SUPPLEMENTARY MATERIAL TO "FORECASTING EXTREME TRAJECTORIES USING SEMINORM REPRESENTATIONS"

BY GILLES DE TRUCHIS[1,A], SÉBASTIEN FRIES[2,B] AND ARTHUR THOMAS[3,C]

[1]Department of Economics, Université d'Orléans, [a]gilles.detruchis@univ-orleans.fr

[2]Department of Econometrics and Data Science, Vrije Universiteit Amsterdam, [b]sebastien.fries@gmail.com

[3]Department of Economics, Université Paris Dauphine - PSL, [c]arthur.thomas@dauphine.psl.eu

Overview of the Supplementary Material. This Supplementary Material complements the main text along five appendices.

- **Appendix A** (Why a seminorm representation?) gives the formal comparison of the seminorm-cylinder representation with the two standard routes to conditional extremes of time series, multivariate regular variation of the path vector and the spectral tail process (Definitions 3–4, Propositions 7–8 and Remarks 5–9). It is summarized in Section 4.4 of the main text.
- **Appendix B** (Forecasting algorithms) extends the pattern-matching algorithm of Section 5.1 to the anticipative AR(1), where the forecast is probabilistic (Algorithm 2), and to general mixed causal–noncausal MAR(r, s) processes (Algorithm 3), together with the risk-threshold decision rule for the predicted crash date.
- **Appendix C** (Monte Carlo study and numerical analysis) reports the finite-sample performance of the forecasting procedures: crash-date forecasts (Appendix C.1), frequency of ties in the pattern-matching step (C.2), comparison of tie-breaking rules (C.3), reversal probabilities (C.4), and visualizations of unit cylinders and directional histograms (C.5–C.6).
- **Appendix D** (Complementary results for the empirical application) documents the SOI application of Section 5: distributional properties of the series, Hill estimator and validity of the diagnostics under infinite variance (D.1), GCoV and AML estimation (D.2–D.3), order selection with AIC and BIC (D.4), and sensitivity of the out-of-sample reversal-date forecasts to the window length m (D.5).
- **Appendix E** (Proofs) contains the proofs of Propositions 1–8, Theorem 1, Corollaries 1–2 and Lemma 1.

A Why a seminorm representation?

In this appendix we aim at clarifying what the seminorm formulation contributes relative to two more standard frameworks for conditional extremes of time series: the multivariate regular variation of the path vector defined in Definition 3, and the spectral tail process anchored at a single coordinate defined in Definition 4. We compare the seminorm framework to these two routes. Proposition 7 quantifies what is lost by anchoring on the single coordinate $\{|X_t| > x\}$: certain trajectory patterns of anticipative processes carry positive spectral mass yet are not charged by the conditional law given that event, while they are captured by the seminorm-pattern conditioning. Proposition 8 then shows that the seminorm conditional limits can be recovered, for any norm on the past block, from the tail measure of (X_t) and the so-called anchoring map of Basrak and Planinić (2021) and Kulik and Soulier (2020); the two routes coincide, and we delineate at the end of this appendix what is owed to the exact α-stable structure. First, let us define the two traditional alternatives of the Conditional Extreme Value Theory (CEVT).

Definition 3 (Multivariate regular variation of the path vector) *An α-stable vector $\boldsymbol{X_t} = (X_{t-m}, \dots, X_{t+h})$ is multivariate regularly varying with index α, i.e. there exists a non-null Radon measure μ on $\mathbb{R}^{m+h+1} \setminus \{\mathbf{0}\}$ such that, for a given norm $\|\cdot\|_\diamond$ on $\mathbb{R}^{m+h+1}$ (the measure μ depends on the chosen norm only through a positive multiplicative constant),*

$$\frac{\mathbb{P}(x^{-1}\boldsymbol{X_t} \in \cdot\,)}{\mathbb{P}(\|\boldsymbol{X_t}\|_\diamond > x)} \xrightarrow[x\to\infty]{v} \mu(\,\cdot\,),$$

in the sense of vague convergence on $\mathbb{R}^{m+h+1} \setminus \{\mathbf{0}\}$, where μ is α-homogeneous: $\mu(cA) = c^{-\alpha}\mu(A)$ for $c > 0$. Writing $\boldsymbol{X_t} = (\overleftarrow{\boldsymbol{X_t}}, \overrightarrow{\boldsymbol{X_t}}) \in \mathbb{R}^{m+1} \times \mathbb{R}^h$, where $\overleftarrow{\boldsymbol{X_t}}$ and $\overrightarrow{\boldsymbol{X_t}}$ stand for past and future path vectors respectively, one may then study, for Borel sets $B \subset \mathbb{R}^{m+1}$ bounded away from the origin and $A \subset \mathbb{R}^h$ with $\mu(B \times \mathbb{R}^h) > 0$ and $\mu(\partial(B \times A)) = \mu(\partial(B \times \mathbb{R}^h)) = 0$,

$$\mathbb{P}\left(\frac{\overrightarrow{\boldsymbol{X_t}}}{x} \in A \,\middle|\, \frac{\overleftarrow{\boldsymbol{X_t}}}{x} \in B\right) \xrightarrow[x\to\infty]{} \frac{\mu(B \times A)}{\mu(B \times \mathbb{R}^h)}. \tag{34}$$

Definition 4 (Spectral tail process anchored at a single coordinate) *For a strictly stationary regularly varying sequence (X_t), Basrak and Segers (2009) construct the spectral tail process $(\Theta_j)_{j\in\mathbb{Z}}$ such that, for any $m \geq 0$, $h \geq 1$,*

$$\mathcal{L}\left(\frac{X_{t-m}}{|X_t|}, \dots, \frac{X_{t+h}}{|X_t|} \,\middle|\, |X_t| > x\right) \xrightarrow[x\to\infty]{d} \mathcal{L}(\Theta_{-m}, \dots, \Theta_h). \tag{35}$$

For a two-sided α-stable moving average as in (11)–(13), the spectral tail process admits the explicit representation $\Theta_j = \Theta_\varepsilon\, d_{J-j}/|d_J|$, $j \in \mathbb{Z}$, where $\Theta_\varepsilon \in \{-1, +1\}$ encodes the extremal

skewness of ε_t and J is an integer-valued index independent of Θ_ε with distribution

$$\mathbb{P}(J = j) = \frac{|d_j|^\alpha}{\sum_{k\in\mathbb{Z}} |d_k|^\alpha}, \qquad j \in \mathbb{Z}, \quad d_j \neq 0. \tag{36}$$

See Mikosch and Wintenberger (2024), and the surrounding single-big-jump heuristic (Kulik and Soulier, 2020) for more details on fundamental properties of the spectral tail process (STP). Those alternatives yield a route to the conditional law of the normalized past and future given that the present coordinate is extreme, without invoking any seminorm structure. However, we claim that our cylinder representation is not in competition with Definitions 3 or 4, but rather provides a distinct and complementary conditioning geometry. Three main features distinguish it.

(a) This first point concerns analytical transparency, and it is worth separating what is owed to the α-stable assumption from what is owed to the seminorm geometry, since the two are easily conflated. In the general regularly varying framework (Definition 3), the limit measure μ is only guaranteed to exist through a convergence-of-measures argument, and its practical use requires nonparametric estimation of the tail measure, on top of the tail-index and extremal-skewness parameters. Within the α-stable class, by contrast, Proposition 1 and Proposition 3($\iota\iota$) yield an exact and finite expression of the spectral measure in terms of the MA coefficients $(d_k)_{k\in\mathbb{Z}}$ and the stability parameters (α, β, σ). This closed form is an advantage of the parametric stable specification over general regular variation; it is not, in itself, specific to seminorms, since the classical Euclidean-sphere spectral measure Γ of $\boldsymbol{X_t}$ already enjoys it. What the seminorm geometry adds is that the same closed form is available directly for the object relevant to forecasting: the cylinder spectral measure $\Gamma^{\|\cdot\|}$ describes the tail behaviour of $\boldsymbol{X_t}$ normalized by the observed past alone, rather than by a full-block norm mixing past and future coordinates (point (b)). The tail conditional limits of Proposition 2 and their MA counterparts (Propositions 4, 5, 6) are thereby analytically explicit. This transparency is of course conditional on the α-stable assumption being appropriate for the data; the genuinely seminorm-specific gains, independent of that caveat, are developed in points (b) and (c).

(b) The spectral tail process (Definition 4) usually anchors on a single coordinate, namely $\{|X_t| > x\}$, which corresponds to the particular choice of seminorm $\boldsymbol{s} \mapsto |s_0|$. Our framework allows any seminorm satisfying (3), for instance $\boldsymbol{s} \mapsto (\sum_{i=-m}^{0} |s_i|^p)^{1/p}$ for $p \in [1, +\infty]$. This freedom matters because different seminorms generate different unit cylinders and thus different conditioning events, and representability on one cylinder does not imply representability on another. Proposition 7 below isolates a family of trajectory

patterns that are charged by the seminorm-cylinder spectral measure but excluded from the single coordinate-anchoring counterpart.

(*c*) Independently of representability, conditioning on a shape $\boldsymbol{X_t}/\|\boldsymbol{X_t}\| \in B(V)$ yields a strictly sharper tail conditional distribution than conditioning on magnitude alone. For past-representable processes (Proposition 4), the magnitude limit is a nondegenerate distribution over countably many trajectory patterns, whereas conditioning additionally on a pattern V_0 as in (27) can collapse this distribution onto a single atom (Proposition 6). This collapse is a property of the coefficient sequence rather than of the normalization: it holds at a target (ϑ_0, k_0) as soon as the past block $\vartheta_0 f(\boldsymbol{d}_{k_0})$ is isolated, up to positive scaling, among the charged past blocks $\vartheta f(\boldsymbol{d}_k)$, which is the content of (27), and this isolation property reads identically under any normalization of the past block. In particular, at a target $k_0 \geq 0$ with $d_{k_0} \neq 0$, the same collapse is obtained from the spectral tail process (35)–(36) by conditioning, through Bayes' formula, on a neighborhood of $\vartheta_0 f(\boldsymbol{d}_{k_0})/|d_{k_0}|$. What the seminorm geometry adds is operational: the normalization $\|\boldsymbol{X_t}\|$ is computable from the observations uniformly in the unknown position k_0 of the peak, post-peak patterns $k_0 \leq -1$ included, whereas the single-coordinate event $\{|X_t| > x\}$ is not triggered by the post-peak patterns isolated in Proposition 7.

We now make points (*b*) and (*c*) precise. The seminorms of interest in this paper satisfy (10), i.e. they measure only the past coordinates of $\boldsymbol{X_t}$. The natural coordinate anchoring of Definition 4 corresponds instead to the seminorm

$$\|(x_{-m}, \ldots, x_0, x_1, \ldots, x_h)\|_\star := |x_0|, \qquad (x_{-m}, \ldots, x_h) \in \mathbb{R}^{m+h+1}, \tag{37}$$

whose kernel is $K^{\|\cdot\|_\star} = \{x \in S_{m+h+1} : \|x\|_\star = 0\}$. For $m \geq 1$, the seminorm (37) satisfies (3) but not (10), so that the framework of Section 3 does not apply to it directly. The following proposition shows that, for any purely anticipative α-stable moving average with $m \geq 1$, there is a genuine representability gap between seminorms satisfying (10) and the single coordinate seminorm (37), and this gap has an explicit asymptotic interpretation.

Proposition 7 *Let (X_t) be a purely anticipative α-stable moving average satisfying (11)–(13), i.e. with coefficients $(d_k)_{k\in\mathbb{Z}}$ such that $d_k = 0$ for all $k \leq -1$, and assume, for simplicity, $\alpha \neq 1$ and $d_k \neq 0$ for all $k \geq 0$. Under these assumptions, $m_0 = 0$ by (18), so that (X_t) is past-representable in the sense of Definition 2 by Proposition 3(ι). Fix $m \geq 1$, $h \geq 1$, let $\boldsymbol{X_t}$ be as in (14), and let $\|\cdot\|_\star$ be the coordinate seminorm (37).*

(ι) *For any seminorm $\|\cdot\|$ satisfying (10), $\boldsymbol{X_t}$ is representable on $C^{\|\cdot\|}_{m+h+1}$, whereas $\boldsymbol{X_t}$ is not representable on $C^{\|\cdot\|_\star}_{m+h+1}$. More precisely, the Euclidean spectral measure Γ of $\boldsymbol{X_t}$ satisfies*

$$\Gamma(K^{\|\cdot\|_\star}) = \sigma^\alpha \sum_{\vartheta\in S_1} \sum_{k=-m}^{-1} w_\vartheta \, \|\boldsymbol{d}_k\|_e^\alpha \; > \; 0, \tag{38}$$

where $\boldsymbol{d}_k = (d_{k+m}, \dots, d_k, d_{k-1}, \dots, d_{k-h})$ and $w_\vartheta = (1+\vartheta\beta)/2$.

($\iota\iota$) *The (at most) $2m$ charged directions excluded from $C^{\|\cdot\|_\star}_{m+h+1}$, namely*

$$\mathcal{D} := \Big\{ \frac{\vartheta \boldsymbol{d}_k}{\|\boldsymbol{d}_k\|_e} \,:\, \vartheta \in S_1,\, k \in \{-m, \dots, -1\} \Big\} \subset K^{\|\cdot\|_\star},$$

correspond to post-peak patterns: for each $(\vartheta_0, k_0) \in S_1 \times \{-m, \dots, -1\}$, the direction $\vartheta_0 \boldsymbol{d}_{k_0}/\|\boldsymbol{d}_{k_0}\|_e$ has its coordinate of index 0 equal to zero, and its nonzero entries are located strictly in past positions $\{-m, \dots, k_0\}$. These directions are asymptotically invisible to coordinate-magnitude conditioning: for any $(\vartheta_0, k_0) \in S_1 \times \{-m, \dots, -1\}$ and any closed neighborhood $\widetilde{A}_{\vartheta_0,k_0} \subset S_{m+h+1}$ of $\vartheta_0 \boldsymbol{d}_{k_0}/\|\boldsymbol{d}_{k_0}\|_e$ in the Euclidean sphere such that $\widetilde{A}_{\vartheta_0,k_0}$ contains no direction $\vartheta \boldsymbol{d}_k/\|\boldsymbol{d}_k\|_e$ with $k \geq 0$,

$$\lim_{x\to\infty} \mathbb{P}\Big(\frac{\boldsymbol{X_t}}{\|\boldsymbol{X_t}\|_e} \in \widetilde{A}_{\vartheta_0,k_0} \,\Big|\, |X_t| > x \Big) = 0. \tag{39}$$

Such neighborhoods exist as soon as $\inf_{k\geq 0} |d_k|/\|\boldsymbol{d}_k\|_e > 0$, which holds for the anticipative AR(1), the anticipative AR(2) with real characteristic roots and the anticipative FWN.

