h, \tau\}$ and equation (45) lead to

$$\begin{pmatrix} a_{hh} & a_{h\tau} \\ a_{\tau h} & a_{\tau\tau} \end{pmatrix} = \begin{pmatrix} P_{hh} & P_{h\tau} \\ P_{\tau h} & P_{\tau\tau} \end{pmatrix}^{-1} \stackrel{\text{def}}{=} \begin{pmatrix} P^{hh} & P^{h\tau} \\ P^{\tau h} & P^{\tau\tau} \end{pmatrix}.$$

Then

$$V\begin{pmatrix} U_h^{*(h,\tau)} \\ U_\tau^{*(h,\tau)} \end{pmatrix} = V\left[\begin{pmatrix} a_{hh} & a_{h\tau} \\ a_{\tau h} & a_{\tau\tau} \end{pmatrix}\begin{pmatrix} U_h^* \\ U_\tau^* \end{pmatrix}\right] = \begin{pmatrix} P^{hh} & P^{h\tau} \\ P^{\tau h} & P^{\tau\tau} \end{pmatrix} \Rightarrow$$

$$r(U_h,\, U_\tau \,|\, U_1,\dots,U_T) = r(U_h^{*(h,\tau)}, U_\tau^{*(h,\tau)}) = \frac{P^{h\tau}}{\sqrt{P^{hh}}\,\sqrt{P^{\tau\tau}}} = \frac{-\,P_{h\tau}}{\sqrt{P_{hh}}\,\sqrt{P_{\tau\tau}}} = -R_{h\tau}.$$

Equation (13) is proved, as well as the equivalence

$$\left(\left[V_U^T\right]^{-1}\right)_{h\tau} = P_{h\tau} \le 0 \Leftrightarrow r(U_h,\, U_\tau \,|\, U_1,\dots,U_T) \ge 0\ (h \ne \tau).$$

From the stationarity of U, the variance–covariance matrix V_U^T is Toeplitz. This property is lost in the inversion, and the entries of the precision matrix are not constant on diagonals. This explains the "generalized" adjective.

Generalized partial autocorrelation coefficients are a tool in the "big data" literature for variable selection. Popular methods such as the "variable selection with the lasso" (Tibshirani, 1996; Meinshausen et al., 2006) use precision matrices and their statistical interpretation. The surge of the "big data" challenges experience rating in non-life insurance because distribution mixing reflects an unobservable information, which receives a residual interpretation with respect to observable information. Second-order moments of random effects decrease when observable information increases.

B.3 A formal proof of Proposition 3

The derivations related to Figure 4 are given first. We have that

$$\gamma_U:\ \text{AR}(1);\ \varphi_1 = 0.5;\ v_{\gamma_U}^3 = 0.5\,c_1 \Rightarrow p\left(v_{\gamma_U}^3\right) = p(c_1) = a_1.$$

$$\text{If } \gamma_U(0) = 1 : V_U^3 = R = \begin{pmatrix} 1 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 \\ 0.25 & 0.5 & 1 \end{pmatrix} = (c_1\,c_2\,c_3).$$

$$(s(c_1)\,s(c_2)\,s(c_3)) = \begin{pmatrix} 7/4 & 2 & 7/4 \end{pmatrix} \Rightarrow$$

$$A = (a_1\, a_2\, a_3) = \begin{pmatrix} \frac{c_1}{s(c_1)} & \frac{c_2}{s(c_2)} & \frac{c_3}{s(c_3)} \end{pmatrix} = \begin{pmatrix} 4/7 & 0.25 & 1/7 \\ 2/7 & 0.5 & 2/7 \\ 1/7 & 0.25 & 4/7 \end{pmatrix}.$$

$$A^{-1} = \begin{pmatrix} 7/3 & -7/6 & 0 \\ -4/3 & 10/3 & -4/3 \\ 0 & -7/6 & 7/3 \end{pmatrix}.$$

A^{-1} is the matrix of barycentric coordinates of the canonical vector basis $\{e_1,\, e_2,\, e_3\}$ in $\{a_1,\, a_2,\, a_3\}$. The lines of A^{-1} are proportional to those of the precision matrix

$$P = R^{-1} = \begin{pmatrix} 4/3 & -4/6 & 0 \\ -2/3 & 5/3 & -2/3 \\ 0 & -4/6 & 4/3 \end{pmatrix}.$$

In Figure 4, the weights used in: $b_h = \pi_h e_h + (1 - \pi_h) a_h$ $(h = 1, 2, 3)$ are equal to $\pi_1 = 1/2$; $\pi_2 = 5/7$; $\pi_3 = 1/2$. From $1/\pi_h = 1 + (\lambda_{T+1-h}\, s(c_h))$ $(h = 1, 2, 3)$, we obtain $\lambda_1 = \lambda_3 = 4/7$; $\lambda_2 = 1/5$.

The matrix of barycentric coordinates of $\{b_1,\, b_2,\, b_3\}$ in $\{a_1,\, a_2,\, a_3\}$ is

$$B = \begin{pmatrix} 4/3 & -4/6 & 0 \\ -2/3 & 5/3 & -2/3 \\ 0 & -4/6 & 4/3 \end{pmatrix}.$$

The columns of B are derived as averages between those of A^{-1} and of the identity matrix I_3 with the weights π_h and $1 - \pi_h$, respectively $(h = 1, 2, 3)$. The matrix of barycentric coordinates of $\{a_1,\, a_2,\, a_3\}$ in $\{b_1,\, b_2,\, b_3\}$ is

$$B^{-1} = \begin{pmatrix} 0.7 & 0.25 & 0.1 \\ 0.2 & 0.5 & 0.2 \\ 0.1 & 0.25 & 0.7 \end{pmatrix}.$$

These positive coordinates imply positive credibilities. Because $p(v^3_{\gamma_U}) = a_1$, the credibilities $\alpha_{3,h}$ $(h = 1, 2, 3)$ are obtained from (47) as

$$\alpha_{3,1} = 0.7 \times \frac{7/8}{7/2} = 0.175;\ \alpha_{3,2} = 0.2 \times \frac{7/8}{7} = 0.025;\ \alpha_{3,3} = 0.1 \times \frac{7/8}{7/2} = 0.025.$$

Formal proof of Proposition 3: a geometrical interpretation of condition (8) (i.e., nonnegative linear filtering) is

$$v^T_{\varphi_U} = \left[V^T_U\right]^{-1} v^T_{\gamma_U} \in \left(\mathbb{R}^+\right)^T \Leftrightarrow v^T_{\gamma_U} = V^T_U\, v^T_{\varphi_U} \in \sum_{1 \le h \le T} \mathbb{R}^+ c_h,$$

where c_h is the h^{th} column of V^T_U. As the autocovariances are positive, $v^T_{\gamma_U}$ belongs to a cone included in the nonnegative orthant.

We use the vocabulary of projective geometry. Write $s(x) = \sum_{h=1}^T x_h$ $\forall x \in \mathbb{R}^T$. For any scalar a, $H_a = s^{-1}(a)$ is an affine hyperplane of $\mathbb{R}^T$. The function

$$p:\, x \longrightarrow \frac{x}{s(x)} = \mathbb{R}\, x \cap H_1$$

maps $\mathbb{R}^T - H_0$ onto H_1. The function p also maps $(\mathbb{R}^+)^T - \{0\}$ onto the simplex $S = (\mathbb{R}^+)^T \cap H_1$. The vertices of S are the elements $e_1, \ldots, e_T$ of the canonical basis of $\mathbb{R}^T$.