($\iota\iota\iota$) *The spectral mass (38), normalized by $\Gamma(S_{m+h+1})$, is the asymptotic fraction of the extremes of the path vector $\boldsymbol{X_t}$ in which the present coordinate is negligible, that is, which fail to trigger $\{|X_t| > \epsilon x\}$ however small the fixed proportion $\epsilon > 0$. Writing $\boldsymbol{s} = (s_{-m}, \dots, s_h)$ for elements of S_{m+h+1}, for every $\epsilon > 0$,*

$$\begin{aligned} \lim_{x\to\infty} \mathbb{P}(|X_t| \leq \epsilon x \,|\, \|\boldsymbol{X_t}\|_e > x) &= \frac{1}{\Gamma(S_{m+h+1})} \int_{S_{m+h+1}} \Big(1 - (|s_0|/\epsilon)^\alpha\Big)_+ \Gamma(d\boldsymbol{s}) \\ &\xrightarrow[\epsilon\to 0]{} \frac{\Gamma(K^{\|\cdot\|_\star})}{\Gamma(S_{m+h+1})} = \frac{\sum_{k=-m}^{-1} \|\boldsymbol{d}_k\|_e^\alpha}{\sum_{k\in\mathbb{Z}} \|\boldsymbol{d}_k\|_e^\alpha} \;\in\; (0,1). \end{aligned} \tag{40}$$

Remark 5 *Proposition 7 makes precise what the single-coordinate conditioning event $\{|X_t| > x\}$ misses, and quantifies it. Its three parts address complementary aspects of this gap. On part (ι), the coordinate seminorm $\|\cdot\|_\star$ fails representability because its kernel $K^{\|\cdot\|_\star}$ carries a strictly positive Euclidean spectral mass, explicitly computed in (38). This mass is exactly the total weight attached to the charged directions $\mathcal{D}$, i.e. precisely those directions whose coordinate of index 0 vanishes.*

Those directions in $\mathcal{D}$, stated in part ($\iota\iota$), correspond to post-peak patterns: an extreme event has already occurred at some date $\tau \in \{t-m, \dots, t-1\}$, so that the present coordinate X_t has

reverted toward central values while the past piece of trajectory $(X_{t-m}, \dots, X_{t-1})$ *still records the trace of the collapsed bubble. These configurations are probabilistically significant under the joint law of* $\boldsymbol{X_t}$*, yet they are not charged by the conditional law of* $\boldsymbol{X_t}$ *given* $\{|X_t| > x\}$*, an event which, by construction, they do not trigger. This is a statement about the conditioning event, not about the information content of the spectral tail process: seen from the date* $t + k_0$ *of the extreme event, the pattern* $\boldsymbol{d}_{k_0}$ *with* $k_0 \leq -1$ *is the spectral tail process anchored at that date,* $(\Theta_{i-k_0})_{-m \leq i \leq h}$ *on* $\{J = 0\}$*, observed over the shifted window* $\{-m - k_0, \dots, h - k_0\}$*, and the shift-invariant tail measure of* (X_t) *recovers it (Proposition 8 below).*

Part $(\iota\iota\iota)$ *turns this observation into a number. In words, a forecaster who waits for the present coordinate* $|X_t|$ *to exceed a fixed fraction* ϵ *of the level* x *misses, asymptotically and whatever* $\epsilon > 0$*, at least the proportion* (40) *of the extremes of the path vector* $\boldsymbol{X_t}$*, and this lower bound is exactly the representability gap of part* (ι)*. At the common level* $\epsilon = 1$*, the first line of* (40) *gives the missed proportion* $1 - \sum_k |d_k|^\alpha / \sum_k \|\boldsymbol{d}_k\|_e^\alpha$*. For* $m = 0$*, the seminorm* (37) *satisfies (10) and the gap vanishes: the single-coordinate anchoring is then a particular case of the seminorm framework, and the phenomenon isolated in this proposition is specific to observation windows of length at least two. The same construction, applied to any seminorm satisfying (10) in place of* $\|\cdot\|_\star$*, is developed in Remark 6.*

Remark 6 *The construction of* (40) *applies to any seminorm* $\|\cdot\|$ *satisfying (10) in place of* $\|\cdot\|_\star$ *and defines, for a general two-sided moving average (11)–(13) and fixed* (m, h)*, the index*

$$\pi_m := \lim_{\epsilon \to 0} \lim_{x \to \infty} \mathbb{P}(\|\boldsymbol{X_t}\| \leq \epsilon x \,|\, \|\boldsymbol{X_t}\|_e > x) = \frac{\Gamma(K^{\|\cdot\|})}{\Gamma(S_{m+h+1})} = \frac{\sum_{k:\, f(\boldsymbol{d}_k) = \boldsymbol{0},\ \boldsymbol{d}_k \neq \boldsymbol{0}} \|\boldsymbol{d}_k\|_e^\alpha}{\sum_{k \in \mathbb{Z}} \|\boldsymbol{d}_k\|_e^\alpha}, \tag{41}$$

which does not depend on the seminorm chosen in (10), is nonincreasing in m*, and vanishes if and only if* $\boldsymbol{X_t}$ *is representable on* $C_{m+h+1}^{\|\cdot\|}$ *(Proposition 1, for* $\alpha \neq 1$*), i.e. if and only if* $m \geq m_0$ *(Proposition 3). It is the asymptotic fraction of the extremes of the path vector that are not announced by the observed window of length* $m + 1$*: conditionally on* $\|\boldsymbol{X_t}\|_e > x$*, the limit law of* $\boldsymbol{X_t}/\|\boldsymbol{X_t}\|_e$ *splits into a component of mass* $1 - \pi_m$ *on which the past block is non-null, so that the extreme is announced by the observed window, and a component of mass* π_m *on which the extreme originates from a shock located strictly after* t*, so that the forecast error is of the order of the extreme itself. For a purely non-anticipative moving average* ($d_k = 0$ *for* $k \geq 1$)*,* $f(\boldsymbol{d}_k) = \boldsymbol{0} \neq \boldsymbol{d}_k$ *at least for* $k = k^* + 1$*, where* $k^* := \max\{k : d_k \neq 0\} \leq 0$ *(and exactly for* $k \in \{1, \dots, h\}$ *when* $d_k \neq 0$ *for all* $k \leq 0$*), whereas* $f(\boldsymbol{d}_{k^*}) \neq \boldsymbol{0}$*, so that* $\pi_m \in (0, 1)$ *for every* $m \geq 0$*, in accordance with Corollary 1. For the purely anticipative processes of Section 4,* $\pi_m = 0$ *for every* $m \geq 1$ *and for every* $m \geq 0$ *when* $d_k \neq 0$ *for all* $k \geq 0$*. The index* π_m *thus gives a quantitative content to the statement that anticipativeness is the source of the*

predictability of extremes.

Now it remains to clarify how our approach relates to the anchoring-map concept introduced by Basrak and Planinić (2021) (see also Last (2023) for the connection between tail processes and tail measures via Palm calculus) and discussed in Kulik and Soulier (2020), in which the conditional extremes of a regularly varying time series may be indexed by exceedance events beyond the single-coordinate anchor $\{|X_t| > x\}$. We first recall the notation of Kulik and Soulier (2020). Generic elements of the sequence space $\mathbb{R}^{\mathbb{Z}}$ are written $y = (y_j)_{j\in\mathbb{Z}}$; $Y = (Y_j)_{j\in\mathbb{Z}}$ denotes the tail process of (X_t), ν its tail measure on $\mathbb{R}^{\mathbb{Z}}$, $Y^* := \sup_{j\in\mathbb{Z}} |Y_j|$, and $\mathcal{E}(y) := \#\{j \in \mathbb{Z} : |y_j| > 1\}$ the number of exceedances of a sequence. The first-exceedance anchor of Kulik and Soulier (2020), Example 5.5.2, is the map

$$A_1(y) := \inf\{j \in \mathbb{Z} : |y_j| > 1\}, \tag{42}$$

with the conventions $\inf\emptyset = +\infty$ and $A_1(y) = -\infty$ when infinitely many exceedances occur on the negative half-line. It satisfies the anchoring axioms (Ai)-(Aii) of Definition 5.5.1 in Kulik and Soulier (2020), and $\theta_e := \mathbb{P}(A_1(Y) = 0)$ is the candidate extremal index of (X_t) (Corollary 5.5.4 there). Let N be a norm on $\mathbb{R}^{m+1}$. Every seminorm satisfying (10) is of the form $\|\boldsymbol{s}\| = N(f(\boldsymbol{s}))$ with $N(\boldsymbol{u}) := \|(\boldsymbol{u}, \boldsymbol{0})\|$, and conversely, so that our seminorm conditioning event $\{\|\boldsymbol{X_t}\| > x\}$ coincides, through (3), with the past-block exceedance event $\{N(\overleftarrow{\boldsymbol{X_t}}) > x\}$, which is measurable with respect to the observed past alone. For $N = \|\cdot\|_\infty$ this event reads $\{\max_{0\le i\le m} |X_{t-i}| > x\}$: an exceedance of the rescaled path falls inside the observation window, and the block first-exceedance map $\inf\{j \in \mathbb{Z} : \|w_j\|_\infty > 1\}$ coincides with A_1. For a general N, it is indexed by the first exceedance $\inf\{j \in \mathbb{Z} : N(w_j) > 1\}$ of the $\mathbb{R}^{m+1}$-valued sequence of past blocks $w_j = (y_{j-m}, \dots, y_j)$, which is an anchoring map for the regularly varying $\mathbb{R}^{m+1}$-valued sequence $(\overleftarrow{\boldsymbol{X}}_j)_{j\in\mathbb{Z}}$ in the sense of Definition 5.5.1 of Kulik and Soulier (2020), that definition being stated on $(\mathbb{R}^d)^{\mathbb{Z}}$ for an arbitrary norm on $\mathbb{R}^d$ (see also Remark 3.1 of Basrak and Planinić (2021)).[1] The following proposition shows that the conditional limit (24) can be recovered from ν for every norm N and every two-sided moving average satisfying (11)-(12), without any recourse to the seminorm structure on $\mathbb{R}^{m+h+1}$ and without any representability condition. For

[1] The same map, embedded in the scalar sequence space $\mathbb{R}^{\mathbb{Z}}$ with the absolute value, fails axiom (Ai) for every ℓ^p-norm with $p < \infty$: for $m = 1$ and $N = \|\cdot\|_2$, the sequence $y_{-5} = y_{-4} = 0.8$, $y_0 = 2$, $y_j = 0$ otherwise, has its first block exceedance at $j = -4$ while $|y_{-4}| = 0.8 < 1 = |y_0| \wedge 1$. This is a feature of the scalar embedding, not of the block anchoring, and it plays no role in what follows: the derivation below only uses A_1 and the tail measure ν.

Borel sets $A \subset \mathbb{R}^{m+h+1}$ and $V \subset \mathbb{R}^{m+1}$, consider

$$\lim_{x\to\infty} \mathbb{P}\left(\frac{\boldsymbol{X_t}}{N(\overleftarrow{\boldsymbol{X_t}})} \in A \,\middle|\, \frac{\overleftarrow{\boldsymbol{X_t}}}{N(\overleftarrow{\boldsymbol{X_t}})} \in V,\ N(\overleftarrow{\boldsymbol{X_t}}) > x \right). \tag{43}$$

In view of Proposition 8 we also split the reference vector $\boldsymbol{d}_k = (d_{k+m}, \dots, d_{k-h})$ of (14)–(15) into past and future blocks, for $k \in \mathbb{Z}$,

$$\overleftarrow{\boldsymbol{d}_k} := (d_{k+m}, \dots, d_k) \in \mathbb{R}^{m+1}, \qquad \overrightarrow{\boldsymbol{d}_k} := (d_{k-1}, \dots, d_{k-h}) \in \mathbb{R}^h,$$

so that $\boldsymbol{d}_k = (\overleftarrow{\boldsymbol{d}_k}, \overrightarrow{\boldsymbol{d}_k})$ and $\overleftarrow{\boldsymbol{d}_k} = f(\boldsymbol{d}_k)$ where f is the past-projection of (22).

Proposition 8 *Let (X_t) be an α-stable moving average satisfying (11)–(13), with $\alpha \neq 1$ or $(\alpha, \beta) = (1, 0)$, let $m \geq 0$, $h \geq 1$, let N be a norm on $\mathbb{R}^{m+1}$ and set $C^N := \{\boldsymbol{s} \in \mathbb{R}^{m+h+1} : N(f(\boldsymbol{s})) = 1\}$. Define the finite measure*

$$\Gamma^N := \sum_{\vartheta \in S_1} \sum_{\ell \in \mathbb{Z}:\ \overleftarrow{\boldsymbol{d}_\ell} \neq \boldsymbol{0}} w_\vartheta\, N(\overleftarrow{\boldsymbol{d}_\ell})^\alpha\, \delta_{\left\{\vartheta \boldsymbol{d}_\ell / N(\overleftarrow{\boldsymbol{d}_\ell})\right\}} \tag{44}$$

on C^N. Let $A \subset \mathbb{R}^{m+h+1}$ and $V \subset \mathbb{R}^{m+1}$ be Borel sets, $B(V) = V \times \mathbb{R}^h$, such that $\Gamma^N(\partial(A \cap B(V))) = \Gamma^N(\partial B(V)) = 0$ and $\Gamma^N(B(V)) > 0$, where ∂ denotes the boundary, relative to C^N, of the trace on C^N of the set considered. Then the conditional tail limit (43)

(ι) exists;

(ιι) admits the closed form

$$\frac{\Gamma^N(A \cap B(V))}{\Gamma^N(B(V))} = \frac{\sum_{\vartheta \in S_1} \sum_{\ell \in \mathbb{Z}} w_\vartheta\, N(\overleftarrow{\boldsymbol{d}_\ell})^\alpha\, \mathbf{1}\{\vartheta \boldsymbol{d}_\ell / N(\overleftarrow{\boldsymbol{d}_\ell}) \in A,\ \vartheta \overleftarrow{\boldsymbol{d}_\ell} / N(\overleftarrow{\boldsymbol{d}_\ell}) \in V\}}{\sum_{\vartheta \in S_1} \sum_{\ell \in \mathbb{Z}} w_\vartheta\, N(\overleftarrow{\boldsymbol{d}_\ell})^\alpha\, \mathbf{1}\{\vartheta \overleftarrow{\boldsymbol{d}_\ell} / N(\overleftarrow{\boldsymbol{d}_\ell}) \in V\}}, \tag{45}$$

where the summands with $\overleftarrow{\boldsymbol{d}_\ell} = \boldsymbol{0}$ vanish by convention; it is derived from the tail-measure identity (5.5.7) of Kulik and Soulier (2020) applied to the first-exceedance anchor A_1, combined with the single-big-jump representation (36), and involves neither an anchoring map attached to N nor a representability condition;

(ιιι) coincides, when (X_t) is (m, h)-past-representable and $\|\boldsymbol{s}\| = N(f(\boldsymbol{s}))$ is the associated seminorm satisfying (10), with the conditional limit of Proposition 4, since then $\Gamma^{\|\cdot\|} = \sigma^\alpha \Gamma^N$ by (23). In particular, for $m \geq 1$, for the anticipative AR(2) with real characteristic roots and for the anticipative FWN, with $A := A_{\vartheta,k}$ and $V := V_0$ as in (27)–(28), it reduces to $\delta_{\vartheta_0}(\vartheta)\delta_{k_0}(k)$ (Proposition 6).

Proposition 8 situates the seminorm-cylinder framework of Propositions 3–4 within the landscape of conditional extreme value theory recalled in Definitions 3 and 4. We organize the

discussion in three remarks: the first two contrast the past-block conditioning of (43) with the block regular variation route of Definition 3 and with the single-coordinate spectral tail process of Definition 4; the third delineates what is owed to the exact α-stable structure.

Remark 7 *Comparison with the multivariate regular variation of the path vector (Definition 3). Normalizing $\boldsymbol{X_t}$ by a genuine norm on the full path vector, as in (2), is misaligned with the forecasting problem: the conditioning event then depends on the unobserved future block $\overrightarrow{\boldsymbol{X_t}}$. The block regular variation reformulation (34) avoids this pitfall, since its conditioning event $\{\overleftarrow{\boldsymbol{X_t}}/x \in B\}$ is measurable with respect to the observed past. Choosing $B = \{\boldsymbol{u} \in \mathbb{R}^{m+1} : N(\boldsymbol{u}) > 1,\ \boldsymbol{u}/N(\boldsymbol{u}) \in V\}$, which is Borel and bounded away from the origin, it even encodes exactly the conditioning event of (43), and Proposition 8 is the explicit form of the limit (34) for stable moving averages: μ is, up to a positive constant, the projected tail measure $\tilde{\nu}$ of the proof, and Γ^N is its polar decomposition with respect to the past-block norm on $\{\boldsymbol{s} : f(\boldsymbol{s}) \neq \boldsymbol{0}\}$. What Definition 3 does not supply by itself is this explicit form, nor the characterization of the set of directions with vanishing past block, $\{\boldsymbol{s} : f(\boldsymbol{s}) = \boldsymbol{0}\}$, which carries the extremes of $\boldsymbol{X_t}$ that no observed window of length $m+1$ announces. Proposition 8 shows that the past-block conditional limit is well defined regardless of the mass of μ on this set; what the absence of such mass characterizes is the representability of $\boldsymbol{X_t}$ on the cylinder (Theorem 1), that is, the exact spectral representation of the characteristic function of $\boldsymbol{X_t}$ in terms of directions of the observed past only. Its failure is quantified in Proposition 7(ι) for the single-coordinate seminorm, and in Remark 6 for a general observation window.*

Remark 8 *Comparison with the spectral tail process anchored at a single coordinate (Definition 4). The route of Definition 4 normalizes by the present coordinate $|X_t|$ and yields the conditional limit law of $(\Theta_{-m}, \ldots, \Theta_h)$ in (35). This single-coordinate anchoring imposes $|\Theta_0| = 1$ a.s. and consequently assigns vanishing asymptotic probability to every sufficiently small closed Euclidean-shape neighborhood of a post-peak direction, $k_0 \in \{-m, \ldots, -1\}$ (Proposition 7($\iota\iota$)), the missed mass being given by (40). The past-block limit (43), by contrast, reaches every pattern index $k_0 \in \{-m, \ldots, h\}$: for $N = \|\cdot\|_\infty$, the conditioning event is triggered as soon as $\max_{0\leq i\leq m} |X_{t-i}| > x$, regardless of which past coordinate realizes the maximum, and for a general N as soon as the past block is large in that norm. This is a difference between conditioning events, not between frameworks. Both limits are functionals of the same tail measure ν, and the proof of Proposition 8 recovers the past-block limit from the single-coordinate tail process through the shift-invariance identity (5.5.7) of Kulik and Soulier (2020), for every norm N and without any anchoring map attached to N. Likewise, at a target $k_0 \geq 0$ the Dirac collapse of*

Proposition 6 is obtained from (35)–(36) by conditioning on a neighborhood of $\vartheta_0 f(\boldsymbol{d}_{k_0})/|d_{k_0}|$ (point (c) above). What is specific to the past-block conditioning is operational: the normalization $N(\overleftarrow{\boldsymbol{X}_t})$ is computable from the observations uniformly in the unknown position k_0 of the peak, post-peak patterns included, whereas the tail-process normalization $|X_t|$ presupposes that the present observation is itself extreme.

Remark 9 *What is owed to exact stability. Proposition 8 holds for every two-sided moving average satisfying (11)–(12), including the anticipative FWN (31) and the mixed causal-noncausal autoregressions of Gouriéroux et al. (2025), and its proof uses the α-stability of the innovations only through the single-big-jump representation (36) of the tail process, which is available for i.i.d. regularly varying innovations with tail index $\alpha > 0$ and tail balance (w_{-1}, w_{+1}) (Mikosch and Wintenberger, 2024). The tail-measure route is thus the natural one for an extension of our conditional limits beyond the stable class, which we leave for future research. The seminorm-cylinder framework, by contrast, exploits the exact α-stable characteristic function: the representation theorem (Theorem 1) is a statement about the law of $\boldsymbol{X}_t$ at every finite scale, not only about its tail, and it identifies past-representability with the necessary and sufficient condition (16) on the coefficient sequence, which ties anticipativeness to the predictability of extremes (Corollary 1 and Remark 6). The two routes coincide on their common domain, the past-block conditional limits, and the seminorm representation delivers these limits in a single closed-form statement (Proposition 4), uniformly across all past-representable specifications and all seminorms satisfying (10).*

B Forecasting algorithms

This appendix presents the pattern-matching forecasting algorithm for general past-representable MAR(r,s) processes. We first treat in detail the anticipative AR(1) (Appendix B.1), the simplest case in which the tail conditional distribution is non-degenerate and the forecast is therefore probabilistic. We then give the general algorithm (Appendix B.2), which requires at least one anticipative root ($s \geq 1$) for the process to be past-representable, is probabilistic when $s = 1$ and deterministic when $s \geq 2$, and nests the anticipative AR(2), the fractionally integrated process and the SOI specification of Section 5.

B.1 Empirical algorithm for the anticipative AR(1)

When (X_t) is an anticipative AR(1), Proposition 5 shows that the 0–1 conditional tail property of Proposition 6 fails. In the case $0 \leq k_0 \leq h$ (the "ongoing bubble" case), the normalized trajectory converges to a discrete distribution charged at several points: the observed exponential growth pattern at rate ρ^{-1} is compatible with any future peak date in $\{0,\dots,h\}$. Accordingly, the forecast produced by the pattern-matching algorithm for the AR(1) is probabilistic, not deterministic.

We therefore modify Algorithm 1 as follows. Steps 1–6 are unchanged, with $\boldsymbol{d}_k$ computed from the AR(1) coefficients $d_k = \rho^k \mathbb{1}_{\{k\geq 0\}}$. Step 7 (identification of $\hat{k}_0$) remains valid if the resulting $\hat{k}_0 \in \{-m,\dots,-1\}$, because in that case Proposition 5 -$(\iota)(b)$ yields a Dirac limit. When $\hat{k}_0 \in \{0,\dots,h\}$, step 7 is replaced by the computation of the Proposition 5 -$(\iota)(a)$ probabilities

$$p_k := |\rho|^{\alpha k}(1-|\rho|^{\alpha}),\ 0 \leq k \leq h-1, \qquad p_h := |\rho|^{\alpha h}, \tag{46}$$

which are reported together with the point forecasts $\hat{X}_{t+k} = \vartheta_0 \hat{c} \rho^{-k} X_t$.

Remark 10 The mixture $(p_0,\dots,p_h)$ is the full predictive distribution of the peak date; the last step above turns it into a single point decision by fixing a risk threshold $p^\star$ and reporting the first date at which the cumulative crash probability exceeds $p^\star$, e.g. $p^\star = 0.95$ means the user is willing to call a crash once it is deemed at least 95% likely to have occurred by that date. This is a decision rule, not a statistical estimator: $p^\star$ encodes the user's risk tolerance and is set by the practitioner, not estimated from the data, exactly as $p^\star \in \{0.90, 0.95, 0.99\}$ is offered as a menu of choices. The problem of choosing $p^\star$ is nonetheless reminiscent, in a decision-theoretic register, of the classical threshold-selection problem in extreme value statistics: just as selecting a high threshold u for tail-index or extremal-index estimation in regularly varying time series trades off bias against variance (too low a threshold biases the estimate, too high a threshold

Algorithm 2 Pattern-matching forecast for the anticipative AR(1)

Require: observed path, sign ϑ_0, window m, horizon h, upper bound $\bar{k}$, seminorm, and the estimated pair $(\hat{\rho}, \hat{\alpha})$

1: run steps 1–6 of Algorithm 1 with $d_j = \hat{\rho}^j \mathbb{1}_{\{j \geq 0\}}$

2: **if** $\hat{k}_0 \in \{-m, \dots, -1\}$ **then**

3: apply steps 8–10 of Algorithm 1; the forecast is a Dirac (Proposition 5 -$(\iota)(b)$)

4: **else** ▷ $\hat{k}_0 \in \{0, \dots, h\}$: ongoing bubble, no deterministic peak date

5: compute the mixture probabilities $(p_0, p_1, \dots, p_{h-1}, p_h)$ as in (46)

6: report the probabilistic forecast $(\hat{k}, \hat{X}_{t+1}, \dots, \hat{X}_{t+h})$ where $\hat{k} \sim \mathrm{Cat}(p_0, \dots, p_h)$ and $\hat{X}_{t+j} = \vartheta_0 \hat{c}\, \hat{\rho}^{\hat{k}-j} \mathbb{1}_{\{j \leq \hat{k}\}}$

7: (decision-oriented output) fix a risk threshold $p^\star \in (0,1)$ chosen by the user and report the threshold-crossing horizon $\hat{k}^\star := \min\{k \in \{0, \dots, h\} : \sum_{j=0}^{k} p_j > p^\star\}$ as the predicted crash date (Remark 10)

8: **end if**

leaves too few exceedances), selecting $p^\star$ trades off the frequency of false alarms against that of missed or late crash calls. We leave as an open question for future research whether $p^\star$ can be chosen in a statistically principled, data-driven way, for instance by calibrating it on the empirical false-alarm/missed-crash rates of past episodes, rather than fixed a priori as a risk-tolerance parameter.

When $m = 0$ (only the current observation is informative), Proposition 5 -$(\iota\iota)$ adds one more mass point at $\vartheta_0 \boldsymbol{d}_{0,-m} / \|\boldsymbol{d}_{0,-m}\|$ with weight $w_{\vartheta_0}/p_{\vartheta_0}$. Algorithm 2 handles this case by enlarging the support of the categorical distribution in step 6.

Remark 11 The probabilistic outputs in (46) coincide, by Proposition 5 , with the reversal probabilities of Fries (2022), Proposition 4.3. The empirical evaluation of these probabilities is reported in Appendix C.4 of the Supplementary Material (Table 10 therein), where they are computed on simulated MAR(1, 1) trajectories and confronted with the Monte Carlo empirical crash probabilities

B.2 Empirical algorithm for the general MAR(r, s)

For a general MAR(r, s) process

$$\Phi(L)\,\Psi(L^{-1})\,y_t = \varepsilon_t, \qquad \Phi(z) = \prod_{i=1}^{r}(1-\lambda_i z), \qquad \Psi(z^{-1}) = \prod_{j=1}^{s}(1-\zeta_j z^{-1}),$$

the pattern-matching device of Section 5.1 extends beyond the anticipative AR(2), provided the process is past-representable. Two preliminary observations, stated in the remark below, delimit its scope; both are direct consequences of the theory of Section 4 and neither needs to be handled inside the algorithm.

Remark 12 By Corollary 2, a MAR(r, s) process is past-representable if and only if $s = \deg(\Psi) \geq 1$, that is, if and only if it has at least one anticipative (noncausal) root. When $s = 0$ (a purely causal AR), the process is not past-representable and admits no forecast in the seminorm-cylinder sense; the condition $s \geq 1$ is therefore a modelling prerequisite, not something the algorithm verifies. Given $s \geq 1$, the number of anticipative roots governs the nature of the forecast. For $s \geq 2$ the tail conditional distribution is $\{0, 1\}$-valued and the forecast is deterministic (Proposition 6, extended to MAR(r, s) through Corollary 2). For $s = 1$ the future side of the coefficient sequence reduces to a single geometric decay at the rate of the unique anticipative root ζ_1 ($\delta_k \propto \zeta_1^k$ for $k > 0$), so the peak date of an ongoing bubble is left undetermined and the forecast is probabilistic, as in the anticipative AR(1) case of Appendix B.1 (Proposition 5). The crash-odds are then geometric in ζ_1 alone, independently of the causal roots (Fries, 2022), whereas the post-peak shape of each candidate path is still governed by the full coefficient sequence (δ_k) and hence by the causal roots λ_i. The causal order r thus requires no dedicated treatment: it enters only through the decay of the coefficients (δ_k) and is already fully encoded in their closed form below via the causal roots λ_i.

The only modification to Algorithm 1 concerns the construction of the reference vector $\boldsymbol{d}_k$ in its step 3: the closed-form AR(2) coefficients in Equation (30) must be replaced by the general closed-form MA(∞) coefficients of the MAR(r, s), which we denote $(\delta_k)_{k\in\mathbb{Z}}$. These coefficients are derived in de Truchis and Thomas (2026, Lemma 1) via a Laurent-series expansion and the residue theorem: **Noncausal coefficients ($k > 0$).**

$$\delta_k \;=\; (-1)^r \sum_{j=1}^{s} \frac{\zeta_j^{(s-1)+k}}{\prod_{m\neq j}^{s}(\zeta_j - \zeta_m)\;\prod_{i=1}^{r}(\lambda_i\zeta_j - 1)}. \tag{47}$$

Causal coefficients ($k \leq 0$).

$$\delta_k \;=\; (-1)^s \sum_{i=1}^{r} \frac{\lambda_i^{(r-1)+|k|}}{\prod_{l\neq i}^{r}(\lambda_i - \lambda_l)\;\prod_{j=1}^{s}(\lambda_i\zeta_j - 1)}. \tag{48}$$

Consequently, step 3 of Algorithm 1 becomes

$$\boldsymbol{\delta}_k = (\delta_{k+m}, \delta_{k+m-1}, \ldots, \delta_k),\;\; k \in \{-m, \ldots, \bar{k}\},$$

where the δ_j's are computed via (47)–(48), so that the causal roots λ_i enter only through the decay of the coefficients and need no separate handling. All other steps of Algorithm 1 are

unchanged, except for the final output, which follows the $s \geq 2$ versus $s = 1$ dichotomy of Remark 12. The resulting Algorithm 3 is a single procedure covering every past-representable MAR(r, s) ($s \geq 1$) satisfying Assumption 1 of de Truchis and Thomas (2026): the anticipative AR(2) and the anticipative FWN as the $s \geq 2$ deterministic case, and the anticipative AR(1) of Appendix B.1 as the simplest $s = 1$ probabilistic case.

Algorithm 3 Pattern-matching forecast for the general MAR(r, s), $s \geq 1$

Require: observed path, sign ϑ_0, window m, horizon h, upper bound $\bar{k}$, seminorm, tail index $\hat{\alpha}$, estimated roots $(\hat{\lambda}_1, \ldots, \hat{\lambda}_r, \hat{\zeta}_1, \ldots, \hat{\zeta}_s)$ with $s \geq 1$

for $k = -m, -m+1, \ldots, \bar{k}$ **do**
  compute δ_j for $j \in \{k, \ldots, k+m,\ k-1, \ldots, k-h\}$ using (47)–(48)
  set $\boldsymbol{\delta}_k := (\delta_{k+m}, \ldots, \delta_k)$ and $\boldsymbol{u}_k := \vartheta_0 \boldsymbol{\delta}_k / \|\boldsymbol{\delta}_k\|$
  store the distance $e(k) := \|\boldsymbol{v}_t - \boldsymbol{u}_k\|_2$ as in Algorithm 1
end for
identify $\hat{k}_0$ as in step 7 of Algorithm 1
if $s \geq 2$ **then** ▷ deterministic case (Proposition 6, via Corollary 2)
  apply steps 8–11 of Algorithm 1 with $\boldsymbol{d}_k$ replaced by $\boldsymbol{\delta}_k$; the forecast is the single Dirac $\vartheta_0 \boldsymbol{\delta}_{\hat{k}_0} / \|\boldsymbol{\delta}_{\hat{k}_0}\|$
else ▷ $s = 1$: single anticipative root $\hat{\zeta}_1$, probabilistic case (Proposition 5)
  if $\hat{k}_0 \in \{-m, \ldots, -1\}$ **then**
    the burst is already observed: apply steps 8–11 as in the $s \geq 2$ case (Dirac limit)
  else ▷ $\hat{k}_0 \in \{0, \ldots, h\}$: ongoing bubble, peak date undetermined
    report the mixture over the peak date $k \in \{0, \ldots, h\}$, with weights $p_k = |\hat{\zeta}_1|^{\hat{\alpha} k}(1 - |\hat{\zeta}_1|^{\hat{\alpha}})$ for $0 \leq k \leq h-1$ and $p_h = |\hat{\zeta}_1|^{\hat{\alpha} h}$
    to each k attach the δ-based path of steps 8–11 of Algorithm 1 evaluated at that k, i.e. $\hat{X}_{t+j}^{(k)} = \vartheta_0 \, \hat{c}_k \, \delta_{k-j}$, $j = 1, \ldots, h$, so that the post-peak decline is governed by the causal roots through δ_{k-j} (this reduces to Algorithm 2 when $r = 0$)
    (decision-oriented output) as in Algorithm 2, fix a risk threshold $p^\star \in (0, 1)$ and report $\hat{k}^\star := \min\{k \in \{0, \ldots, h\} : \sum_{j=0}^{k} p_j > p^\star\}$ as the predicted crash date (Remark 10)
  end if
end if

C Monte Carlo study and numerical analysis

In this section, we apply our theoretical results by proposing two forecasting procedures and demonstrating their finite-sample performance. Our simulation study focuses on two forecasting exercises: crash date prediction in Appendix C.1, together with the associated robustness checks on the role of m and on the k_0 pattern-matching ties in Appendices C.2 and C.3, and reversal probability prediction in Appendix C.4. Furthermore, Appendix C.5 illustrates the unit cylinder geometry through numerical simulations, while Appendix C.6 presents the associated directional histograms that visualize the concentration patterns in extreme observations.

C.1 Forecasting crash dates

The Monte Carlo exercise of this appendix assesses the finite-sample performance of the pattern-matching crash-date forecast of Algorithm 1 on simulated MAR processes. We do not repeat its steps here; recall only that, by Proposition 6, when X_t is anticipative enough the normalized observed segment satisfies $\boldsymbol{X_t}/\|\boldsymbol{X_t}\| \approx \vartheta_0 \boldsymbol{d}_{k_0}/\|\boldsymbol{d}_{k_0}\|$, which lets Algorithm 1 identify the position k_0 of that segment within the coefficient sequence (d_k) and extrapolate the future path up to the burst date. The sign ϑ_0 is treated as known, as it can be read off the data. Two design parameters are expected to affect the accuracy and are varied below: the window length $m \in \{1, 3, 5, 7, 9, 11\}$ and, since departures from Gaussianity matter (see Remark 2, point $(\iota\iota)$), the tail index $\alpha \in \{0.9, 1.2, 1.5, 1.8\}$.

We simulate three specifications, MAR(0,2), MAR(1,2) and MAR(0,3); the MAR(1,1) is excluded because Proposition 6 does not apply to it, yielding probabilities rather than a degenerate path (the object of Appendix C.4). We first report the purely anticipative AR(2), the specification retained in the empirical Section 5, generated as $X_t = 0.7X_{t+1} + 0.1X_{t+2} + \varepsilon_t$ with $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, 0, 0.5, 0)$. For each simulated path we take the bubble peak as $\max(x_t)$, treat as unobserved the values following it together with the $\lceil N \times 0.01 \rceil$ periods preceding the burst, and run the forecast over 1000 trajectories for $N \in \{250, 500, 1000\}$ (i.e. $k_0 \in \{3, 5, 10\}$). The sample size N leaves the asymptotic performance unchanged but controls the quantile of the last in-sample observation used to trigger the forecast, reported in Table 3: for instance, at $N = 1000$ and $\alpha = 1.5$ that observation sits only at the 0.91 quantile, so the realizations of X_t are only moderately large relative to the asymptotic requirement $x \to \infty$.

Accordingly, in the simulation results, we report the labels "High", "Quite High", "Moderately High", rather than the sample sizes. For each simulation, we compute the bias as the difference between the predicted crash date and the true simulated date.

The results are reported in Table 4. First, they shed light on the crucial role of limit theory,

Table 3: Quantile of the last in-sample observation

N/α	0.9	1.2	1.5	1.8
250	0.99	0.99	0.99	0.94
500	0.98	0.98	0.94	0.89
1000	0.97	0.96	0.91	0.78

as the predicted crash date is considerably more biased when the shape of the trajectory is inferred from an observation that corresponds to a moderately high quantile. Second, for a given m, the performance deteriorates as α increases, thereby involving quantiles far from the asymptotic theory (e.g., $q_{X_t} \approx 0.78$ when $N = 1000$ and $\alpha = 1.8$) and introducing more noise. Our theory states that when $x \to \infty$, $m = 1$ can be sufficient. However, in practice, the simulation study reveals that the optimal selection of m is not obvious, as it interacts in a complex manner with the tail index α. For instance, when X_t is very high and the tail index is close to 1, a shorter m improves the performance of the forecasting procedure. Conversely, if the tail index is close to 2, it is slightly better to select a medium-range m. The same analysis holds for X_t that is large or moderately large.