We write $a_h = p(c_h)$ $(h = 1, \ldots, T)$. As V_U^T is positive definite, $\{c_h\}_{h=1,\ldots,T}$ is a vector basis of $\mathbb{R}^T$ and $\{a_h\}_{h=1,\ldots,T}$ is an affine basis of H_1. Hence

$$x \in H_1 \Rightarrow x = \sum_{h=1}^{T} \pi_h^x \, a_h, \text{ with } \sum_{h=1}^{T} \pi_h^x = 1.$$

The $(\pi_h^x)_{h=1,\ldots,T}$ are the barycentric coordinates of x in the affine basis $\{a_h\}_{h=1,\ldots,T}$. The convex hull of $\{a_h\}_{h=1,\ldots,T}$ is denoted by $CH(a_1, \ldots, a_T)$ and is the subset of H_1 defined from nonnegative barycentric coordinates in the affine basis $\{a_h\}_{h=1,\ldots,T}$. A new geometrical interpretation of (8) is obtained from:

$$v_{\gamma_U}^T \in \sum_{1 \le h \le T} \mathbb{R}^+ c_h \Leftrightarrow p\left(v_{\gamma_U}^T\right) \in CH(p(c_1), \ldots, p(c_T)) = CH(a_1, \ldots, a_T). \tag{46}$$

Nonnegative credibilities are interpreted in a similar way. The vector of stacked credibilities per period is denoted by v_α^T. We assume that $\lambda_1, \ldots, \lambda_T > 0$. From equation (4), we have $v_{\gamma_U}^T = \left[\Lambda_T^{-1} + V_U^T\right] v_\alpha^T$. Hence

$$v_{\gamma_U}^T = \sum_{1 \le h \le T} \alpha_{T,h} \, d_h, \text{ with } d_h = \frac{e_h}{\lambda_{T+1-h}} + c_h. \text{ Then} \tag{47}$$

$$\underset{1 \le h \le T}{vec}(\alpha_{T,h}) = v_\alpha^T \in \left(\mathbb{R}^+\right)^T \Leftrightarrow v_{\gamma_U}^T \in \sum_{1 \le h \le T} \mathbb{R}^+ d_h, \Leftrightarrow p\left(v_{\gamma_U}^T\right) \in CH(b_1, \ldots, b_T), \tag{48}$$

$$\text{with } b_h = p\left(d_h\right) = \pi_h e_h + (1 - \pi_h) a_h;\ \pi_h = \frac{1}{1 + \left(\lambda_{T+1-h} \, s(c_h)\right)} \ \forall h = 1, \ldots, T. \tag{49}$$

Nonnegative credibilities are obtained from the nonnegative filtering condition given in (46) if $CH(a_1, \ldots, a_T) \subset CH(b_1, \ldots, b_T)$. The weights π_h defined in (49) are positive, hence $b_h \in]a_h, e_h[\forall h = 1, \ldots, T$. Besides, $CH(a_1, \ldots, a_T) \subset CH(e_1, \ldots, e_T) = S$, where S is the simplex.

The nonpositivity of the off-diagonal coefficients of the precision matrix (i.e., condition (10)) also receives a geometrical interpretation. The equations

$$e_h = \sum_{1 \le \tau \le T} \pi_\tau^{e_h} \, a_\tau \ (h = 1, \ldots, T) \tag{50}$$

provide the barycentric coordinates of the vectors of the canonical vector basis in the affine basis $\{a_h\}_{h=1,\ldots,T}$. Then $e_h = \sum_{\tau=1}^{T} \frac{\pi_\tau^{e_h}}{s(c_\tau)} c_\tau$ $(h = 1, \ldots, T)$. As $(c_\tau)_{\tau=1,\ldots,T}$ are the columns of V_U^T, the precision matrix P is defined by $P_{\tau,h} = \frac{\pi_\tau^{e_h}}{s(c_\tau)}$ $\forall h, \tau = 1, \ldots, T$. Hence, condition (10) is related to the nonpositivity of the barycentric coordinates $\pi_\tau^{e_h}$, if $\tau \neq h$. As $\sum_{\tau=1}^{T} \pi_\tau^{e_h} = 1$ $\forall h = 1, \ldots, T$ in (50), we have $e_h - a_h = \sum_{\tau=1}^{T} \pi_\tau^{e_h} \ (a_\tau - a_h)$ and

$$e_h - a_h \in \sum_{\tau \neq h} \mathbb{R}^- (a_\tau - a_h) \ \forall h = 1, \ldots, T. \tag{51}$$

As $b_h \in]a_h, e_h[\ \forall h = 1, \ldots, T$, equation (51) also holds if e_h is replaced by b_h.

From (48), the credibilities per period are nonnegative if $p\left(v_{\gamma_U}^T\right) \in CH(b_1, \ldots, b_T)$. As nonnegative linear filtering implies $p\left(v_{\gamma_U}^T\right) \in CH(a_1, \ldots, a_T)$, Proposition 3 is proved in the wide sense from the following result.

Proposition 12 *Let $\{a_h\}_{h=1,\ldots,T}$ be an affine basis of H_1. If $(b_h)_{h=1,\ldots,T}$ are elements of H_1 such that*

$$b_h - a_h \in \sum_{\tau \neq h} \mathbb{R}^-(a_\tau - a_h) \ \forall h = 1, \ldots, T \tag{52}$$

(i.e. b_h is at the opposite of a_h with respect to the $(a_\tau)_{\tau \neq h}$ $\forall h = 1, \ldots, T$), then $\{b_h\}_{h=1,\ldots,T}$ is an affine basis of H_1 and $CH(a_1, \ldots, a_T) \subset CH(b_1, \ldots, b_T)$.

We write: $b_h - a_h = \sum_{\tau \neq h} \mu_{\tau,h} \times (a_\tau - a_h)$, with $\mu_{t,h} \leq 0 \ \forall h, \tau$ from (52). The barycentric coordinates of b_h in the affine base $\{a_\tau\}_{\tau=1,\ldots,T}$ are $\pi_\tau^{b_h} = \mu_{\tau,h}$ if $\tau \neq h$, and $\pi_h^{b_h} = 1 - \sum_{\tau \neq h} \mu_{\tau,h}$. These barycentric coordinates are the entries of a matrix Π, with

$$\Pi = (I_T - B) \times D, \ D = \underset{h=1,\ldots,T}{diag} \left(\pi_h^{b_h}\right), \tag{53}$$

and: a) $B_{h,h} = 0 \ \forall h = 1, \ldots, T$; b) $B_{\tau,h} = \frac{-\pi_\tau^{b_h}}{\pi_h^{b_h}} = \frac{-\mu_{\tau,h}}{1 - \sum_{\tau \neq h} \mu_{\tau,h}}$ if $\tau \neq h$.

We have: $B \geq 0$ entrywise and $\sum_{\tau=1}^T B_{\tau,h} < 1 \ \forall h = 1, \ldots, T$ (B is column substochastic). Hence, $I_T - B$ is a nonsingular $M-$matrix (see Lemma 2). The matrix Π is nonsingular, and $\{b_h\}_{h=1,\ldots,T}$ is an affine basis of H_1. The barycentric coordinates of $(a_h)_{h=1,\ldots,T}$ in the affine basis $\{b_\tau\}_{\tau=1,\ldots,T}$ are the entries of $\Pi^{-1} = D^{-1} \times (I_T - B)^{-1} = \sum_{n=0}^{+\infty} D^{-1} B^n$. The entries of Π^{-1} are nonnegative entrywise. A convex hull is defined from nonnegative barycentric coordinates, hence $CH(a_1, \ldots, a_T) \subset CH(b_1, \ldots, b_T)$ and the credibilities are nonnegative.

The positivity result also holds in the strict sense. If the entries of the precision matrix are negative on the subdiagonal and on the superdiagonal, the corresponding barycentric coordinates in (50) are also negative. The same result holds for the barycentric coordinates of b_h with respect to $\{a_\tau\}_{\tau=1,\ldots,T}$, as $\pi_h > 0$ in (49) for $h = 1, \ldots, T$. Applying Lemma 2 in the strict sense implies positive barycentric coordinates of a_h with respect to $\{b_\tau\}_{\tau=1,\ldots,T}$. As $p(v_{\gamma_U}^T) \in CH(a_1, \ldots, a_T)$, the barycentric coordinates in $\{b_\tau\}_{\tau=1,\ldots,T}$ of $p(v_{\gamma_U}^T)$ are positive. Therefore, the credibilities $\alpha_{T,h}$ $(h = 1, \ldots, T)$ are positive as

$$v_{\gamma_U}^T = \sum_h \alpha_{T,h} \, d_h \Rightarrow p(v_{\gamma_U}^T) = \frac{\sum_h \alpha_{T,h} \, d_h}{\sum_h \alpha_{T,h} \, s(d_h)} = \sum_h \frac{\alpha_{T,h} \, s(d_h)}{\sum_\tau \alpha_{T,\tau} \, s(d_\tau)} \, b_h. \ \blacksquare$$

B.4 Proof of Proposition 6

The conditions given in (23) ensure the stationarity of U because the roots of the lag polynomial $\Phi(z) = 1 - \sum_{1 \leq h \leq p} \varphi_h \, z^h$ (i.e., defined by $\Phi(L) \, U^c = \varepsilon$) lie outside the unit disk of $\mathbb{C}$. The double index on the filtering coefficients is used in the proof, and φ_h is denoted by $\varphi_{p,h}$ $(h = 1, \ldots, p)$. As ε is the innovation of U^c, we obtain $\varphi_{T,h} = \varphi_{p,h}$ if $T \geq p \geq h$ and $\varphi_{T,h} = 0$ if $T \geq h > p$. Let us prove that the conditions $\varphi_{p,h} \geq 0 \ \forall h = 1, \ldots, p$ are hereditary. The Levinson–Durbin recursion is used backwards, with

$\varphi_{p,h} = \varphi_{p-1,h} - \varphi_{p,p}\varphi_{p-1,p-h}$; $\varphi_{p,p-h} = \varphi_{p-1,p-h} - \varphi_{p,p}\varphi_{p-1,h}$ for $h = 1, \ldots, p-1$. Then, $\varphi_{p,h} \geq 0$, $\varphi_{p,p-h} \geq 0 \Rightarrow \varphi_{p-1,h} \geq \varphi_{p,p}\varphi_{p-1,p-h} \geq \varphi_{p,p}^2\varphi_{p-1,h}$, and $\varphi_{p,p}^2 < 1$ (from (20)) implies $\varphi_{p-1,h} \geq 0 \, \forall h = 1, \ldots, p-1$. From backward induction, the conditions given in (23) entail a level N2 specification for U. They are obviously necessary.

B.5 Precision matrices of an AR(p) sequence

Proposition 1 is illustrated by the precision matrix associated with an AR(1) sequence. This matrix is tridiagonal and such a result is extended to an AR(p) sequence. The derivation starts from a spherization of the AR(p) sequence. Write $v_U^T = \underset{1 \leq t \leq T}{vec} U_t$, and let $(L_T)_{T \in \mathbb{N}^*}$ be a sequence of lower triangular matrices, with $V\left(L_T \, v_U^T\right) = \sigma_\varepsilon^2 \, I_T \; \forall T \in \mathbb{N}^*$. The constant σ_ε^2 is the variance of the innovation of U. The precision matrices are obtained from the $(L_T)_{T \in \mathbb{N}^*}$ by

$$V\left(L_T \, v_U^T\right) = L_T \, V_U^T \, L_T^{'} \Rightarrow P = \left[V_U^T\right]^{-1} = \frac{L_T^{'} \, L_T}{\sigma_\varepsilon^2} \, . \tag{54}$$

We now determine P for $T \geq 2p$. The definition of the AR(p) sequence provides a natural solution for the t^{th} line of L_T if $t > p$, with $\left(L_T \, v_U^T\right)_t = \varepsilon_t$. We want to bypass the first p lines of L_T, which are not detailed in the proof. The southeast block of P, which is a square matrix of order $T-p$, is obtained from the scalar products of the last $T-p$ columns of L_T. These scalar products are known because L_T is lower triangular. The precision matrix is centrally symmetric (see the proof of the Levinson–Durbin recursion). This provides the northwest block of P. The northeast and southwest square blocks of order p must still be derived, and one derivation is enough due to the symmetry of P. These blocks are derived from the scalar products of one of the first p columns of L_T, and one of the last p columns. As L_T is lower triangular, the unknown entries of L_T are eliminated in the scalar products that lead to the yet undefined blocks of P if $T \geq 2p$. Then, $L_T^{'} \, L_T$ and the precision matrix are completely determined.

The entries of $L_T^{'} \, L_T$ and P are now detailed. First, $|h - \tau| > p \Rightarrow P_{h\tau} = 0$. Hence, the precision matrix is $(2p+1)$-diagonal (which extends the tridiagonal denomination of P if $p = 1$). The off-diagonal entries of $L_T^{'} \, L_T$ that are located at the border are equal to $-\,\varphi_1, \ldots, -\,\varphi_p$ if we start from the northwest or southeast corners. They are nonpositive if U reaches level S. The first condition given in (23) is obtained. On the h^{th} diagonal above or below the main diagonal ($1 \leq h \leq p$), the entries of $L_T^{'} \, L_T$ increase from $-\varphi_h$ (at the border of the matrix) to $-\varphi_h + \sum_{1 \leq \tau \leq p-h} \varphi_\tau \, \varphi_{\tau+h}$. Then the entries are constant, which leads to (24). We assume $T \geq 2p$ in these derivations but this does not restrict the result owing to the hereditarity of level S specifications.

We detail the spherization operator L_T, a lower triangular matrix. We have

$$
L_T = \begin{pmatrix}
? & 0 & \dots & 0 & \dots & \dots & 0 & 0 & 0 & 0 & 0 \\
 & \ddots & \ddots & \vdots & & & & & \vdots & \vdots & \vdots \\
? & \dots & ? & 0 & & & & & & & \\
? & \dots & ? & ? & 0 & & & 0 & & & \vdots \\
-\varphi_p & \dots & -\varphi_2 & -\varphi_1 & 1 & \ddots & & & & & \\
0 & \ddots & \ddots & & \ddots & \ddots & 0 & & \vdots & \vdots & \vdots \\
\vdots & \ddots & \ddots & & \dots & \ddots & 1 & 0 & 0 & 0 & 0 \\
\vdots & & \ddots & \ddots & \ddots & & \ddots & \ddots & \ddots & \vdots & \vdots \\
0 & 0 & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 1 & 0 & 0 \\
0 & & & & 0 & \ddots & \dots & & -\varphi_1 & 1 & 0 \\
0 & \dots & 0 & \dots & \dots & 0 & -\varphi_p & \dots & -\varphi_2 & -\varphi_1 & 1
\end{pmatrix}
$$

In this example, we see that the $p \times p$ northwest block of L_T can be bypassed in the derivation of $L_T^{'} L_T$ if $T \geq 2p$, with the central symmetry property of the precision matrix $\left[V_U^T\right]^{-1} = \frac{L_T^{'} L_T}{\sigma_\varepsilon^2}$.

On the h^{th} diagonal above or below the main diagonal ($1 \leq h \leq p$), the entries of $L_T^{'} L_T$ increase from $-\varphi_h$ to $-\varphi_h + \sum_{1 \leq \tau \leq p-h} \varphi_\tau \, \varphi_{\tau+h}$ if we leave the border of the matrix. Then the entries are constant.

The matrix $L_T^{'} L_T$ is represented below.

$$
L_T^{'} L_T = \begin{pmatrix}
1 & -\varphi_1 & \dots & \dots & -\varphi_p & 0 & 0 & 0 & 0 & 0 & 0 \\
-\varphi_1 & 1+\varphi_1^2 & \ddots & \vdots & & \ddots & 0 & & \vdots & \vdots & \vdots \\
\vdots & \dots & \ddots & \ddots & & & \ddots & & & & \\
\vdots & \dots & & \ddots & \ddots & & & \ddots & & & \vdots \\
-\varphi_p & \dots & & & \ddots & \ddots & & & \ddots & 0 & 0 \\
0 & \ddots & \ddots & & & \ddots & \ddots & & \vdots & \ddots & 0 \\
\vdots & \ddots & \ddots & & \dots & & \ddots & \ddots & & & -\varphi_p \\
\vdots & & \ddots & \ddots & \ddots & & & \ddots & \ddots & \vdots & \vdots \\
0 & 0 & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 1+\varphi_1^2+\varphi_2^2 & -\varphi_1+\varphi_1\varphi_2 & -\varphi_2 \\
0 & & & & 0 & \ddots & \dots & & -\varphi_1+\varphi_1\varphi_2 & 1+\varphi_1^2 & -\varphi_1 \\
0 & \dots & 0 & \dots & \dots & 0 & -\varphi_p & \dots & -\varphi_2 & -\varphi_1 & 1
\end{pmatrix}
$$

From (54), the precision matrix is a positive multiple of $L_T^{'} L_T$. The positivity conditions on autoregressive specifications given in Propositions 6 and 7 are then obtained.

B.6 Proof of equation (26)

We assume that $E(W) = 0$ entrywise without loss of generality, as an intercept is eliminated in the definition of U. The useful result is

$$Z \sim N(0,\, \sigma^2) \Rightarrow E[\exp(Z)] = \exp(\sigma^2/2).$$

Then

$$W_t \sim N(0,\, \gamma_W(0)) \Rightarrow E[\exp(W_t)] = \exp(\gamma_W(0)/2).$$

In the fully specified framework of Section 4, every finite linear combination of the entries of W is Gaussian. For any lag h, we have

$$W_t + W_{t+h} \sim N(0,\, 2(\gamma_W(0) + \gamma_W(h))) \Rightarrow$$

$$E[\exp(W_t + W_{t+h})] = \exp(\gamma_W(0) + \gamma_W(h)) \Rightarrow$$

$$E(U_t\, U_{t+h}) = \frac{E\left[\exp(W_t + W_{t+h})\right]}{E[\exp(W_t)] \times E[\exp(W_{t+h})]} = \exp(\gamma_W(h)).$$

Then, $Cov(U_t,\, U_{t+h}) = E(U_t\, U_{t+h}) - 1 = \exp(\gamma_W(h)) - 1$ for any lag h implies $\gamma_U = \exp(\gamma_W) - 1$ entrywise.