Table 5 presents the results to determine the crash date for a MAR(1,2) model specified as $X_t = 0.7X_{t+1} + 0.1X_{t+2} + 0.4X_{t-1} + \varepsilon_t$. We want to check if adding a causal part impacts the estimation of the crash date. Theoretically, it should lead to no difference, as only the persistence in the anticipative part matters in estimating the crash date. This is verified, as Tables 4 and 5 display similar results except for the choice of m. For the MAR(1,2), choosing a lower m is always better. We also apply the same procedure to an anticipative AR(3) process, specified as $X_t = 0.8X_{t+1} + 0.2X_{t+2} - 0.1X_{t+3} + \varepsilon_t$, to check if adding persistence in the anticipative part improves the estimation of the crash date. Table 6 showcases the results for this process. Overall, the bias is smaller in any case compared to the anticipative AR(2). However, in comparison to AR(2) and even for the MAR(1,2), the choice of m is not clear-cut. It seems better to choose a large or medium-range m when α is close to 2, and for small α, it is better to choose a small m.

For $\alpha = 1.8$ (the regime closest to the estimated SOI tail index $\hat{\alpha} \approx 1.9$), the bias in the Moderately High regime of Table 4 is close to flat for $m \geq 3$, ranging from 8.355 ($m = 3$) to 8.470 ($m = 11$), a span of about 0.12 relative to a level of 8.4: m indeed matters little there. For lower α, however, the bias increases monotonically and non-negligibly with m: e.g., at $\alpha = 0.9$ it rises from 6.850 ($m = 1$) to 7.542 ($m = 11$), a span of about 0.7. This near-flatness therefore holds specifically in the light-tailed regime relevant to the SOI application, not across the board.

Table 4: Bias for the crash date predictor for the purely anticipative AR(2) process

High						
α/m	1	3	5	7	9	11
0.9	0.1540	0.3590	0.5310	0.5570	0.6730	0.7280
1.2	0.5850	0.7950	0.8500	0.9880	0.9860	1.0040
1.5	1.0170	1.0510	1.1280	1.1500	1.2330	1.2470
1.8	1.2660	1.2480	1.3130	1.3040	1.3200	1.3570
Quite High						
α/m	1	3	5	7	9	11
0.9	1.7670	2.0210	2.2130	2.3010	2.3880	2.4680
1.2	2.5120	2.6530	2.7960	2.8500	2.9580	2.9460
1.5	2.9550	3.0560	3.1160	3.1160	3.1470	3.1750
1.8	3.2970	3.2250	3.2760	3.3070	3.3310	3.2860
Moderately High						
α/m	1	3	5	7	9	11
0.9	6.8500	7.1450	7.2460	7.3340	7.4360	7.5420
1.2	7.6960	7.8270	7.9320	7.9790	8.0530	8.0810
1.5	8.1690	8.1730	8.2160	8.2390	8.2910	8.2860
1.8	8.4230	8.3550	8.3640	8.3790	8.4210	8.4700

Notes: The simulated process is the following:$X_t = 0.7X_{t+1} + 0.1X_{t+2} + \varepsilon_t$, where $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, 0, 0.5, 0)$. High, Quite High, and Moderately High correspond to a number of simulated observations of 250, 500, and 1000, respectively.

This simulation-based pattern is corroborated by the out-of-sample robustness exercise on the real SOI episodes reported in Appendix D.5.

C.2 Frequency of ties in the k_0 pattern-matching step

Algorithm 1 selects $\hat{k}_0 = \arg\min_k \|\boldsymbol{v}_t - \vartheta_0 \boldsymbol{d}_k/\|\boldsymbol{d}_k\|\|_2$. For each of the 1000 replications per (α, m, N) cell described above, we record the two smallest values $e_{(1)} \leq e_{(2)}$ of $e(k) = \|\boldsymbol{v}_t - \vartheta_0 \boldsymbol{d}_k/\|\boldsymbol{d}_k\|\|_2$ and declare a near-tie whenever the relative gap $(e_{(2)} - e_{(1)})/e_{(1)}$ falls below $\eta = 0.05$. Table 7 reports the resulting tie frequency $\eta = 0.05$. Table 7 reports the resulting tie frequency.

Contrary to the intuition that a longer window m should improve the discrimination between candidate patterns $\boldsymbol{d}_k$, the tie frequency increases sharply and essentially monotonically with m, for every α and every regime. This is a geometric feature of the design, not a finite-sample accident: for a smoothly decaying MA(∞) coefficient sequence such as the AR(2)'s, two windows $\boldsymbol{d}_k = (d_{k+m}, \ldots, d_{k-h})$ and $\boldsymbol{d}_{k+1} = (d_{k+1+m}, \ldots, d_{k+1-h})$ overlap in m of their $m+1$ coordinates; as m grows, this overlap dominates the (normalized) shape of both vectors, which

Table 5: Bias for the crash date predictor for the MAR(1,2) process

High						
α/m	1	3	5	7	9	11
0.9	0.3170	0.5870	0.6650	0.7870	0.8420	0.8880
1.2	0.7240	0.9630	1.0490	1.1050	1.1620	1.1180
1.5	1.0880	1.2100	1.2760	1.2320	1.2530	1.2990
1.8	1.2670	1.3400	1.3860	1.3630	1.3720	1.3740
Quite High						
α/m	1	3	5	7	9	11
0.9	1.9840	2.3310	2.5250	2.5570	2.5540	2.6780
1.2	2.6130	2.8560	2.9550	2.9940	3.0280	3.0480
1.5	3.0080	3.2010	3.2180	3.2230	3.3120	3.2940
1.8	3.3040	3.3480	3.3440	3.3550	3.3930	3.4100
Moderately High						
α/m	1	3	5	7	9	11
0.9	7.1650	7.2600	7.5560	7.5900	7.7090	7.6910
1.2	7.8420	7.9810	8.1190	8.1320	8.1760	8.1810
1.5	8.1990	8.2750	8.2870	8.3690	8.3500	8.3910
1.8	8.4080	8.4180	8.4180	8.4300	8.4470	8.4670

Notes: The simulated process is the following:$X_t = 0.7X_{t+1} + 0.1X_{t+2} + 0.4X_{t-1} + \varepsilon_t$, where $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, 0, 0.5, 0)$. High, Quite High, and Moderately High correspond to a number of simulated observations of 250, 500, and 1000, respectively.

Table 6: Bias for the crash date predictor for the purely anticipative AR(3) process

High						
α/m	1	3	5	7	9	11
0.9	-0.3180	-0.3170	-0.4670	-0.6250	-0.6740	-0.5570
1.2	-0.1440	-0.3650	-0.5310	-0.4950	-0.4850	-0.3820
1.5	0.3630	-0.1000	-0.2420	-0.2150	-0.2960	-0.1490
1.8	0.8220	0.2710	-0.0890	-0.1710	0.0480	0.0160
Quite High						
α/m	1	3	5	7	9	11
0.9	0.2200	0.2790	0.4240	0.3490	0.4730	0.5830
1.2	0.9180	0.8680	0.9250	1.0030	1.0500	1.1030
1.5	1.8640	1.6100	1.4960	1.4720	1.5320	1.5520
1.8	2.6780	2.2010	1.8550	1.8870	1.9600	2.0040
Moderately High						
α/m	1	3	5	7	9	11
0.9	3.7370	4.2240	4.4710	4.6080	4.7000	5.0470
1.2	5.5510	5.7180	5.7680	5.8700	5.9220	5.9730
1.5	6.7790	6.5850	6.5550	6.5830	6.6000	6.6310
1.8	7.7710	7.2760	7.0830	7.1390	7.1680	7.1380

Notes: The simulated process is the following:$X_t = 0.8X_{t+1} + 0.2X_{t+2} - 0.1X_{t+3} + \varepsilon_t$, where $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, 0, 0.5, 0)$. High, Quite High, and Moderately High correspond to a number of simulated observations of 250, 500, and 1000, respectively.

Table 7: Frequency of near-ties ($\eta = 0.05$) in the k_0 pattern-matching step, MAR(0,2), 1000 replications per cell.

High						
α/m	1	3	5	7	9	11
0.9	0.045	0.440	0.542	0.656	0.714	0.780
1.2	0.070	0.434	0.588	0.670	0.786	0.803
1.5	0.085	0.497	0.637	0.738	0.787	0.850
1.8	0.104	0.520	0.689	0.794	0.851	0.912
Quite High						
α/m	1	3	5	7	9	11
0.9	0.055	0.455	0.583	0.694	0.733	0.785
1.2	0.066	0.455	0.604	0.690	0.771	0.795
1.5	0.075	0.463	0.658	0.746	0.811	0.842
1.8	0.162	0.606	0.766	0.821	0.866	0.908
Moderately High						
α/m	1	3	5	7	9	11
0.9	0.067	0.476	0.587	0.668	0.762	0.801
1.2	0.081	0.503	0.635	0.727	0.792	0.827
1.5	0.160	0.541	0.709	0.773	0.842	0.879
1.8	0.349	0.703	0.807	0.884	0.922	0.954

become increasingly collinear regardless of sample size. Enlarging m therefore does not make the pattern-matching step more discriminating in this design; if anything, it makes ties more frequent. This has a direct consequence for the empirical application. The baseline $m = 10$ used for the SOI in Section 5 falls between the $m = 9$ and $m = 11$ columns of Table 7. In the Moderately High regime with $\alpha = 1.8$ (closest to $\hat{\alpha} \approx 1.9$), the tie frequency is already 0.922 at $m = 9$ and 0.954 at $m = 11$. In other words, at the window length actually retained for the empirical application, ties are the rule rather than the exception: the tie-breaking rule of Section 5.4 is not a safeguard for a marginal edge case, it is the mechanism that resolves the pattern-matching step in the vast majority of replications, not a rare occurrence. This should be qualified, however, by the out-of-sample robustness exercise of Appendix D.5: on the real SOI episodes, the forecast remains reasonably accurate whichever window length is used, provided the theoretical minimum $m \geq 1$ is respected. Ties are therefore frequent but not necessarily harmful to forecast accuracy. We leave to future research a proper statistical treatment of the choice of m, of the tie frequency, and more generally of statistical inference

for this pattern-matching approach and its geometric properties; our purpose here is only to illustrate the practical applicability of the theoretical probability results derived in the main paper.

C.3 Does the tie-breaking rule matter? Comparison with the second-best candidate

Section 5.4 breaks ties by retaining the candidate with the smallest $|k|$, on the grounds that it corresponds to the piece of trajectory closest to the current date and hence the most informative one. Table 8 compares, conditional on a declared tie ($\eta = 0.05$), the mean bias obtained under this rule ($\hat{k}_0$) against the mean bias that would be obtained by retaining the second-best candidate instead, for $\alpha = 1.8$.

Table 8: Mean bias of the retained candidate ($\hat{k}_0$, arg-min rule) versus the second-best candidate, conditional on a declared tie ($\eta = 0.05$), for $\alpha = 1.8$, MAR(0,2).

Regime	m	1	3	5	7	9	11
High	$\hat{k}_0$ bias	1.298	0.940	1.000	1.122	1.248	1.306
	2nd-best bias	-0.625	-1.052	-0.975	-0.868	-0.711	-0.671
	tie freq.	0.104	0.520	0.689	0.794	0.851	0.912
Quite High	$\hat{k}_0$ bias	3.333	3.012	3.030	3.139	3.263	3.313
	2nd-best bias	1.506	1.018	1.042	1.157	1.301	1.343
	tie freq.	0.162	0.606	0.766	0.821	0.866	0.908
Moderately High	$\hat{k}_0$ bias	8.524	8.216	8.305	8.290	8.368	8.433
	2nd-best bias	6.559	6.222	6.340	6.304	6.418	6.455
	tie freq.	0.349	0.703	0.807	0.884	0.922	0.954

Both rules carry a substantial, comparable-magnitude bias in absolute terms (the reference crash date is not recovered exactly in either case, which is expected outside the strict $x \to \infty$ asymptotic regime). The retained rule ($\hat{k}_0$) is not uniformly closer to zero bias than the alternative: in the Moderately High regime, for instance, $\hat{k}_0$'s bias is somewhat larger in absolute value than the second-best candidate's for every m reported. We do not interpret this as evidence that the current rule is wrong, the "smallest $|k|$" criterion remains a reasonable default on qualitative grounds, but the numerical evidence for its superiority over the natural alternative is, at present, weak. This should temper any claim that the specific tie-breaking convention is empirically validated; it is a practical convention, not a data-driven choice.

C.4 Forecasting reversal probabilities

Our main theoretical result (Proposition 6) demonstrates that for sufficiently anticipative processes such as the AR(2), the conditional distribution of the future path degenerates asymptotically to a single pattern, indexed by k_0, under the assumed model. However, accurately identifying k_0 from finite, potentially noisy observations can be challenging (see Remark 2). In this section, we evaluate a simpler, probabilistic forecasting approach that serves as a benchmark. This method does not require the identification of k_0 and instead focuses on the question: "Given that an extreme event is detected, based solely on the magnitude of the recent past trajectory exceeding a quantile q and a known sign ϑ_0, what is the average probability of a reversal (crash) at horizon h ?"This approach is in the spirit of Proposition 4.3 in Fries (2022), adapted to our framework where conditioning uses the seminorm of the past trajectory vector. Since we condition only on magnitude and sign, without pinpointing the specific pattern k_0, the resulting forecast is a probability reflecting an average over all possible underlying patterns consistent with the observed magnitude.

We simulate $M = 100$ trajectories of length $N = 10^6$ observations for four different specifications of MAR processes: MAR(1,1), MAR(0,2), MAR(1,2), and MAR(0,3). The large sample size ensures sufficient observations in the extreme tails, allowing us to evaluate probabilities conditional on high quantiles (up to 99.99%). We focus here on the purely anticipative AR(2) (MAR(0,2)) process: $X_t = 0.7X_{t+1} + 0.1X_{t+2} + \varepsilon_t$, where $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(1.5, 0, 0.5, 0)$. We set the past trajectory length $m = 2$ and its magnitude using the seminorm $||\boldsymbol{X_t}||$. We evaluate reversal probabilities at horizons $h \in \{1, 5, 10\}$.

We define the two following sets. The conditioning set C_{q,ϑ_0} includes all time points t where the magnitude of the past trajectory exceeds the q-th quantile of its distribution and the sign matches ϑ_0

$$C_{\vartheta_0}(q) = \{t : \|\boldsymbol{X_t}\| \geq q \text{ and } \operatorname{sign}(X_t) = \vartheta_0\}.$$

We consider $\vartheta_0 = 1$ for positive bubbles. Furthermore, q represents the empirical quantile (e.g., 90th, 99th percentile) of the distribution of X_t estimated from the simulated data. The reversal event set $R_h(\delta)$ includes any time point t where the normalized future value $X_{t+h}/\|\boldsymbol{X_t}\|$ returns to the central region $[-\delta, \delta]$ at horizon h

$$R_h(\delta) = \left\{t : \left|\frac{X_{t+h}}{\|\boldsymbol{X_t}\|}\right| \leq \delta\right\}$$

Consistent with Fries (2022), we set the threshold $\delta = 0.2$. We estimate the conditional proba-

bility of reversal $\hat{p}_{q,\vartheta_0}(h)$ using the relative frequency over the simulation sample:

$$\hat{p}_{q,\vartheta_0}(h) = \frac{\sum_{t=1}^{N-h} \mathbb{1}_{\left\{t \in R_h(\delta) \cap C_{\vartheta_0}(q)\right\}}}{\sum_{t=1}^{N-h} \mathbb{1}_{\left\{t \in C_{\vartheta_0}(q)\right\}}} \tag{49}$$

This estimator measures the probability

$$\mathbb{P}\left(\left|\frac{X_{t+h}}{\|\boldsymbol{X_t}\|}\right| \leq \delta \;\middle|\; \|\boldsymbol{X_t}\| \geq q, \operatorname{sign}(X_t) = \vartheta_0\right).$$

Note that while a theoretical limit for this probability exists as $q \to \infty$, its explicit formula is complex due to the seminorm conditioning and averaging over patterns. Hence, we cannot provide an analog version of the $p_{\infty,h}$ reported in Fries (2022).

Table 9 reports the average $\hat{p}_q$ (across $M = 100$ simulations) and the corresponding 95% empirical confidence intervals for the MAR(0,2) process.

Table 9: Empirical crash probabilities of bubbles generated by the MAR(0,2)

	$h = 1$	$h = 5$	$h = 10$
$\hat{p}_{0.9}$ (95% CI)	17.247 (17.061-17.414)	52.369 (51.946-52.794)	69.239 (68.742-69.677)
$\hat{p}_{0.99}$ (95% CI)	27.413 (26.490-28.191)	73.396 (71.932-74.640)	90.438 (89.254-91.511)
$\hat{p}_{0.999}$ (95% CI)	29.490 (26.763-33.122)	77.712 (72.964-82.796)	94.708 (92.124-97.136)
$\hat{p}_{0.9999}$ (95% CI)	30.141 (24.000-38.017)	79.247 (69.003-89.492)	95.782 (87.993-100.000)

Notes: Empirical average (Mean) and 95% confidence intervals (95%-CI) of the estimated probabilities $\hat{p}_q$ computed using (49) on M = 100 simulated trajectories of $N = 10^6$ observations, for q several empirical quantiles of $||\boldsymbol{X_t}||$. These probabilities are reported in percent. We set $m = 2$. The simulated process is: $X_t = 0.7X_{t+1} + 0.1X_{t+2} + \varepsilon_t$ where $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(1.5, 0, 0.5, 0)$

One notices that the empirical probabilities $\hat{p}_q$ converge as the quantile q increases, approaching a limit probability that is generally not 0 or 1. For instance, at $h = 5$, $\hat{p}_q$ increases from 52.4% at the 90th percentile to 79.2% at the 99.99th percentile. This limit represents the average probability of reversal conditional on observing an extreme magnitude, averaged over all patterns k compatible with such magnitude. Unsurprisingly, the confidence intervals widen considerably for the highest quantiles due to the scarcity of observations. The analysis is repeated for the MAR(1,1), MAR(1,2), and MAR(0,3) processes in Tables 10, 11, and 12, respectively.

The results for the MAR(1,1) process, presented in Table 10, exhibit markedly different characteristics compared to the purely anticipative AR($p \geq 2$) processes. Notably, the empirical reversal probabilities $\hat{p}_q$ are substantially lower, particularly at shorter horizons (e.g., $h = 1$ and $h = 5$), even for very high quantiles q. For instance, at $q = 0.9999$ and $h = 5$, $\hat{p}_q$ is only 0.053%, compared to 79.2% for the MAR(0,2) process under similar conditions (Table 9). This empirical finding aligns directly with the theoretical predictions outlined in Proposition 5 (Case

Table 10: Comparison of empirical crash probabilities at horizons $h = 1, 5, 10$ of bubbles generated by the MAR(1,1)

	$h = 1$	$h = 5$	$h = 10$
$\hat{p}_{0.9}$ (95% CI)	0.478 (0.441-0.518)	19.961 (19.563-20.414)	53.569 (52.96-54.266)
$\hat{p}_{0.99}$ (95% CI)	0.064 (0.017-0.126)	5.005 (4.181-5.728)	64.874 (63.214-66.255)
$\hat{p}_{0.999}$ (95% CI)	0.007 (0.000-0.109)	0.521 (0.000-1.559)	69.284 (64.843-72.694)
$\hat{p}_{0.9999}$ (95% CI)	0.006 (0.000-0.010)	0.053 (0.000-0.567)	67.561 (53.115-75.074)

Notes: Empirical average (Mean) and 95% confidence intervals (95%-CI) computed using (49) on M = 100 simulated trajectories of $N = 10^6$ observations, for q several empirical quantiles of $||\boldsymbol{X_t}^{past}||$. We set m=2. The simulated process is $X_t = 0.9X_{t+1} + 0.1X_{t-1} + \varepsilon_t$, where $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(1.5, 0, 0.5, 0)$

ι, a). Unlike the MAR(0,2) process where observing the past pattern $f(\boldsymbol{X_t})/\|\boldsymbol{X_t}\|$ (for $m \geq 1$) uniquely identifies the underlying index k_0 and leads to a deterministic $\{0, 1\}$-valued conditional tail distribution (Proposition 6), the MAR(1,1) structure retains inherent uncertainty. Specifically, Proposition 5 shows that even when the past pattern corresponding to an ongoing exponential growth phase ($k_0 \in \{0, \ldots, h\}$) is identified, the future path is not deterministic. The conditional distribution assigns distinct, non-degenerate probabilities to different future scenarios: peaking at horizon $k \in \{0, \ldots, h-1\}$ with probability $|\rho|^{\alpha k}(1 - |\rho|^{\alpha})$, or continuing beyond horizon h with probability $|\rho|^{\alpha h}$. Therefore, the low empirical probabilities observed in Table 10 reflect this fundamental property: even conditioning on extreme magnitude $\|\boldsymbol{X_t}\| \geq q$ (which implicitly selects trajectories following the exponential growth pattern), the probability of an imminent reversal ($k < h$) remains low due to the high persistence ($\rho = 0.9$) chosen for the anticipative component in this simulation. This contrasts sharply with the MAR(0,2) case, where identifying the pattern (even implicitly through high magnitude conditioning) quickly resolves the uncertainty, leading to high reversal probabilities once the trajectory approaches its peak. This highlights that while magnitude-based conditioning provides probabilistic insights for the MAR(1,1), it cannot achieve the deterministic predictability achievable for processes covered by Proposition 6.

Overall, the simulations confirm that conditioning solely on the magnitude of the past trajectory provides meaningful probabilistic forecasts of reversals, which converge as the conditioning quantile increases. This benchmark highlights the information contained in the magnitude alone, setting the stage for the asymptotically deterministic forecasts achievable, under the assumed model, through pattern identification as demonstrated in Appendix C.1.

Table 11: Comparison of empirical crash probabilities at horizons $h = 1, 5, 10$ of bubbles generated by the MAR(1,2)

	$h = 1$	$h = 5$	$h = 10$
$\hat{p}_{0.9}$ (95% CI)	10.429 (10.133-10.743)	59.71 (58.816-60.517)	82.535 (81.343-83.754)
$\hat{p}_{0.99}$ (95% CI)	9.066 (8.436-9.815)	66.83 (63.998-69.511)	90.299 (87.706-92.689)
$\hat{p}_{0.999}$ (95% CI)	8.534 (5.782-10.661)	67.486 (59.013-78.226)	91.695 (83.716-99.396)
$\hat{p}_{0.9999}$ (95% CI)	7.815 (0.000-13.437)	68.23 (45.776-100.000)	92.863 (70.983-100.000)

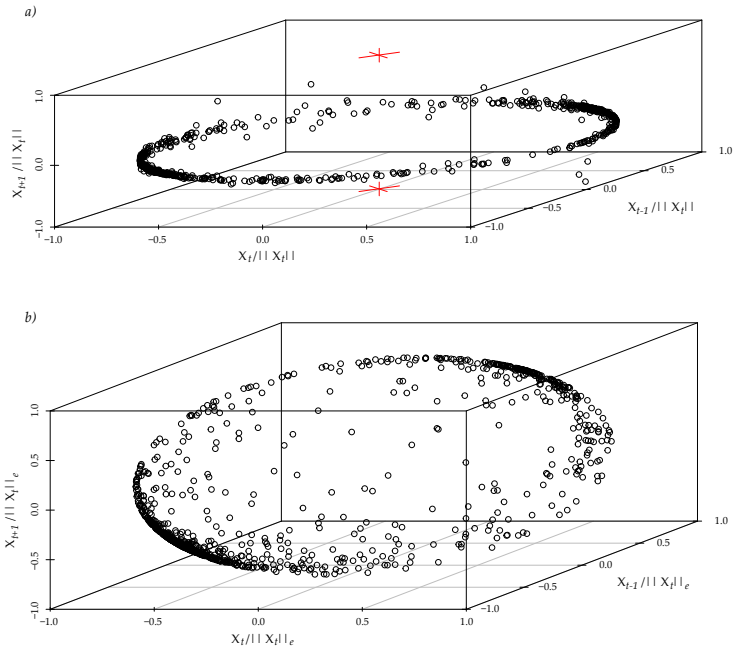
Notes: Empirical average (Mean) and 95% confidence intervals (95%-CI) computed using (49) on M = 100 simulated trajectories of $N = 10^6$ observations, for q several empirical quantiles of $||\boldsymbol{X_t}^{past}||$. We set m=2. The simulated process is $X_t = 0.7X_{t+1} + 0.1X_{t+2} + 0.4X_{t-1} + \varepsilon_t$, where $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(1.5, 0, 0.5, 0)$

Table 12: Comparison of empirical crash probabilities at horizons $h = 1, 5, 10$ of bubbles generated by the anticipative AR(3)

	$h = 1$	$h = 5$	$h = 10$
$\hat{p}_{0.9}$ (95% CI)	8.599 (8.433-8.743)	34.573 (34.128-35.056)	54.792 (54.087-55.440)
$\hat{p}_{0.99}$ (95% CI)	13.740 (13.093-14.345)	50.585 (48.61-52.659)	75.168 (73.035-77.219)
$\hat{p}_{0.999}$ (95% CI)	14.612 (12.53-16.693)	53.400 (47.320-58.863)	78.857 (72.304-85.104)
$\hat{p}_{0.9999}$ (95% CI)	15.287 (10.716-22.733)	54.813 (41.167-72.491)	79.757 (63.015-99.580)

Notes: Empirical average (Mean) and 95% confidence intervals (95%-CI) computed using (49) on M = 100 simulated trajectories of $N = 10^6$ observations, for q several empirical quantiles of $||\boldsymbol{X_t}^{past}||$. We set m=2. The simulated process is $X_t = 0.8X_{t+1} + 0.2X_{t+2} - 0.1X_{t+3} + \varepsilon_t$, where $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(1.5, 0, 0.5, 0)$

Figure 2: Unit cylinder and unit sphere representations of $X_t = 0.7X_{t+1} + 0.1X_{t+2} + \varepsilon_t$



C.5 Visualisation of the unit cylinder

In the spirit of the information-encoding interpretation of spectral measures, we consider an α-stable vector $\boldsymbol{X_t} = (X_{t-1}, X_t, X_{t+1})$ where X_t is an anticipative AR(2) specified as in C.4. X_t being past-representable, it admits a representation on the unit-cylinder, as the Theorem 1 applies. Furthermore, as discussed in 6, its spectral measure exhibits the following asymptotic behavior

$$\mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t}, A\Big|B(V_0)\Big) \underset{x\to\infty}{\longrightarrow} \frac{\Gamma^{\|\cdot\|}\left(A \cap \left\{\frac{\vartheta_0 \boldsymbol{d}_{k_0}}{\|\boldsymbol{d}_{k_0}\|}\right\}\right)}{\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta_0 \boldsymbol{d}_{k_0}}{\|\boldsymbol{d}_{k_0}\|}\right\}\right)}.$$

and hence $\mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t}, A\Big|B(V_0)\Big)$ is either 1 or 0. This peculiar $\{0,1\}$ tail conditional distribution leads to the following graphical representation on the unit-cylinder (see Figure 2.a). The simulation of X_t is performed for a sample size $n = 1000$.

We clearly see that $C_3^{\|\cdot\|}$ spans all directions of $\mathbb{R}^3$ but the ones of $(0,0,-1)$ and $(0,0,+1)$. This is of no consequence as the representability property holds and implies that $\Gamma\Big(\{(0,0,-1),(0,0,+1)\}\Big) = 0$ as $x \to \infty$. In other words, the seminorm representability reflect the fact that extreme realizations of X_{t+1} never occur conditionally to small realizations of X_{t-1} and X_t. Those inaccessible coordinates are indicated by the two red cross. In the opposite case where we represent $\boldsymbol{X_t}$ on the unit sphere, S_3 spans all directions of $\mathbb{R}^3$ and describes any potential tail dependence of (X_{t-1}, X_t, X_{t+1}). This includes the tail dependence between X_{t+1}

and the past, which reflects the odd (and rare, as depicted in Figure 2.b) situation where the realisation of X_{t+1} is extreme whereas immediate past realizations are not.

C.6 Directional histograms

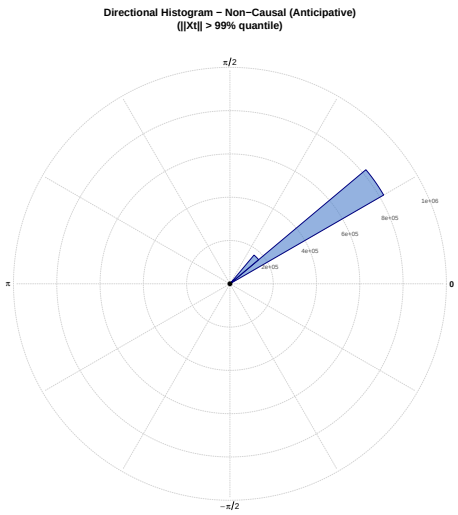
We present here the directional histograms that visualize the unit cylinder representation for extreme tail observations. These histograms provide numerical evidence for the theoretical results established in Section 2 and demonstrate the fundamental difference between causal and anticipative processes when extreme events occur. The directional histograms are constructed using simulated data from both causal and non-causal AR(2) processes with α-stable innovations. Specifically, we consider:

$$\begin{aligned} \text{MAR(0,2):} \quad & X_t = 0.7X_{t+1} + 0.1X_{t+2} + \varepsilon_t, \\ \text{MAR(2,0):} \quad & X_t = 0.7X_{t-1} + 0.1X_{t-2} + \varepsilon_t, \end{aligned}$$

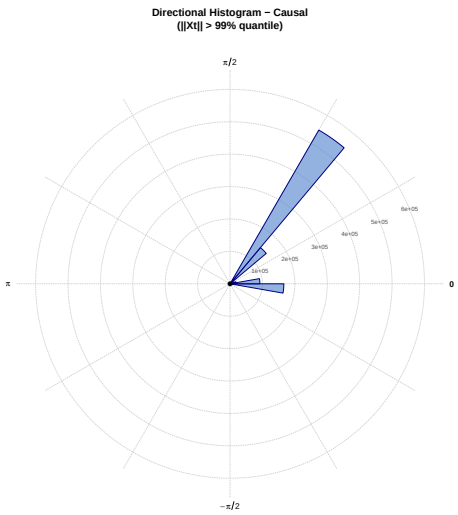
where $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(1.5, 1, 0.5, 0)$ represents α-stable innovations with tail index $\alpha = 1.5$, skewness parameter $\beta = 1$, scale parameter $\sigma = 0.5$, and location parameter $\mu = 0$. For each process, we simulate $N = 10^8$ observations to ensure sufficient data points in the extreme tail. The directional histograms are constructed by first filtering observations where $\|X_t\|$ exceeds the 99% quantile to focus on extreme events. For these filtered observations, we compute the angles $\theta_t = \arctan(X_{t-1}/X_t)$ for the bivariate vectors (X_t, X_{t-1}), which represent the directional information of consecutive observations. The angular domain $[-\pi, \pi]$ is then partitioned into 36 equal bins to create a detailed angular resolution, and the frequency counts within each bin are displayed as radial bars in polar coordinates to visualize the directional concentration patterns.

In Figure 3a, the extreme observations exhibit a highly concentrated angular distribution, with the majority of mass concentrated around specific angular directions. This concentration pattern reflects the deterministic nature of the tail conditional distribution established in Proposition 6. The directional concentration provides visual confirmation that extreme trajectories in anticipative processes follow predictable patterns. In contrast, Figure 3b shows that the causal process displays a more dispersed angular distribution across the unit circle. While some directional preference is observed, the distribution is significantly less concentrated than in the non-causal case. This dispersion is consistent with the theoretical result that causal processes do not admit seminorm representations (Corollary 1), particularly as it charges the x-axis with y zero coordinates.

To further examine the robustness of the directional concentration patterns, we also construct histograms using alternative lag structures. Specifically, we consider the bivariate vectors



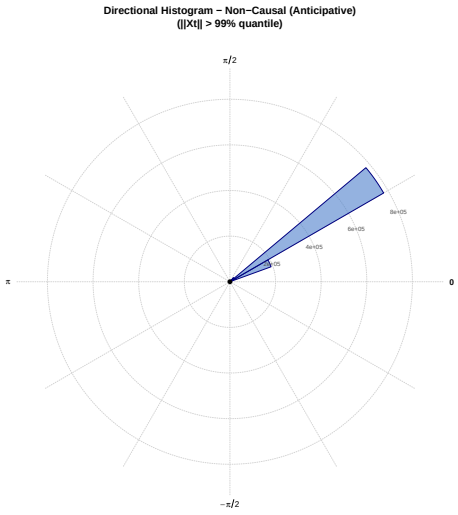
(a) MAR(0,2)



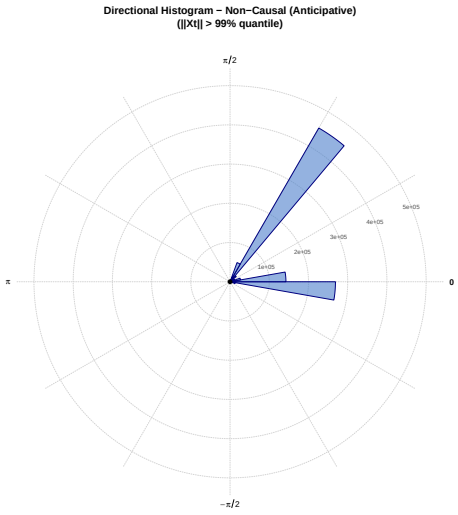
(b) MAR(2,0)

Figure 3: *Directional histograms for extreme observations ($\|X_t\| > 99\%$ quantile) from causal and non-causal AR(2) processes. The histograms display the angular distribution of bivariate vectors (X_t, X_{t-1}) in polar coordinates, where the radial distance represents frequency counts and the angular coordinate represents* $\arctan(X_{t-1}/X_t)$.

(X_t, X_{t-2}) instead of (X_t, X_{t-1}), which allows us to investigate whether the directional patterns persist across different temporal relationships within the same process.



(a) MAR(0,2): (X_t, X_{t-2})



(b) MAR(2,0): (X_t, X_{t-2})

Figure 4: *Directional histograms for alternative lag structures, showing the robustness of the directional concentration patterns in non-causal processes across different temporal relationships.*

Figure 4 presents additional directional histograms with these alternative lag structures to demonstrate the robustness of our findings. The persistent concentration patterns in the

non-causal case across different lag specifications further validate the theoretical predictions regarding the deterministic nature of extreme trajectories in anticipative processes. Notably, even when considering the relationship between X_t and X_{t-2}, the directional concentration remains pronounced for the non-causal process, while the causal process continues to exhibit even more dispersed angular patterns.