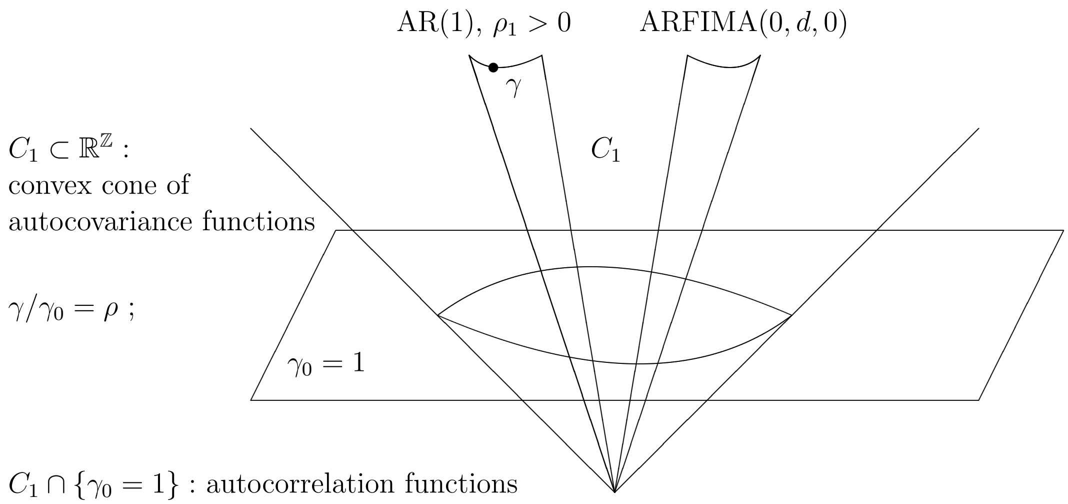
B.7 Interplay between semiparametric and fully specified models on second-order stationary random effects

Figure 5 describes two sets. The first is the convex cone of autocovariance functions and is denoted by C_1. This is the parameter domain for the second-order semiparametric analysis of stationary random effects. The convex set C_2 is much larger, and relates to the probability distributions on $\mathbb{R}^{\mathbb{Z}}$ that are second-order stationary. The subset $C_2^+ = \{P \in C_2,\ P\left((\mathbb{R}^+)^{\mathbb{Z}}\right) = 1\}$ is used for fully specified models of Poisson mixtures with dynamic random effects.

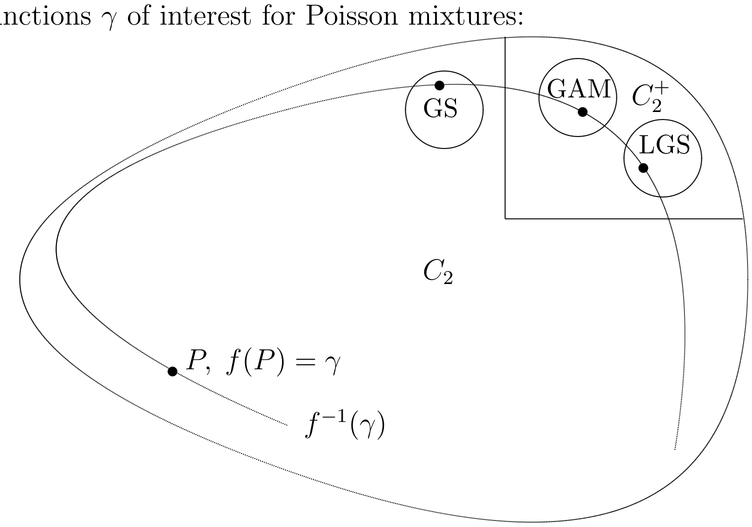
Let P be a probability distribution in C_2, and let $\gamma = f(P)$ be the related autocovariance function. The set $f^{-1}(\gamma)$ always intersects the set GS of Gaussian stationary distributions. To see this, perform Cholesky decompositions on nested variance–covariance matrices related to consecutive variables and then apply the corresponding lower matrices to a strong Gaussian standard white noise. Denote LGS as the set of stationary log-Gaussian distributions. Equation (26) means:

$$\forall \gamma \in C_1 :\ f^{-1}(\gamma) \cap \mathrm{LGS} \neq \emptyset \Leftrightarrow \log(1 + \gamma) \in C_1.$$

In Section 4, we verify with the Levinson–Durbin recursion that this condition holds at the horizon of a century and on a grid of parameters for the AR(p) autocovariance functions ($p = 1, 2, 3$) that reach level N2, and for the ARFIMA$(0, d, 0)$ specifications.



C_2 : convex set of second-order stationary probability distributions on $\mathbb{R}^{\mathbb{Z}}$.
$f:\ C_2 \longrightarrow C_1,\ f(P) = \gamma$
$C_2^+ = \{P \in C_2,\ P\left((\mathbb{R}^+)^{\mathbb{Z}}\right) = 1 \Leftrightarrow \mathrm{supp}(P) \subset (\mathbb{R}^+)^{\mathbb{Z}}\}$.
Autocovariance functions γ of interest for Poisson mixtures:
$f^{-1}(\gamma) \cap C_2^+ \neq \emptyset$



GS, GAM, LGS: sets of distributions of the Gaussian, gamma (including autoregressive gamma), and log-Gaussian type. The sets GAM and LGS are included in C_2^+.
If γ is AR(1), with $\rho_1 (= \gamma_1/\gamma_0) > 0$, then $f^{-1}(\gamma)$ intersects GS, GAM and LGS.

Figure 5: Interplay between semiparametric and fully specified models on second-order stationary random effects.

B.8 The deterministic component of a stationary sequence analyzed in both the time and frequency domains

A deterministic sequence X is defined by a null innovation, hence by an autoregressive equation without innovation of the type $X_t = \sum_{h=1}^{+\infty} \varphi_h \, X_{t-h}$. Deterministic sequences can be obtained from $AR(p)$ specifications related to lag polynomials with roots on the unit circle, which are conjugate because the lag polynomial is real. With $p = 1, 2$, we obtain:
$p = 1 :$ $(I - L)\, X_t = X_t - X_{t-1} = 0 \Rightarrow$ X_t is time-invariant. This is the solution retained in the paper in a multiplicative setting.
$p = 2 :$ $(I - e^{i\theta} L)\,(I - e^{-i\theta} L) X_t = 0$ (with $e^{i\theta} \neq 1) \Leftrightarrow X_t = 2\cos(\theta)\, X_{t-1} - X_{t-2}$. The solutions are of the type

$$X_t = A \sin(\theta t) + B \cos(\theta t),$$

where A and B are random variables. The autocovariance function oscillates without vanishing, and the same result holds for larger values of p.
A characterization of deterministic sequences requires an analysis in the frequency domain. The time domain is $\mathbb{Z}$ in our setting, and the frequency domain is defined by a dual approach on groups (Rudin (2017)). In a discrete time framework, the frequency domain is equal to $\widehat{\mathbb{Z}} = \mathbb{R}/2\pi\mathbb{Z}$. This domain is identified with the interval $[-\pi,\, \pi[$, and is defined as the set of characters $\chi_\theta(h) = e^{i\theta h}$ $(\theta \in \mathbb{R},\ h \in \mathbb{Z})$. The time domain $\mathbb{Z}$ can be seen as a group G endowed with Haar measures (i.e., not identically zero, nonnegative, and invariant with respect to translations). From the invariance property, it is clear that every Haar measure on $\mathbb{Z}$ is equal to the counting measure times a positive constant. Haar measures are equal up to a positive and multiplicative constant for every locally compact and Abelian group. The characters on G constitute the dual group $\widehat{G}$. If G is the time domain, then $\widehat{G}$ is the frequency domain.