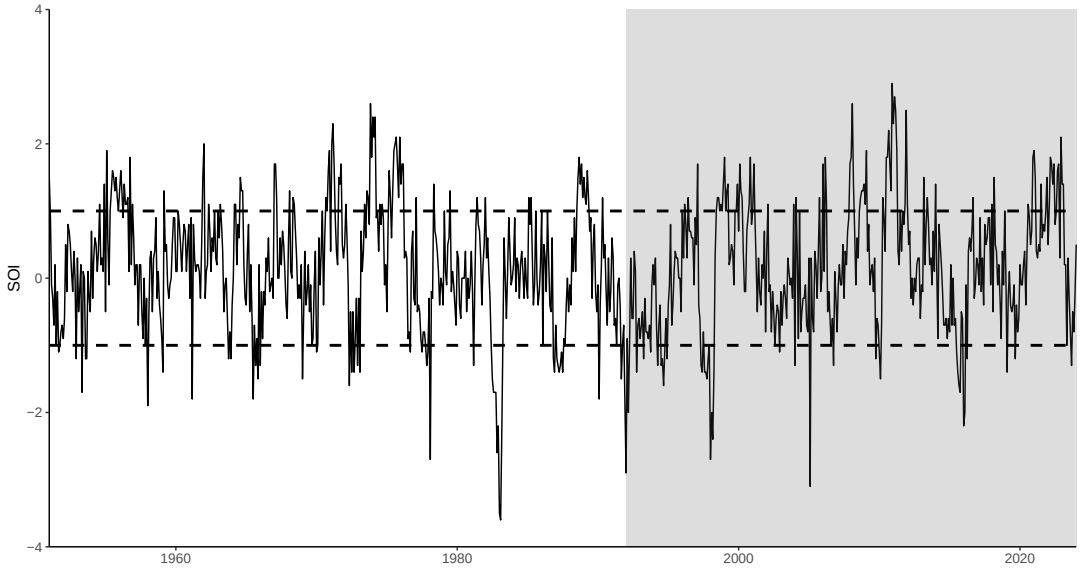
D Complementary results for the empirical application

This appendix presents the estimation results for the main empirical application discussed in Section 5 It includes the GCov estimation of Gourieroux and Jasiak (2023), presented in Appendix D.2, and the parameter estimates obtained using the alternative approximate maximum likelihood (AML) method of Lanne and Saikkonen (2011), adapted to α-stable distributions, presented in Appendix D.3.

D.1 Distributional Properties and Autocorrelation Structure of the SOI

Figure 5 shows that SOI values range approximately from -4 to 2. Most observations are clustered around 0, reflecting the index's central tendency, while fewer observations are found at the extremes, underscoring the rarity of extreme El Niño (SOI below -1) or La Niña (SOI above 1) events. Notably, an SOI of -1 corresponds to the 10th percentile, while 1 corresponds to the 80th percentile.

Figure 5: Southern Oscillation Index (SOI)

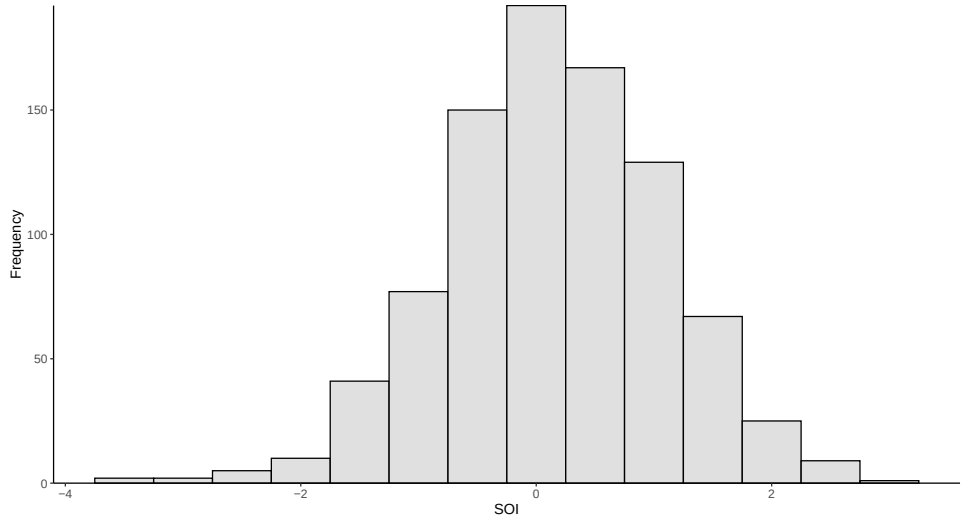


Notes: The shaded area represents the out-of-sample data. The horizontal black dashed lines indicate values of 1 *and* -1*. An El Niño event is defined as an SOI below* -1*, while a La Niña event corresponds to an SOI above* 1*, each persisting for at least three periods.*

To assess the distributional properties of the SOI, Figure 6 presents the histogram of the data. The distribution exhibits slight negative skewness (-0.2024) and excess kurtosis (3.4823), suggesting departures from normality. To formally test for non-Gaussianity, we conduct a Jarque-Bera test. The results, summarized in Table 13, yield a test statistic of $X^2 = 14.4890$

with a p-value of 0.0007, strongly rejecting the null hypothesis of normality at conventional significance levels.

Figure 6: Distribution of the Southern Oscillation Index (SOI)



Notes: The histogram shows that most observations are clustered around 0, with fewer observations at the extremes, confirming the rarity of extreme El Niño and La Niña events.

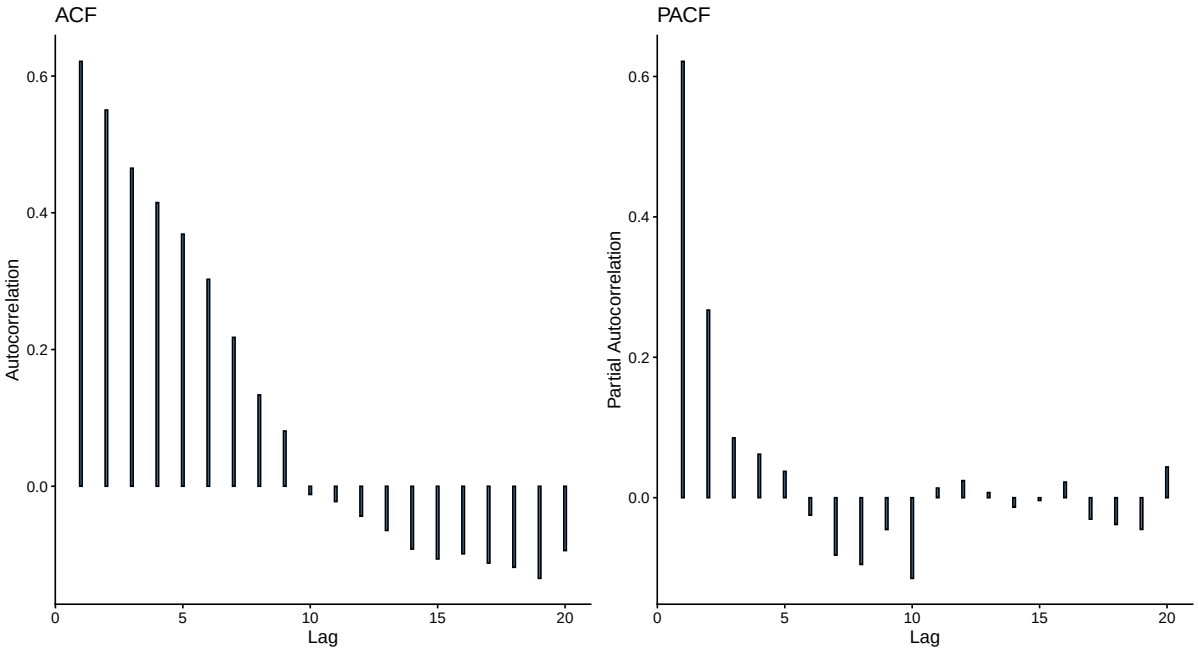
Table 13: Descriptive Statistics and Jarque-Bera Test Results for SOI Data

Statistic	**Value**
Skewness	−0.2024
Kurtosis	3.4823
Jarque-Bera Statistic (X^2)	14.4890
Degrees of Freedom	2
p-value	0.0007

Figure 7 displays the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) for the in-sample SOI data, without any confidence band (see the discussion above on why the classical Gaussian band is not appropriate here, and why we do not substitute a simulated one). The PACF exhibits pronounced spikes at lags 1 and 2 followed by a sharp drop, suggesting an autoregressive structure of order 2. Based on this analysis, we set $p_{\text{ap}} = p+q = 2$. This specification is further examined using the Akaike and Bayesian Information Criteria (AIC, BIC), as reported in Table 15 and discussed further, together with an order-3 robustness check, in Appendix D.4.

A nonparametric check of heavy tails. The Jarque-Bera test above establishes non-

Figure 7: ACF and PACF of the In-Sample SOI Data

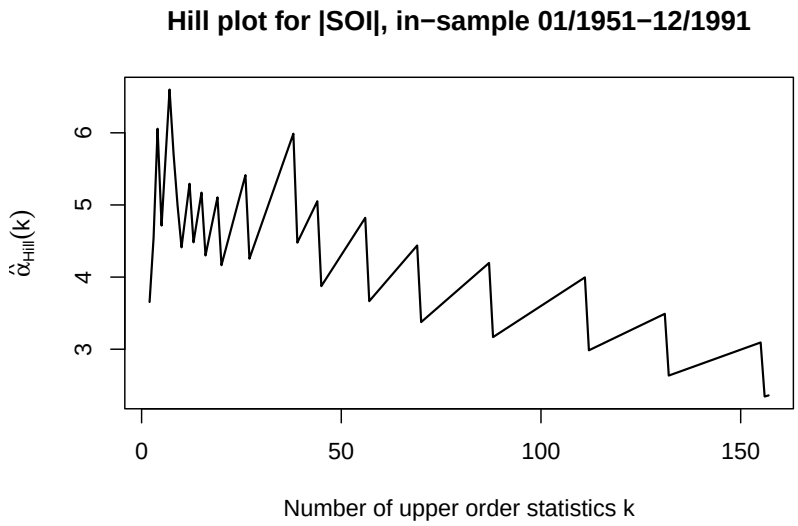


Gaussianity but not heavy-tailedness per se, a platykurtic or skewed light-tailed distribution would also reject normality. As a distributional-assumption-free check, we compute the Hill estimator (Hill, 1975) of the tail index on the in-sample SOI series ($n = 492$), applied to $|X_t|$. Figure 8 shows the Hill plot $\hat{\alpha}_{\text{Hill}}(k)$ for k ranging from the top 2 to the top 164 (one third of the sample) order statistics. The estimate ranges from $\hat{\alpha}_{\text{Hill}}(30) = 4.73$ down to $\hat{\alpha}_{\text{Hill}}(100) = 3.60$, i.e. noticeably higher (lighter tail) than the $\hat{\alpha} \approx 1.94$–$1.98$ obtained from the parametric α-stable fit (Tables 16 and 14). Estimating the positive and negative tails separately gives $\hat{\alpha}_{\text{Hill}}(30) = 4.71$ (SOI$>$ 0) versus 3.48 (SOI$<$ 0), a gap broadly consistent with the small negative skewness reported above.

We do not read this exercise as confirming the α-stable, infinite-variance assumption underlying our model; if anything, it tempers it. It is a well-documented fact that the Hill estimator is itself biased and unstable when the true tail index is close to 2, since the regular-variation approximation on which it relies only holds asymptotically far into the tail (Embrechts et al., 1997). Independent support for a moderate, finite-variance form of heavy-tailedness in the SOI comes from Ausloos and Ivanova (2001), who fit a power law $\mathbb{P}(X > x) \sim x^{-\mu}$ to SOI increments and estimate $\mu = 3.30 \pm 0.06$ for fluctuation amplitudes between 1.5 and 2.8: this value convincingly rejects the Gaussian ($\mu = 2$) but is also well outside the range $\mu \in (0, 2)$ that a genuinely infinite-variance, non-Gaussian stable law would require. Taken together, our Hill exercise and this external estimate suggest that the SOI is heavy-tailed relative to the Gaussian, but plausibly has finite variance, the strict α-stable, infinite-variance assumption we adopt

(Tables 16, 14) is best understood as a tractable parametric idealization rather than a literal claim about the population moments of the SOI.

Figure 8: Hill plot for $|X_t|$, in-sample SOI, 01/1951–12/1991



Notes: $\hat{\alpha}_{\text{Hill}}(k)$ *computed on* $|X_t|$ *as a function of the number* k *of upper order statistics used.*

Why adopt the α-stable assumption regardless? Given that the evidence above does not pin down the SOI as literally infinite-variance, one may ask why we retain the α-stable specification rather than a more general finite-variance heavy-tailed model. The reason is not descriptive accuracy but analytical tractability of the forecast. Building predictive densities for mixed causal-noncausal processes has long been considered difficult precisely because of their dependence on future values, and much of the existing literature relies on simulation-based, Bayesian, or otherwise numerical procedures, e.g. Lanne et al. (2012) or Gouriéroux and Jasiak (2016), which Hecq and Voisin (2021) document can become computationally demanding at long horizons and fail to capture tail dynamics accurately. Even Gourieroux and Jasiak (2026), who derive an exact closed-form one-step-ahead predictive density from the Markov property of causal-noncausal VAR models, still rely on a numerical kernel to approximate that density in practice; and Dumitrescu et al. (2026) push the numerical route further by training a Mixture Density Network to approximate the predictive density directly from data, without assuming α-stability, at the cost of retraining whenever the estimated process or sample changes. The CEV-based framework of Gouriéroux et al. (2025) ultimately relies on nonparametric estimation of the tail index and asymmetry parameters governing the limit measure, rather than on an explicit closed-form expression tied to a single distributional assumption. Among all these frameworks, the α-stable assumption is, to our knowledge, the only one that yields an exact,

closed-form characterization of the tail conditional distribution of the future path (Proposition 6 and its counterparts), reducing forecasting to a single explicit formula. The α-stable assumption should therefore be read as a deliberate trade-off: it sacrifices some descriptive realism about the exact tail index in exchange for an exact, computationally trivial forecasting formula, which is the object this paper is built around.

Do ACF/PACF, BIC, Jarque-Bera and Ljung-Box remain valid under infinite variance? The specification search reported below uses the sample ACF/PACF of the raw series to suggest an order, the BIC to confirm it, and the Jarque-Bera and Ljung-Box statistics on the residuals to check the fit. All four tools are classically justified under a finite-variance, typically Gaussian, asymptotic regime, which is precisely the regime we reject for the SOI.

Sample ACF/PACF. Davis and Resnick (1986) establish the limit theory for the sample autocovariance and autocorrelation functions of linear processes with regularly varying, possibly infinite-variance, innovations (building on the moving-average limit theory of Davis and Resnick (1985)); the sample ACF/PACF remain consistent for their population counterparts in this setting, but their fluctuations are governed by a stable, not Gaussian, limit law with a slower-than-$\sqrt{n}$ rate when $\alpha < 2$, so that the classical $\pm 1.96/\sqrt{n}$ band is not the correct finite-sample benchmark. Following the practice of Andrews et al. (2009), who fit noncausal α-stable AR models and explicitly note that "even if a process has infinite second-order moments, the sample correlations and partial correlations can still be useful for identifying a suitable model for the data" (their Section 4.2, citing. Adler, Feldman and Gallagher, 1998), we use the sample ACF/PACF of the SOI only as an informal, exploratory device to suggest the order p_{ap}, not as a formal test with a confidence-band statement attached to any individual bar. Accordingly, Figures 7 and 11 report the raw-series and residual ACF/PACF without any confidence band, neither the classical (invalid) Gaussian one, nor a simulated or bootstrapped one. A band obtained by simulating from the fitted α-stable model, in the spirit of Andrews et al. (2009)'s own residual diagnostics (their Figure 4b–c), could certainly be constructed; we do not report one here because the pattern that matters for order selection, a sharp drop in magnitude from lag 2 onward, is visually unambiguous in Figure 7, and is subsequently confirmed by evidence that does not depend on any band construction: the causal/noncausal log-likelihood comparison, the portmanteau/GCoV residual checks, and the AIC/BIC robustness comparison against an order-3 alternative, all reported in Appendix D.4.

BIC. The classical consistency of BIC for autoregressive order selection is proved under finite-variance, typically Gaussian, likelihood asymptotics. To our knowledge no analogous consistency result for BIC has been established for infinite-variance autoregressive processes.

By contrast, Knight (1989) proves that AIC is a consistent order-selection criterion for infinite-variance (heavy-tailed) autoregressive processes, even though it is inconsistent in the finite-variance case. We therefore recompute the order selection using AIC alongside BIC in Appendix D.4. For the SOI, AIC and BIC do not fully agree: BIC is minimized at $p_{ap} = 2$ (as reported in Table 15), while AIC is minimized at $p_{ap} = 3$, with AIC/BIC/HQ values that are, however, nearly flat over $p_{ap} \in \{2, 3, 4\}$ (see Appendix D.4 for the exact values). We retain $p_{ap} = 2$, not on the strength of BIC alone, but on the combined weight of (i) the PACF pattern, (ii) the log-likelihood comparison across causal/noncausal splits of the α-stable model itself (which unambiguously favours the purely anticipative AR(2) among all splits of order 2, see Appendix D.4), and (iii) the portmanteau and GCoV specification checks reported below. We also note that, when the third lag is added, its coefficient is only borderline significant at the 5% level under GCoV ($\hat{\phi}_3 = 0.1122$, $t \approx 2.10$; Appendix D.4), which reinforces the choice of $p_{ap} = 2$.