A Fourier transform of $f \in L^1(G)$ is defined by $\widehat{f}(\chi) = \int_G f(g)\, \overline{\chi(g)}\, d\mu(g)\ \forall \chi \in \widehat{G}$. In this equation, μ is a Haar measure on G and the Fourier transform depends on a positive coefficient c such as $\mu = c\,\mu_0$, with μ_0 a reference Haar measure. In a time series context, the functions f are autocovariances if $G = \mathbb{Z}$ or spectral densities if $G = \widehat{\mathbb{Z}} = \mathbb{R}/2\pi\mathbb{Z}$. They are even because we restrict our considerations to real-valued random variables. Hence, $\overline{\chi(g)}$ could be replaced by $\chi(g)$ in the Fourier transform, which defines an inverse Fourier transform $\check{f}$.

A Fourier transform of a summable sequence γ (with a short memory if γ is an autocovariance function) is defined by $s(\theta) = \widehat{\gamma}(\theta) = c_1 \sum_{h\in\mathbb{Z}} e^{-i\theta h}\, \gamma(h)$, $\theta \in [-\pi,\, \pi[$; $c_1 > 0$. Herglotz's theorem implies

$$\gamma \text{ is an autocovariance function } \Leftrightarrow\ s = \widehat{\gamma} \geq 0 \text{ on } [-\pi,\, \pi[. \tag{55}$$

In this case, $s = \widehat{\gamma}$ is a spectral density. If X is a stationary sequence such as $\gamma = \gamma_X$, then we have $\gamma_X = \check{\widehat{\gamma_X}} = \check{s}_X = c_2 \int_{-\pi}^{\pi} e^{ih\theta}\, s_X(\theta)\, d\theta$ if $c_1\, c_2 = 1/2\pi$. We will take $c_2 = 1/2\pi$ and $c_1 = 1$, in which case the variance of the innovation of X defined from (27) is the geometric mean of the spectral density s_X. Then

$$\forall h \in \mathbb{Z},\, \forall \theta \in [-\pi,\, \pi[: \gamma_X(h) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ih\theta}\, s_X(\theta)\, d\theta\ ;\ \ s_X(\theta) = \sum_{h\in\mathbb{Z}} e^{-i\theta h}\, \gamma_X(h). \tag{56}$$

An autocovariance function $\gamma_X \notin L^1(\mathbb{Z})$ can still be represented by a nonnegative spectral measure S_X, with

$$\gamma_X(h) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ih\theta} \, dS_X(\theta) \; \forall h \in \mathbb{Z}. \tag{57}$$

Spectral theory provides a characterization of deterministic time series.

Let us start from a time-invariant sequence X. We have $S_X = 2\pi \, \gamma_X(0) \, \delta_0$ from (57), hence the spectral measure is proportional to a Dirac mass located at zero. If X is related to an $AR(2)$ specification without innovation (i.e. $(I - e^{i\theta} L) \, (I - e^{-i\theta} L) X_t = 0$ with $e^{i\theta} \neq 1$), then the spectral measure is a linear combination of Dirac masses located at θ and $-\theta$. For higher values of p, the spectral measure has a finite support.
If $p = +\infty$, consider the component of S_X that is absolutely continuous with respect to any Haar measure on the frequency domain (i.e. the trace of the Lebesgue measure on $\mathbb{R}/2\pi\mathbb{Z}$, up a positive constant). The corresponding density is denoted by s_X. If the Haar measure on the time domain is the counting measure in the definition of the spectral density, then the variance of the innovation I^X of a stationary sequence X is equal to

$$\gamma_{I^X}(0) = \exp \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \log(s_X(\theta)) \, d\theta \right].$$

The variance of the innovation is the geometric mean of the spectral density, as the Haar measure on the frequency domain that is dual to the counting measure is a probability measure. Then

$$\gamma_{I^X}(0) = 0 \Leftrightarrow \int_{-\pi}^{\pi} \log(s_X(\theta)) \, d\theta = -\infty,$$

which is the Kolmogorov–Szegö condition for X to be deterministic. This condition is fulfilled, for instance, if the spectral measure equals zero for a non empty and open set of the frequency domain.
A characterization of deterministic time series from the autocovariance function would be more relevant for practitioners. The nonconstant autocovariance functions derived by the author from spectral densities following the Kolmogorov–Szegö condition always oscillate around zero. Hence, they cannot be retained in the credibility model because they do not reach the positivity level N1.

B.9 Ergodicity properties of stationary time series

Stationary time series have properties that relate to ergodicity in the mean. Consider the dynamic random effect U defined in Section 5. If $U = Q$, U is purely nondeterministic and the autocovariance function γ_U vanishes. The mean square convergence of time averages $\overline{U}^T = \frac{1}{T} \sum_{t=1}^{T} U_t$ ($T \in \mathbb{N}^*$) towards $E(U) = 1$ follows. This weak ergodicity result is easily obtained from the variance of $\overline{U}^T$, equal to the mean of the entries of the variance–covariance matrix V_U^T. This variance is also a mean of γ_U with a triangular kernel. As γ_U vanishes and as the weights of the kernel go to zero for any lag when the length goes to infinity, the limit of $V(\overline{U}^T)$ is equal to zero. The convergence almost everywhere of $\overline{U}^T$ towards $E(U) = 1$ is obtained if U is strictly stationary. This strong ergodicity result is obtained from subadditive ergodicity concepts (Kingman, 1968; Liggett, 1985).

If the sequence U is strictly stationary but not ergodic, then there is still a limit almost everywhere for $\overline{U}^T$ from Birkhoff's theorem. However, this limit is not a constant. In this paper, the random effect is defined as $U_t = P\,Q_t$. In a fully specified context (e.g. with log-Gaussian sequences), we may assume strict stationarity, and $\overline{U}^T = P\,\overline{Q}^T$ converges almost everywhere towards $P \times E(Q) = P$. This result is exploited in Table 3, with between–within derivations on the random effects.

B.10 Proof of Proposition 8

Equation (30) implies $V_U^T = (1+\sigma_P^2)\left[V_Q^T + \left(\frac{\sigma_P^2}{1+\sigma_P^2}\,J_T\right)\right]$, with $J_T = 1_T\,1_T'$. Then

$$\left(I_T + uv'\right)^{-1} = I_T - \frac{uv'}{1+u'v}\ \forall u, v \in \mathbb{R}^T,\ u'v \neq -1 \Rightarrow \tag{58}$$

$$\left[V_Q^T + a\,J_T)\right]^{-1} = \left[V_Q^T\right]^{-1} - \frac{a\left[V_Q^T\right]^{-1} 1_T\,\left(\left[V_Q^T\right]^{-1} 1_T\right)'}{1+\left(a\,1_T'\left[V_Q^T\right]^{-1} 1_T\right)}. \tag{59}$$

Equation (58) is proved with $\left(uv'\right)\left(uv'\right) = u'v\left(uv'\right)$. Equation (59) is obtained from $\left[V_Q^T + a\,J_T\right]^{-1} = \left(I_T + uv'\right)^{-1}\left[V_Q^T\right]^{-1}$, with $u = a\left[V_Q^T\right]^{-1} 1_T$, $v = 1_T$. The first part of Proposition 8 is proved because level S in a wide sense (resp. in a strict sense) is related to the nonpositivity of off-diagonal entries (resp. to the supplementary negativity of entries for the subdiagonal and the superdiagonal) for precision matrices. The hereditarity is obtained from (22), with $\sum_{1\le h\le T}\varphi_{T,h} < 1$ and $\varphi_{T,h} \ge 0\ \forall h = 1,\ldots,T$.

Let Q be an AR(p) sequence that reaches level S. With the notations of Section B.5, we write $c = L_T'\,L_T 1_T = \sigma_\varepsilon^2\left[V_Q^T\right]^{-1} 1_T$, with $T > 2p$. We have a) $c = \widetilde{c}\ (c_h = c_{T+1-h}\ \forall h = 1,\ldots,T)$; b) $c_1 = 1 - \sum_{1\le h\le p}\varphi_h$; c) $c_{h+1} = c_h - \varphi_h\,c_1\ \forall h = 1,\ldots,p$; d) $c_{p+1} = c_1^2$; e) $c_h = c_{p+1}$ if $p+1 \le h \le T-p$. Hence, c_h decreases from c_1 to $c_{p+1} = c_1^2$ as h increases from 1 to $p+1$, which implies $\left[V_Q^T\right]^{-1} 1_T \in (\mathbb{R}^+)^T$. The result also holds for $T \le 2p$, due to hereditarity.

A geometrical interpretation of the condition $\left[V_Q^T\right]^{-1} 1_T \in (\mathbb{R}^+)^T$ is obtained with the framework defined in Sections A.2 and B.3. If $V_Q^T = (c_1^v \ldots c_T^v)$, this condition is equivalent to $p(1_T) \in CH(p(c_1^v),\ldots,p(c_T^v))$. Because $p(1_T) = 1_T/T$, the latter condition means that the center of the simplex belongs to the aforementioned convex hull.

B.11 Limit credibility in the short memory case

Consider equation (18) applied to X. The filtering coefficients (i.e., the credibilities) and the autocorrelation coefficients are nonnegative. Then, $(\widetilde{v_\varphi^T} \mid v_\rho^T) \ge 0 \Rightarrow 0 \le pac_{T+1} \le \frac{\rho_{T+1}}{\sin^2(\psi_T)}$. For positive lags, we have $\gamma_U = \gamma_X$ and $\rho_X = \gamma_U/\gamma_X(0)$. As $\sin^2(\psi_T)$ decreases with T, $\sin^2(\psi_T) \ge \lim_{\tau\to+\infty}\sin^2(\psi_\tau) = \gamma_{I^X}(0)/\gamma_X(0)$.

A lower bound is needed for $\gamma_{I^X}(0)$. Equation (16) implies $E_t^T = A_t + \left[U_t^c - \sum_{h=1}^T \varphi_{T,h}^X\ X_{t-h}\right]$. Then $V(E_t^T) \ge V(A_t) = 1/\lambda$, as A is a white noise sequence uncorrelated with U. The sequence E^T is stationary because it was obtained from X by linear filtering. Hence,

$$\gamma_{E^T}(0) \ge \gamma_A(0) = \frac{1}{\lambda}\ \forall T \in \mathbb{N}^* \Rightarrow \lim_{T\to+\infty}\gamma_{E^T}(0) = \gamma_{I^X}(0) \ge \frac{1}{\lambda}\ \forall \lambda > 0,$$

$$\text{and: } 0 \leq pac_{T+1} \leq \frac{\rho_{T+1}}{\sin^2(\psi_T)} \leq \frac{\gamma_U(T+1)}{\gamma_X(0)} \frac{1}{\gamma_{I^X}(0)/\gamma_X(0)} \leq \lambda\, \gamma_U(T+1).$$

Then, $\gamma_U = \gamma_Q \in L^1(\mathbb{Z}) \Rightarrow pac = pac_X \in L^1(\mathbb{N}^*)$. As γ_X is positive definite, we have $pac_X \in [0,1[^{\mathbb{N}^*}$. From (33), the limit credibility t_α^∞ is less than one by a classic result on infinite products.

Equation (34) is obtained from (16), written as $\Phi_T(L)\, X_t = E_t^T,$ with $\Phi_T(z) = 1 - \sum_{1\leq h\leq T} \varphi_{T,h}\, z^h$. The spectral density is transformed by linear filtering as follows:

$$E_t^T = \Phi_T(L)\, X_t \Rightarrow s_{E^T}(\theta) = \left|\Phi_T(e^{-i\theta})\right|^2\, s_X(\theta) \tag{60}$$

for any $T \in \mathbb{N}^*$ and any $\theta \in [-\pi,\, \pi[$. Equation (56) implies $s_X(0) = \sum_{h\in\mathbb{Z}} \gamma_X(h) = \frac{1}{\lambda} + \sum_{h\in\mathbb{Z}} \gamma_Q(h) = \frac{1}{\lambda} + \|\gamma_Q\|_1$ as $U = Q,\ \gamma_Q \geq 0$ and $\gamma_X = \gamma_U + \frac{\delta_0}{\lambda}$. Writing (60) with $\theta = 0$ yields $\Phi_T(1) = 1 - \sum_{1\leq h\leq T} \varphi_{T,h} = 1 - t_\alpha^T$. Hence,

$$\left(1 - t_\alpha^T\right)^2 \left((1/\lambda) + \|\gamma_Q\|_1\right) = s_{E^T}(0).$$

As the mean square convergence of E_t^T towards I_t^X holds $\forall t \in \mathbb{N}^*$, we have $\lim_{T\to+\infty} \gamma_{E^T}(h) = \gamma_{I^X}(h) = \gamma_{I^X}(0)\, \delta_0(h)\ \forall h \in \mathbb{Z}$ because the innovation I^X is a white noise process. The limit and the sums that define the spectral densities can be interchanged under condition (35) by the dominated convergence theorem, which yields (34).

Although condition (35) is not discussed in this paper, we have verified it numerically on AR(p) specifications (with $p = 1, 2, 3$). An example is given in the document that comments the programs, Section 6.

B.12 Cesàro convergence

From (30), $\lim_{h\to\pm\infty} \gamma_Q(h) = 0$, and from $\gamma_X = \gamma_U + (\delta_0/\lambda)$, we have $\lim_{h\to\pm\infty} \gamma_X(h) = \sigma_P^2$.

The hth column of V_X^T is denoted by $c_X^{T,h}$. If 1_T is the intercept, then

$$\left(1_T'\, V_X^T\right)_h = 1_T'\, c_X^{T,h}\ \forall h \Rightarrow 1_T'\, V_X^T\ \leq \max_{1\leq h\leq T} \left(1_T'\, c_X^{T,h}\right)\ 1_T'\ ,$$

where the inequality is taken entrywise. Nonnegative credibilities imply

$$1_T'\, V_X^T\ v_{\varphi_X}^T \leq \max_{h=1,\ldots,T} \left(1_T'\, c_X^{T,h}\right)\ 1_T'\, v_{\varphi_X}^T.$$

Equation (15) implies $V_X^T\ v_{\varphi_X}^T = v_{\gamma_X}^T$. With $1_T'\, v_{\varphi_X}^T = t_\alpha^T$ and $1_T'\ v_{\gamma_X}^T = \sum_{1\leq h\leq T} \gamma_X(h)$, we obtain

$$t_\alpha^T \geq \frac{\frac{1}{T}\ \sum_{1\leq h\leq T} \gamma_X(h)}{\frac{1}{T}\ \max_{\tau=1,\ldots,T} \sum_{1\leq h\leq T} \gamma_X(h-\tau)}.$$

The proposition follows from

$$\lim_{T\longrightarrow\infty} \frac{1}{T} \sum_{1\leq h\leq T} \gamma_X(h) = \lim_{T\longrightarrow\infty}\ \max_{\tau=1,\ldots,T} \frac{1}{T} \sum_{1\leq h\leq T} \gamma_X(h-\tau) = \sigma_P^2 > 0. \tag{61}$$

Equation (61) implies $t_\alpha^\infty \geq 1$, and the result as $t_\alpha^\infty \leq 1$.

The first limit in (61) follows from a Cesàro convergence result $\lim_{T\longrightarrow\infty} \frac{1}{T}\sum_{h=1}^{T}\gamma_X(h) = \lim_{h\longrightarrow\infty}\gamma_X(h)$. The second limit relies on uniform Cesàro convergence. The proof of the uniform convergence follows.

We obviously can assume that $\lim_{h\to+\infty}\gamma_X(h) = 0$ without loss of generality. However, this cannot be assumed at the inception of the proof because of the ratio.

The mean of γ_X on a non-empty set S, $S \subset \mathbb{Z}$ is denoted by $\overline{\gamma_X}^S$. Let us prove

$$\lim_{T\to+\infty}\left[\max_{S,\,S\subset\mathbb{Z},\,|S|=T}\left|\overline{\gamma_X}^S\right|\right] = 0, \tag{62}$$

where $|S|$ denotes the cardinality of S. This uniformity result is stronger than our primary goal. Then $\lim_{h\to+\infty}\gamma_X(h) = 0 \Rightarrow \forall\varepsilon > 0,\ \exists\, n_\varepsilon \in \mathbb{N}^*\ /\ |\gamma_X(h)| \leq \varepsilon/2\ \forall h \in \mathbb{Z},\ |h| \geq n_\varepsilon$. The set $S\cap\,]-n_\varepsilon,\, n_\varepsilon[$ is denoted by S_ε. Then

$$\overline{\gamma_X}^S = \left[\frac{|S_\varepsilon|}{|S|}\,\overline{\gamma_X}^{S_\varepsilon}\right] + \left[\left(1 - \frac{|S_\varepsilon|}{|S|}\right)\overline{\gamma_X}^{S-S_\varepsilon}\right],$$

with $|S_\varepsilon| \leq 2n_\varepsilon - 1$. The sets S_ε or $S - S_\varepsilon$ may be empty, in which case we may associate them to any given average because the corresponding weight is null. Hence,

$$\left|\overline{\gamma_X}^S\right| \leq \left[\frac{|S_\varepsilon|}{|S|}\,\|\gamma_X\|_\infty\right] + \left[\left(1 - \frac{|S_\varepsilon|}{|S|}\right)\frac{\varepsilon}{2}\right].$$

Then $\left|\overline{\gamma_X}^S\right| \leq \frac{(2n_\varepsilon-1)\,\|\gamma_X\|_\infty}{|S|} + \frac{\varepsilon}{2}$, and $\left|\overline{\gamma_X}^S\right| \leq \varepsilon$ if $|S| \geq (4n_\varepsilon - 2)\,\|\gamma_X\|_\infty/\varepsilon$.