Jarque-Bera and Ljung-Box. These statistics have χ^2 limiting null distributions derived under a finite-variance (Gaussian, for JB; finite-fourth-moment, for the standard Ljung-Box derivation) asymptotic regime, which does not formally cover α-stable innovations with $\alpha < 2$. We use JB only in its qualitative role, as a rejection of normality, not as a formal test of a specific alternative, consistently with our claim that the SOI is non-Gaussian rather than literally α-stable (see the Hill discussion above). For the residual diagnostics that matter for model adequacy (whiteness of the residuals and of their squares), we already report, alongside Ljung-Box, the portmanteau test of Jasiak and Manafi Neyazi (2026), which is designed for semiparametric models with i.i.d., not necessarily finite-variance, errors, and is therefore a more appropriate formal test in this setting than Ljung-Box itself.

Estimating mix-causal processes is challenging, and only a few estimation procedures are available. For the sake of robustness, we rely on two different methods: the semi-parametric Generalized Covariance Estimator (see Gourieroux and Jasiak, 2023), referred to as GCoV, and the approximate maximum likelihood method (AML), introduced by Lanne and Saikkonen (2011). Both methods face limitations due to the assumption that the error term follows a $\mathcal{S}(\alpha, \beta, \sigma, 0)$ distribution. For AML, Andrews et al. (2009) demonstrates that while asymptotic results exist, the limit distribution is largely intractable because the rate of convergence depends on the tail α parameter.[2] Gourieroux and Jasiak (2023) shows that the consistency and asymptotic properties of the semi-parametric estimator require the existence of the first four moments of the residuals, which is not the case for α-stable laws, as only the first $2\alpha + 1$ unconditional

[2] Bootstrap procedures are proposed in Andrews et al. (2009) and further extended in Cavaliere et al. (2020). However, these procedures are applicable only to purely anticipative processes, not to mixed processes.

moments exist. However, as pointed out in Remark 2 on p. 1318 of Gourieroux and Jasiak (2023), if the error term has no finite fourth-order moment, the consistency and asymptotic properties of the GCoV estimator are preserved if some nonlinear transformation of the error terms exists. In both cases, we extend the procedure to accommodate an α-stable distribution. However, a formal analysis of these extensions, along with their associated asymptotic theory, is deferred to future research. We start by reporting the GCoV results in the subsection D.2 and include the AML results in the subsection D.3. Nonetheless, the retained specification, a purely anticipative AR(2), and the coefficients estimated using AML are similar.

D.2 GCoV Estimation of Gourieroux and Jasiak (2023)

Let us begin by recalling the $MAR(p,q)$ model from Lanne and Saikkonen (2011), which imposes a multiplicative representation of the two-sided alpha-stable MA(∞) form in equation 11.[3] This representation corresponds to Corollary 2, where $\Theta = H = 1$. It is referred to as a mixed-causal autoregressive process, $MAR(p,q)$, assuming an α-stable distributed error term. This process is defined as follows:

$$\Psi(F)\Phi(B)X_t = \varepsilon_t, \qquad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha,\beta,\sigma,0), \tag{50}$$

where $\Phi(B) = 1 - \phi_1 B - \cdots - \phi_q B^q$, $\Psi(F) = 1 - \psi_1 F - \cdots - \psi_p F^p$, have their roots outside the unit circle so that

$$\Phi(z) \neq 0 \quad \text{for } |z| \leq 1 \quad \text{and} \quad \Psi(z) \neq 0 \quad \text{for } |z| \leq 1.$$

This assumption ensures stationarity in both the causal and noncausal components. A key challenge in this context is the identification of p and q. One solution, proposed by Lanne and Saikkonen (2011) and Hecq et al. (2017, 2020), is to first select $p_{ap} = p + q$ by relying on the ACF and PACF of the SOI in the training sample, and then confirm the choice using the BIC criterion.[4]

Figure 7, displays the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) for the in-sample SOI data. Based on the PACF, we set $p_{ap} = p+q$, which is equal to 2 in this case. We also apply standard information criteria, such as BIC, to validate this choice.[5] Once p_{ap} is selected, we follow the procedure outlined by Lanne and Saikkonen (2011) and (see

[3] In the approach of Gourieroux and Jasiak (2023), no such restriction is imposed. In the univariate case, both representations are equivalent. However, in the multivariate case, certain conditions must be met for the equivalence of the two representations (see Giancaterini, 2023).

[4] p_{ap} corresponds to the number of lags in the all-pass representation of the $MAR(p,q)$ model, which is known to correspond exactly to $p+q$ (see Fries and Zakoian, 2019).

[5] Table 15 confirms the selection of 2.

Hecq et al., 2020) to select the best specification for the SOI. This is done by estimating all the $2^{p_{ap}} - 1$ possible combinations for $MAR(p,q)$ using the likelihood, along with the ACF of both the residuals and the squared residuals.[6] Theoretically, only the true specification should result in the i.i.d. nature of the residuals. This is reflected in a clear ACF and PACF for both the residuals and the squared residuals. Achieving this is the primary goal of the GCoV estimator. Specifically, this estimator relies on the i.i.d. assumption for the errors as the parameters of interest and minimizes a residual-based portmanteau statistic. It is a one-step estimator that has been shown to be consistent, asymptotically normally distributed, and semi-parametrically efficient. More precisely, the estimator minimizes a portmanteau-type objective function involving the autocovariances of nonlinear transformations of model errors, viewed as functions of the model parameters. If $\theta = (\Phi, \Psi) \in \Theta$ represents the set of parameters of our MAR model from equation (50), where Θ is the entire parameter space, the GCoV estimator minimizes the following portmanteau statistic:

$$\hat{\theta} = \underset{\Theta}{\operatorname{argmin}} \sum_{h=1}^{\tilde{H}} Tr\left[\hat{\gamma}_g(h;\theta)\hat{\gamma}_g(0;\theta)^{-1}\hat{\gamma}_g(h;\theta)'\hat{\gamma}_g(0;\theta)^{-1}\right] \tag{51}$$

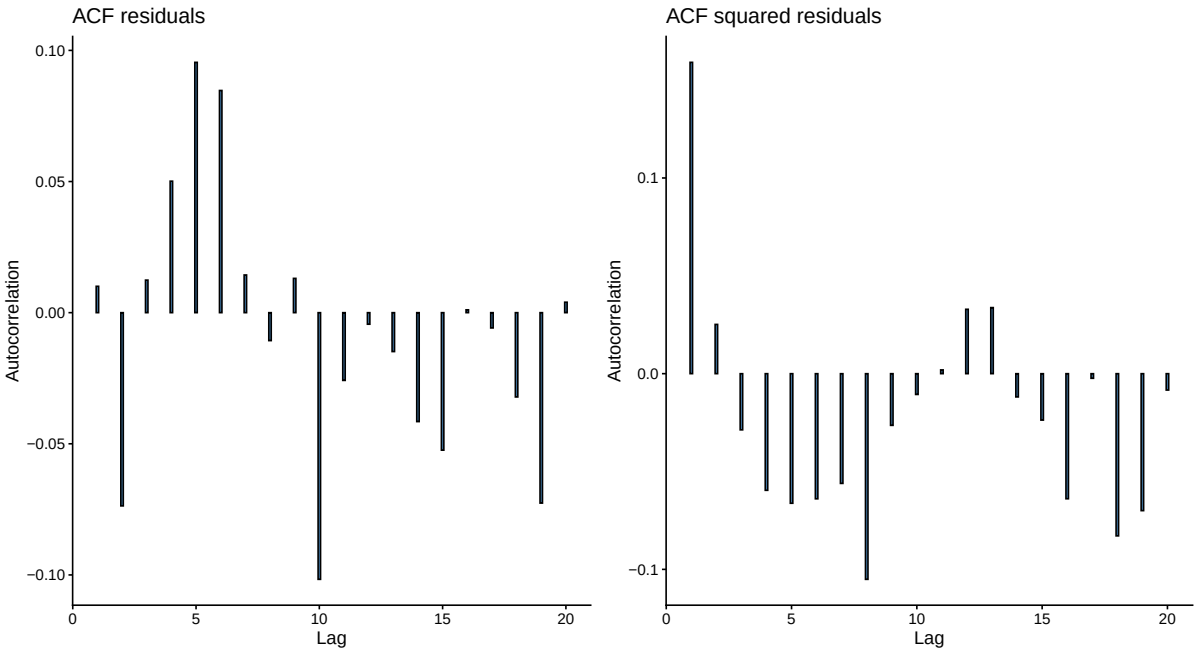
where $\tilde{H}$ is the highest selected lag, $\hat{\gamma}_g(h;\theta)$ is the sample autocovariance between $g(\varepsilon_t)$ and $g(\varepsilon_{t-h})$, with $g(\varepsilon_t) = [g_1(\varepsilon_t), \cdots, g_K(\varepsilon_t)]$, where ε_t, is the residuals obtained using equation (50).

Since only $2\alpha + 1$ moments exist for the α-stable law, it is impossible to rely on nonlinear transformations such as ε_t^k with $k > 1$ and k being an integer. This also violates the assumption of the existence of fourth-order moments, as required by the GCoV approach, to be consistent and asymptotically normally distributed. To address the issue of moment existence, as pointed out in Remark 2 on p. 1318 of Gourieroux and Jasiak (2023), if ε_t has no finite fourth-order moment, the consistency and asymptotic properties of the GCoV estimator are preserved if the transformed errors $g_k(\varepsilon_t)$ have finite fourth-order moments. We choose the following nonlinear transformation for the residuals: $g_k = \log(|\varepsilon_t|)^k$, where $k \in \{0, \ldots, K\}$. Indeed, Corollary 3.6 in Nolan (2021) ensures that g_k, $k \in \{0, \ldots, K\}$ exists even if ε_t does not have finite fourth-order moments. With this choice of g_k, we are able to compute the standard errors associated with the parameters in θ using the formula in Corollary 1 of Gourieroux and Jasiak (2023). If θ_0 represents the parameters estimated by GCoV, the standard errors can be computed as described in Gourieroux and Jasiak (2023), requiring the first-order derivative of $\gamma(h;\theta_0)$. We use finite differences to approximate this first-order derivative. Proving and assessing the performance of the GCoV approach for univariate and multivariate α-stable mixed-causal models is left for

[6] See again Table 15 , which confirms the choice of purely anticipative AR(2)

future research. Recall that the aim of the GCoV approach is to target i.i.d. residuals.[7] We test all $2^{p_{ap}} - 1$ possible combinations for MAR(r, s) with different combinations of $H = \{1, 2, 3\}$ and $K = \{1, 2, 3\}$. Subsequently, we test for the i.i.d.-ness of the residuals, focusing specifically on the autocorrelations of both the residuals and the squared residuals. The only choice of (r, s, K, H) leading to close i.i.d. residuals is the parameter set $(0, 2, 2, 2)$. Figure 9 displays the ACF of the residuals, showing no significant autocorrelation, which is confirmed by the results of the Ljung-Box test on the residuals, as reported in Table 14. Figure 9 also shows the ACF of the squared residuals. A barely significant autocorrelation at lag one is observed, leading to the rejection of the null hypothesis in the Ljung-Box test for the squared residuals (see Table 14).

Figure 9: ACF of Residuals and Squared Residual using GCoV



However, we also implement the Portmanteau Test from Jasiak and Manafi Neyazi (2026), which is a residual-based specification test for semiparametric models with i.i.d. errors. The i.i.d. nature of the residuals is confirmed by the non-rejection of the null hypothesis for this test (see Table 14). Table 14 presents the estimated parameters for ψ_1 and ψ_2, which closely resemble those obtained using the AML approach, as detailed in Table 16.

GCoV is a semi-parametric approach that does not rely on any distributional assumptions, unlike our approach, which assumes that ε_t is α-stable. First, the Jarque-Bera test in Table 14 confirms that the residuals are indeed non-Gaussian. Next, we fit an α-stable distribution

[7] The residuals are extracted from the SOI data using the estimated parameters and the equation (50)

Table 14: AR(2) Estimation for SOI, using GCov

ψ_1	ψ_2	α	β	σ
0.4224	0.2924	1.9754	-0.0216	0.4735
(0.0480)	(0.0487)	(0.0018)	(1.6763)	(0.0011)

Residuals test	Stats	$CV_{\alpha=5\%}$	p-value
LB-Test on ε_t (lag=5)	8.5836	7.8147	0.0354
LB-Test on ε_t^2 (lag=5)	17.142	7.8147	0.0007
Jasiak and Manafi Neyazi (2026)'s test	5.5933	12.5915	0.4702
JB-Test	21.261	5.9900	0.0000

Notes: Estimated parameters of the α-stable anticipative AR(2) process associated with the SOI series for the period 01/1951–12/1991. Standard deviations are provided in parentheses. The specification tests are as follows: the Ljung-Box test on residuals, the Ljung-Box test on the squared residuals, the Portmanteau test of Jasiak and Manafi Neyazi (2026), and a non-Gaussianity test.

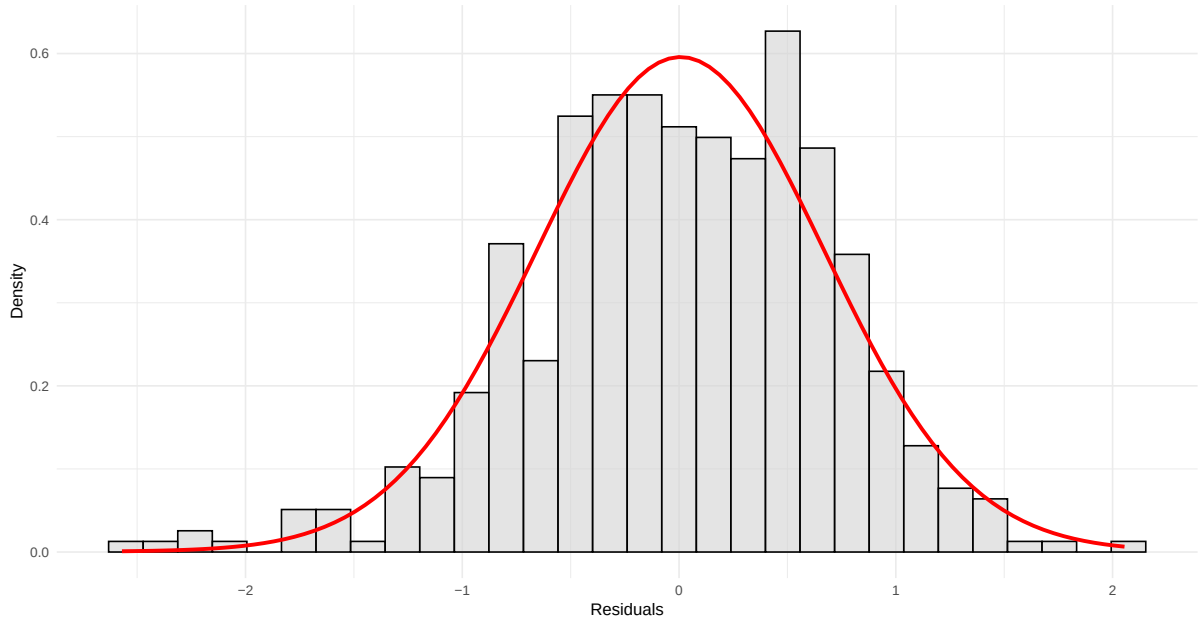
$\mathcal{S}(\alpha, \beta, \sigma, 0)$ to the residuals using the characteristic function-based estimation for α-stable distributions described in Nolan (2021). The estimated parameters are reported in Table 14. These results are consistent with those obtained using the AML approach (see Table 16). Figure 10, confirms the good fit of the estimated α-stable distribution. However, we encounter an identification issue with the β parameters, though this is not significant. This is unsurprising, as β is inherently difficult to identify. This challenge arises from the fact that it becomes nearly impossible to reliably estimate β when α is close to 2. As α approaches 2, the distribution increasingly resembles a Gaussian distribution, which is symmetric. In this limit, the skewness parameter β has a negligible effect on the shape of the distribution, making it practically unidentifiable due to the dominance of the symmetric properties inherent in the Gaussian limit of α-stable laws.