B.13 Length of the memory and spectral measures of the random effects

Let S_U be the spectral measure of the random effect U defined by (57), and consider the related distribution function $C_U(\theta) = S_U([-\pi,\,\theta[)$ defined on $[-\pi,\,\pi]$. The function C_U is increasing, since S_U is a nonnegative measure from Herglotz's theorem. This function is bounded, as $C_U(\pi) = S_U([-\pi,\,\pi[) = 2\pi\,\gamma_U(0)$ from (57), with $h = 0$. The regularity level of this increasing function at $\theta = 0$ in the frequency domain depends on the length of the memory in U and determines the limit credibility. Three cases may occur.

1. There is a jump of C_U for $\theta = 0$ if $\lim_{h\to\pm\infty}\gamma_U(h) = \sigma_P^2 > 0$. The constant autocovariance function γ_P is related to a spectral measure of the Dirac type from (57), with $S_P = 2\pi\,\sigma_P^2\,\delta_0$. The length of the memory is at the highest level, and the limit credibility is equal to one from Proposition 11.

2. The function C_U is differentiable for $\theta = 0$ if $\sigma_P^2 = 0$ and $\gamma_Q \in L^1(\mathbb{Z})$. Indeed, $\gamma_U = \gamma_Q$, and $C_U'(0) = \sum_{h\in\mathbb{Z}}\gamma_U(h)$ from (56). From Proposition 10, the limit credibility is less than one. This short memory level is reached with the AR(p) models defined in Section 3.4.

3. The intermediate level in terms of regularity (i.e., C_U is continuous but not differentiable for $\theta = 0$) corresponds to $\sigma_P^2 = 0$ and $\sum_{h\in\mathbb{Z}}\gamma_Q(h) = +\infty$. This long memory level is obtained with ARFIMA$(0, d, 0)$ semiparametric specifications on $U = Q$. Equation (34) suggests that $t_\alpha^\infty = 1$. An example supporting this intuition is given in Section 7.2.

B.14 Stationary time series of the AR(3) type, and their filtering coefficients

Let us explicit the stationarity conditions for the AR(3) family from the Levinson–Durbin recursion used backwards. First, we have the obvious condition $|pac_3| = |\varphi_{3,3}| = |\varphi_3| < 1$, with $\varphi_{3,h} = \varphi_h$ $(h = 1, 2, 3)$. From the Levinson–Durbin recursion, we obtain

$$(1): \quad \varphi_1 = \varphi_{3,1} = \varphi_{2,1} - \varphi_3\varphi_{2,2}\ ;$$

$$(2): \quad \varphi_2 = \varphi_{3,2} = \varphi_{2,2} - \varphi_3\varphi_{2,1}\ .$$

Then, $\varphi_3(1) + (2) \Leftrightarrow \varphi_{2,2} = pac_2 = \frac{\varphi_2+\varphi_1\varphi_3}{1-\varphi_3^2}$, and the second condition is

$$|pac_2| = |\varphi_{2,2}| < 1 \Leftrightarrow |\varphi_1\varphi_3 + \varphi_2| < 1 - \varphi_3^2.$$

We also obtain $(1) + \varphi_3(2) \Leftrightarrow \varphi_{2,1} = \frac{\varphi_1+\varphi_2\varphi_3}{1-\varphi_3^2}$. From the first step of the Levinson–Durbin recursion, we have $\varphi_{2,1} = \varphi_{1,1} \times (1 - \varphi_{2,2}) \Rightarrow \frac{\varphi_{2,1}}{1-\varphi_{2,2}} = \varphi_{1,1} = pac_1 = \frac{\varphi_1+\varphi_2\varphi_3}{1-\varphi_3^2-\varphi_1\varphi_3-\varphi_2}$.

The denominator of the last ratio is equal to $1-\varphi_{2,2}$ and is positive as $|pac_2| = |\varphi_{2,2}| < 1$. Hence the stationarity conditions: $|pac_h| < 1$ for $h = 3, 2, 1$ are

$$|\varphi_3| < 1;\ |\varphi_1\varphi_3 + \varphi_2| < 1 - \varphi_3^2;\ |\varphi_1 + \varphi_2\varphi_3| < 1 - \varphi_3^2 - \varphi_1\varphi_3 - \varphi_2. \tag{63}$$

Let us interpret the last condition derived from $|pac_1| < 1$: we have
$\varphi_1 + \varphi_2\varphi_3 < 1 - \varphi_3^2 - \varphi_1\varphi_3 - \varphi_2 \Leftrightarrow (\varphi_1 + \varphi_2) \times (1 + \varphi_3) < 1 - \varphi_3^2$
$\Leftrightarrow \varphi_2 < 1 - (\varphi_1 + \varphi_3)$, and $-\varphi_1 - \varphi_2\varphi_3 < 1 - \varphi_3^2 - \varphi_1\varphi_3 - \varphi_2 \Leftrightarrow \varphi_2 < 1 + (\varphi_1 + \varphi_3)$.
To summarize, we obtain the equivalence

$$|pac_1| < 1 \Leftrightarrow \varphi_2 < 1 - |\varphi_1 + \varphi_3|\,. \tag{64}$$

From (63), we have $|pac_2| < 1 \Leftrightarrow |\varphi_1\varphi_3 + \varphi_2| < 1-\varphi_3^2$. The inequality $\varphi_1\varphi_3+\varphi_2 < 1-\varphi_3^2$ is always fulfilled from (64). Indeed, we have

$$\varphi_2 < 1 - |\varphi_1 + \varphi_3| \Rightarrow \varphi_2 < 1 - \varphi_1\varphi_3 - \varphi_3^2 = 1 - [\varphi_3 \times (\varphi_1 + \varphi_3)]\,,$$

as $\varphi_3 \times (\varphi_1 + \varphi_3) \le |\varphi_3| \times |\varphi_1 + \varphi_3| \le |\varphi_1 + \varphi_3|$, with $|\varphi_3| = |pac_3| < 1$.
The only working inequality derived from $|pac_2| < 1$ is $\varphi_3^2 - 1 < \varphi_1\varphi_3 + \varphi_2 \Leftrightarrow \varphi_3^2 - \varphi_1\varphi_3 - 1 < \varphi_2$. The domain related to stationary AR(3) sequences is then defined by

$$|\varphi_3| < 1;\ \varphi_3^2 - \varphi_1\varphi_3 - 1 < \varphi_2 < 1 - |\varphi_1 + \varphi_3|\,. \tag{65}$$

The graphs of the two functions of φ_1 and φ_3 that bound φ_2 look respectively like a saddle and a roof. The constraint $|\varphi_3 - \varphi_1| < 2$ is easily derived from (65). The parameter domain is below the roof and above the saddle, and its projection on coordinates $1-3$ is the parallelogram defined by $|\varphi_3| < 1$ and $|\varphi_3 - \varphi_1| < 2$. The representation is given in Figure 6, with a red roof and a green saddle.

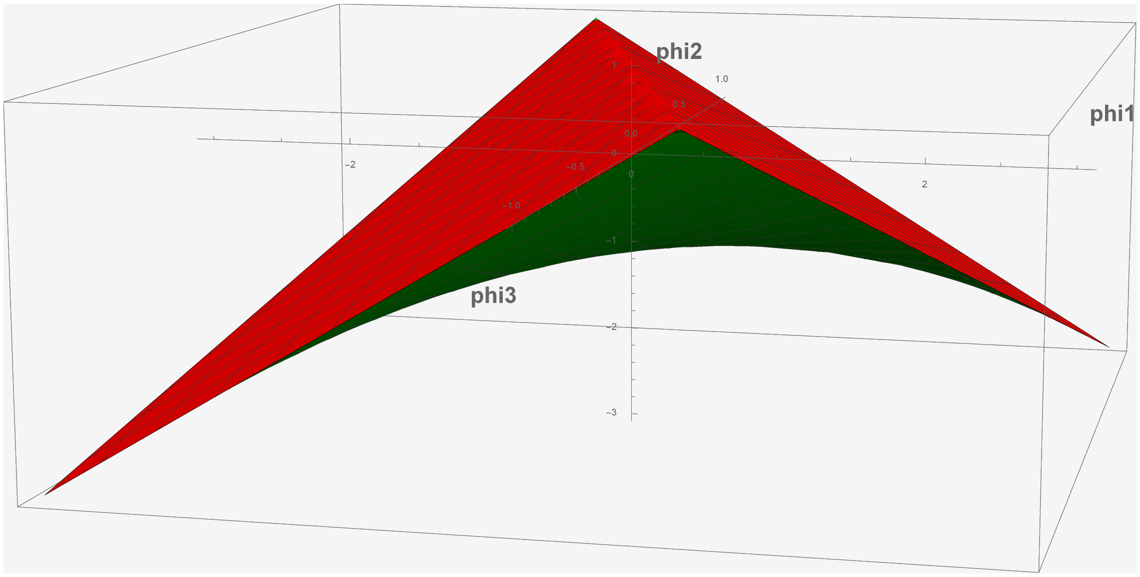


Figure 6: Set of filtering coefficients for stationary AR(3) time series.

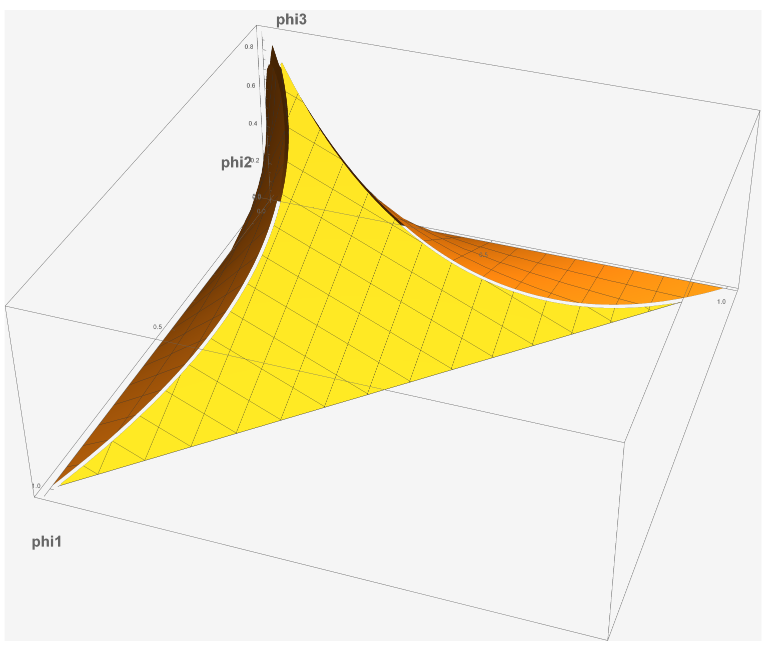


Figure 7: AR(3) sequences that reach level S.

Let us represent the subsets related to distributions that reach levels N2 and S. The intersection of the preceding set with the nonnegative orthant is related to distributions that reach level N2. As seen in Proposition 6, the related set is defined by $\varphi_1, \varphi_2, \varphi_3 \geq 0;\ \varphi_1 + \varphi_2 + \varphi_3 < 1$. The representation is obvious. The subset related to level S is related to the additional constraints $\varphi_1 \geq \varphi_1\varphi_2 + \varphi_2\varphi_3$ and $\varphi_2 \geq \varphi_1\varphi_3$. The set related to level S is described by Figure 7. From left to right, we first see the face related to the binding constraint $\varphi_2 = \varphi_1\varphi_3$. The estimation given in the case study is located in this face. Then the frontier related to the linear constraint $\varphi_1 + \varphi_2 + \varphi_3 < 1$ is seen, and the second nonlinear constraint generates the face on the right side of the figure.

B.15 Limits for the between and within sample variances of the random effects

The variable $y = U$, with $U_{i,t} = P_i\, Q_{i,t}$ from (29) is defined on a virtual and balanced panel dataset, with $T_i = T\ \forall i = 1, \ldots, m$. The total variance of U splits into a between and a within variance as follows:

$$\frac{1}{m\,T}\sum_{i,t}(U_{i,t} - U_{\bullet\bullet})^2 = \frac{1}{m}\sum_{i}(U_{i,\bullet} - U_{\bullet\bullet})^2 + \frac{1}{mT}\ \sum_{i,t}(U_{i,t} - U_{i,\bullet})^2. \tag{66}$$

The within variance is a sample average on the policyholders of the variable

$$\frac{1}{T}\sum_{t}(U_{i,t} - U_{i,\bullet})^2 = \frac{\left(vU_i^T\right)'\ C_T'\, C_T\ vU_i^T}{T},$$

with $vU_i^T = vec_{1\leq t\leq T}\ (U_{i,t})$. The matrix $C_T = I_T - \frac{J_T}{T}\ \left(J_T = 1_T\, 1_T'\right)$, is symmetric and idempotent, hence

$$E\left[\frac{1}{T}\sum_{t}(U_{i,t} - U_{i,\bullet})^2\right] = \frac{\text{Trace}\left[C_T\ E\left(vU_i^T\left(vU_i^T\right)'\right)\right]}{T} = \frac{\text{Trace}\left[C_T\ \left(J_T + V_U^T\right)\right]}{T}.$$

As $V_U^T = \sigma_P^2\, J_T + (1+\sigma_P^2)\, V_Q^T$ (see Section B.10), and $C_T\, J_T = 0$, we have

$$C_T\ \left(J_T + V_U^T\right) = C_T\, V_U^T = (1+\sigma_P^2)\, C_T\, V_Q^T.$$

Then

$$E\left[\frac{1}{T}\sum_{t}(U_{i,t} - U_{i,\bullet})^2\right] = (1+\sigma_P^2)\ \left[\gamma_Q(0) - \frac{\text{Trace}\left(J_T V_Q^T\right)}{T^2}\right]. \tag{67}$$

The ratio $\frac{\text{Trace}\left(J_T V_Q^T\right)}{T^2} = V(Q_{i,\bullet})$ equals the mean of the entries of V_Q^T. As the within variance is a sample average, the convergence towards the expectation given in (67) holds when m goes to infinity. The convergence holds almost everywhere if the random effects $(U_{i,t})_{t\in\mathbb{N}^*}$ are i.i.d. across policyholders in fully specified models following the semiparametric constraints. In a semiparametric setting, the mean square convergence follows from fourth order conditions on the moments of U. The variance $V(Q_{i,\bullet})$ goes to zero when T goes to infinity, from the weak ergodicity result mentioned in Section B.9. This justifies the last line of Table 3.

The total variance equals

$$\frac{1}{m\,T}\sum_{i,t}(U_{i,t} - U_{\bullet\bullet})^2 = \frac{1}{m}\sum_{i}\left[\frac{1}{T}\sum_{t}U_{i,t}^2\right] - U_{\bullet\bullet}^2,$$

and converges almost everywhere towards $E(U^2) - E^2(U) = \gamma_U(0) = \sigma_P^2 + (1+\sigma_P^2)\,\gamma_Q(0)$. The convergence holds almost everywhere or mean square, depending on the assumptions on the random effects discussed earlier. Hence, the limit almost everywhere or mean square of the between variance is equal to

$$\gamma_U(0) - (1+\sigma_P^2)\ \left[\gamma_Q(0) - \frac{\text{Trace}\left(J_T V_Q^T\right)}{T^2}\right] = \sigma_P^2 + \frac{(1+\sigma_P^2)\ \text{Trace}\left(J_T V_Q^T\right)}{T^2}.$$

From a weak ergodicity result applied on Q, the limit of the between variance of U is σ_P^2 when T goes to infinity.

References

Kingman, J. F. (1968). The ergodic theory of subadditive stochastic processes, *Journal of the Royal Statistical Society. Series B,* **30**(3): 499–510.

Liggett, T. M. (1985). An improved subadditive ergodic theorem, *The Annals of Probability,* **13**(4): 1279–1285.

Meinshausen, N., Bühlmann, P. et al. (2006). High-dimensional graphs and variable selection with the lasso, *The Annals of Statistics* **34**(3): 1436–1462.

Rudin, W. (2017). *Fourier Analysis on Groups*, New York: Courier Dover Publications.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 267–288.