D.3 AML estimation of Lanne and Saikkonen (2011)

This section, presents the results on in-sample SOI data using the AML estimation procedure from, Lanne and Saikkonen (2011) and Hecq et al. (2020, 2017). As is indicated by the Figure 7, we set $p_{ap} = 2$, and we estimate all the possible $2^{p_{ap}}) - 1$ combinaison of $MAR(p, q)$. More precisely, the parameters of all the $MAR(p, q)$ are subsequently estimated using a modified version of the MARX package suitable for α-stable laws (Hecq et al., 2020, 2017). For the $MAR(p, q)$ as in eq (50), the ensemble of parameters to be estimated is $\theta = (\Psi, \Phi, \alpha, \beta, \sigma) \in \Theta$ then the Lanne and Saikkonen (2011) AML estimator, is defined as:

$$\hat{\theta} = \underset{\Theta}{\operatorname{argmax}} \sum_{t=p+1}^{T-q} \ln f^{-1} \left[\Psi(F)\Phi(B)X_t; (\alpha, \beta, \sigma, 0)\right] \tag{52}$$

Figure 10: Histogram of residuals using GCoV



Notes: The figure illustrates the histogram of residuals with an overlaid density curve. The red line represents the density of an α-stable law with the parameters: $\alpha = 1.9754$, $\beta = -0.0216$, $\sigma = 0.4735$.

where $f(.;(\alpha,\beta,\sigma,0))$ denotes the probability function of the ε_t.[8] Standard deviations are estimated using finite differences of the gradient and Hessian for the parameters in the right space. As shown by Andrews et al. (2009), this normal approximation is exact for the parameters of the α-stable innovation distribution, but only indicative for the autoregressive coefficients, whose ML estimators have a non-Gaussian, analytically intractable limiting distribution; the corresponding standard errors should therefore be interpreted with caution. Table 15, according to the likelihood criteria, shows that the best specification for causal-non-causal models is the anticipative AR(2)

$$(1 - \psi_1 F - \psi_2 F^2)X_t = \varepsilon_t,\ \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha,\beta,\sigma,0)$$

Table 16 reports the estimation results for the anticipative AR(2) parameters and the associated parameters of the α-stable distribution for the error term. Additionally, we present a set of descriptive statistics and validation tests on the residuals and the squared residuals. From Table 16, the Ljung-Box (LB) tests indicate that all autocorrelation in the residuals of the AR(2) model has been removed. This conclusion is further supported by the ACFs of the

[8] Lanne and Saikkonen (2011) shows that the AML approach is consistent if f admits a Lebesgue representation, which is the case for the α-stable law.

Table 15: Identification of Noncausal Processes for SOI

BIC	MAR(2,0)	MAR(1,1)	MAR(0,2)
p_{pseudo}		Likelihood	
2	-513.7138	-511.3617	-508.8777

residuals, shown in Figure 11. However, Figure 11 reveals a barely significant autocorrelation at lag one in the squared residuals, which is confirmed by the rejection of the null hypothesis for the Ljung-Box test on the squared residuals, as reported in Table 16.

The Jarque-Bera (JB) test in Table 16 indicates that the residuals are indeed non-Gaussian ($\alpha < 2$), this is consistent with an estimated $\alpha = 1.93$ $(1.90E-4)$. However, we have an estimated β of -0.99 $(2.68E-2)$ which is barely significant. This is not surprising as β is hard to identify, explained by the fact that it is impossible to reliably identify the coefficient β when α is too close to 2. This is because, as α gets close to 2, the distribution increasingly resembles a Gaussian distribution, which is symmetric. In this limit, the skewness parameter β has a diminishing effect on the shape of the distribution, rendering it practically unidentifiable due to the dominance of the symmetric properties inherent in the Gaussian limit of α-stable laws. The goodness of fit of the estimated parameters for the α-stable law is confirmed by the associated estimated density and the histogram of the residuals shown in Figure 12.

Table 16: AR(2) AML estimation for SOI

ψ_1	ψ_2	α	β	σ
0.4476	0.2750	1.9390	-0.9970	0.4696
(0.0000)	(0.0000)	(0.0001)	(0.0268)	(0.0000)

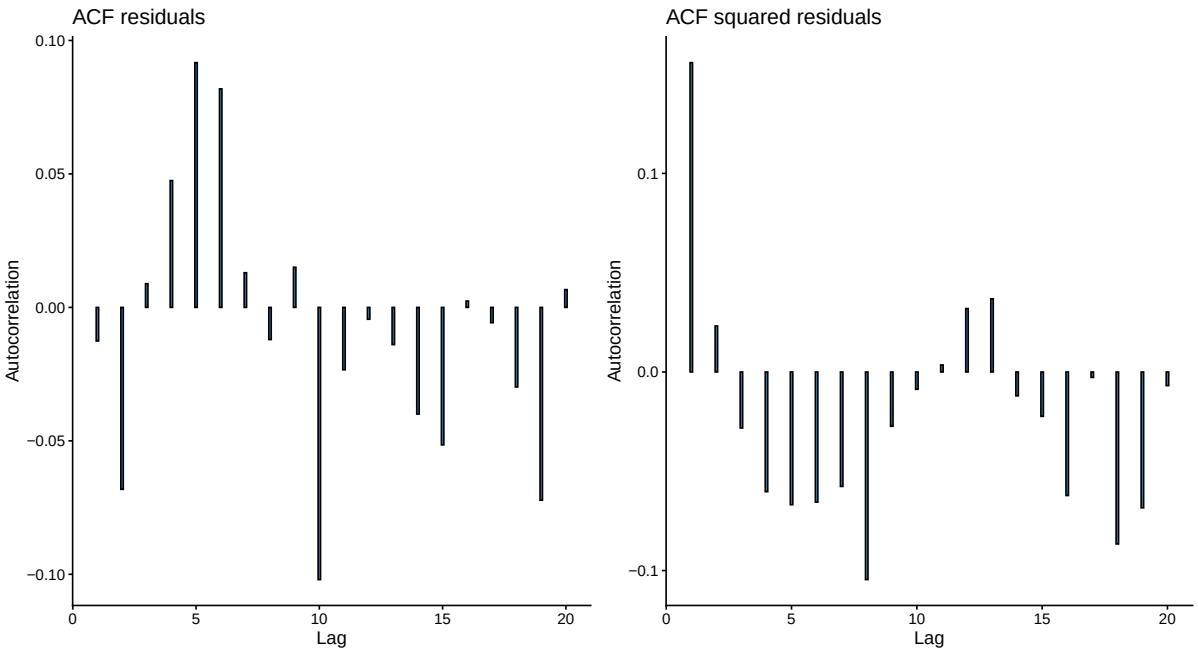
Residuals test	Stats	$CV_{\alpha=5\%}$	p-value
LB-Test on ε_t (lag=5)	7.6878	7.8147	0.0529
LB-Test on ε_t^2 (lag=5)	16.641	7.8147	0.0008
JB-Test	20.6780	5.9900	0.0000

Notes: Estimated parameters of α-stable anticipative AR(2) process associated with the SOI series for the period 01/1951 - 12/1991. Standard deviations are in parentheses.

In Figure 11, the left panel shows the ACF of the residuals computed using the AML approach, confirming their lack of serial correlation, while the right panel presents the ACF of the squared residuals, indicating only a barely significant autocorrelation at lag one. This confirms the good accuracy of the anticipative AR(2) model.

Figure 12, helps to visually assess the fit of the α-stable model to the residuals. The histogram and the density curve align well, it suggests that the model is appropriately capturing

Figure 11: ACF of Residuals and Squared Residual using AML



the underlying distribution of the residuals.

D.4 Order selection under infinite variance: AIC vs. BIC, and the order-3 robustness check

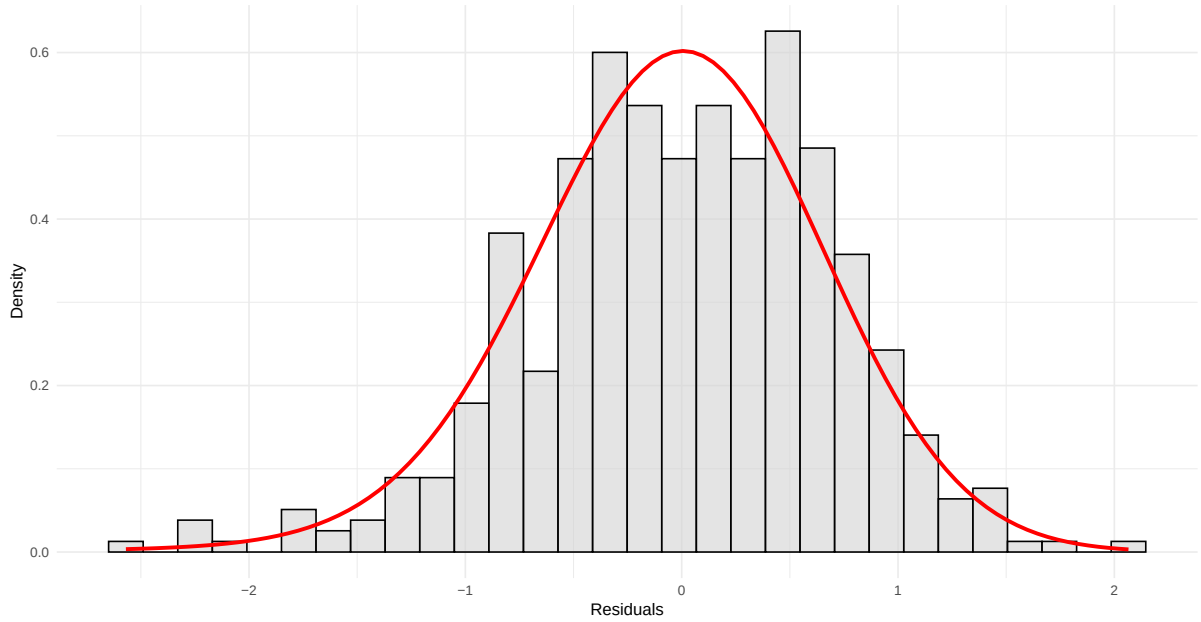
Table 17 reports the exact AIC, BIC and HQ values for $p_{ap} \in \{0, 1, 2, 3, 4\}$ referenced above, computed on the raw (OLS-filtered) series as a preliminary order-selection step.

Table 17: AIC, BIC and HQ for the SOI, $p_{ap} \in \{0, \ldots, 4\}$.

	$p_{ap} = 0$	$p_{ap} = 1$	$p_{ap} = 2$	$p_{ap} = 3$	$p_{ap} = 4$
BIC	2.6823	2.1956	2.1349	2.1386	2.1475
AIC	2.6738	2.1785	2.1092	**2.1043**	2.1046
HQ	2.6771	2.1852	2.1193	**2.1178**	2.1214

BIC is minimized at $p_{ap} = 2$; AIC and HQ are both minimized at $p_{ap} = 3$, though the AIC gap between $p_{ap} = 2$ and $p_{ap} = 3$ (2.1092 vs. 2.1043) is small relative to the gap between $p_{ap} = 1$ and $p_{ap} = 2$ (2.1785 vs. 2.1092). Using the α-stable log-likelihood (not an OLS/Gaussian proxy) to compare the three possible splits of $p_{ap} = 2$ into causal and noncausal orders, we obtain log-likelihoods of -513.731 (MAR(2,0), purely causal), -511.190 (MAR(1,1))

Figure 12: Histogram of residuals using AML



Notes: The figure illustrates the histogram of residuals with an overlaid density curve. The red line represents the density of an α-stable law with the parameters: $\alpha = 1.9390$, $\beta = -0.9970$, $\sigma = 0.4696$.

and -508.578 (MAR(0,2), purely anticipative). The purely anticipative specification retained is the unambiguous maximizer among the three.

Robustness: the order-3 alternative. Since AIC and HQ point to $p_{ap} = 3$, we re-run the same causal/noncausal split selection with $p_{ap} = 3$. The α-stable log-likelihood again unambiguously favours the purely anticipative specification, MAR(0,3), with log-likelihoods -510.507, -510.624, -506.636 and -504.801 for MAR(3,0), MAR(2,1), MAR(1,2) and MAR(0,3) respectively. Table 18 reports the fitted MAR(0,3) coefficients under GCoV. The AML estimator of Appendix D.3 gives a similar fit ($\hat{\phi}_1 = 0.4127$, $\hat{\phi}_2 = 0.2238$, $\hat{\phi}_3 = 0.1125$, $\hat{\alpha} = 1.930$, $\hat{\beta} = -0.990$, $\hat{\sigma} = 0.464$): both estimators agree closely on the third-lag coefficient in particular (0.1122 GCoV vs. 0.1125 AML), and on the overall pattern of rapidly decaying coefficients, though they disagree more on the individual split between ϕ_1 and ϕ_2 and on (α, β), consistently with the GCoV/AML discrepancy already documented for the retained AR(2) in Appendix D.2.

Under GCoV, the third lag has a t-statistic of 2.10 ($\hat{\phi}_3$/s.e. $= 0.1122/0.0536$), i.e. borderline significant at the 5% level, an entirely independent (semiparametric, distribution-free) confirmation of the "weak, not decisive" reading of the AIC/BIC gap. Comparing the retained AR(2) against the AML MAR(0,3) alternative using the actual (likelihood-based (LL), not OLS-proxy)

Table 18: Fitted purely anticipative MAR(0,3) for the SOI, GCoV estimator.

	ϕ_1	ϕ_2	ϕ_3	α	β	σ
Estimate	0.3415	0.2873	0.1122	1.9668	-0.0257	0.4693
s.e.	(0.0473)	(0.0489)	(0.0536)	–	–	–

AIC and BIC, AIC $= 2k + 2 \times$ LL and BIC $= k\log(n) + 2 \times$ LL with $k = 5$ (AR(2)) or $k = 6$ (MAR(0,3)) and $n = 492$, gives $\text{AIC}_{AR(2)} = 1027.16$ vs. $\text{AIC}_{MAR(0,3)} = 1021.60$ (difference 5.55 in favour of the order-3 model), and $\text{BIC}_{AR(2)} = 1048.15$ vs. $\text{BIC}_{MAR(0,3)} = 1046.79$ (difference 1.35, again in favour of order 3 but far more modest once the extra parameter is penalised). Taken together with the GCoV evidence above, the case for a third lag is real but weak, and does not overturn the retained specification.

We nonetheless retain the AR(2) specification for three reasons. First, the third-lag coefficient ($\hat{\phi}_3 \approx 0.112$) is markedly smaller than $\hat{\phi}_1$ and $\hat{\phi}_2$ under both estimators, consistent with a rapidly decaying dependence structure that AR(2) already captures the bulk of; adding a third lag refines the fit rather than revealing a qualitatively different dynamic. Second, the BIC gap of 1.35 is modest on the usual Kass–Raftery scale for evidence differences, i.e. "weak" evidence at best for the extra lag, and is dwarfed by the AIC/BIC gap separating $p_{ap} = 1$ from $p_{ap} = 2$ above.

D.5 Sensitivity of the out-of-sample reversal-date forecasts to the window length m

This section complements Table 1, which reports the reversal-date forecasts obtained with the baseline $m = 10$ for the 11 out-of-sample El Niño/La Niña episodes. We recompute the same forecasting exercise, using the fitted GCoV anticipative AR(2) for the SOI (Table 14, $\hat{\psi}_1 = 0.4224$, $\hat{\psi}_2 = 0.2924$), same $\bar{k} = 30$, under three additional benchmarks: the theoretical minimum $m = 1$, a medium benchmark $m = 5$, and a data-driven $\hat{m}$, chosen for each episode in $\{2, \ldots, 10\}$ as the value minimising the pattern-match distance $e_m(\hat{k}_0)$ normalised by $(m + 1)$ (to make the criterion comparable across window lengths). Table 19 reports, for each episode, the forecasted peak date and the forecast error in months relative to the true peak date, under each benchmark.

The baseline $m = 10$ achieves the lowest mean absolute error (0.64 months), followed by $m = 5$ and the data-driven $\hat{m}$ (0.91 each), and $m = 1$ (1.55). This is consistent with the baseline choice, but it should be read together with Appendix C.2: at $m = 10$, ties are declared

Table 19: Sensitivity of the out-of-sample reversal (peak) date forecast to the window length m, GCoV estimator, for the 11 El Niño/La Niña episodes of Table 1 (main text). Forecast error is measured in months relative to the true peak date. $\hat{m}$ is the data-driven choice in $\{2, \ldots, 10\}$ minimising the normalised pattern-match distance $e_m(\hat{k}_0)/(m+1)$.

Type	Start	True peak	$m = 1$		$m = 5$		$m = 10$		$\hat{m}$	
			Peak	Err.	Peak	Err.	Peak	Err.	Peak ($\hat{m}$)	Err.
El Nino	12/1991	01/1992	01/1992	0	01/1992	0	01/1992	0	01/1992 ($\hat{m} = 2$)	0
El Nino	07/1994	09/1994	08/1994	-1	11/1994	2	09/1994	0	08/1994 ($\hat{m} = 2$)	-1
La Nina	11/2007	02/2008	06/2008	4	04/2008	2	02/2008	0	04/2008 ($\hat{m} = 3$)	2
El Nino	12/2009	02/2010	04/2010	2	03/2010	1	01/2010	-1	01/2010 ($\hat{m} = 7$)	-1
La Nina	07/2010	09/2010	08/2010	-1	08/2010	-1	08/2010	-1	08/2010 ($\hat{m} = 2$)	-1
La Nina	11/2010	12/2010	01/2011	1	01/2011	1	01/2011	1	01/2011 ($\hat{m} = 7$)	1
El Nino	07/2015	10/2015	08/2015	-2	10/2015	0	09/2015	-1	10/2015 ($\hat{m} = 2$)	0
La Nina	11/2021	01/2022	02/2022	1	02/2022	1	12/2021	-1	03/2022 ($\hat{m} = 3$)	2
La Nina	02/2022	03/2022	03/2022	0	04/2022	1	04/2022	1	04/2022 ($\hat{m} = 10$)	1
La Nina	08/2022	10/2022	02/2023	4	10/2022	0	10/2022	0	10/2022 ($\hat{m} = 10$)	0
La Nina	11/2022	12/2022	01/2023	1	01/2023	1	01/2023	1	01/2023 ($\hat{m} = 4$)	1
Mean absolute error (months)			1.55		0.91		0.64		0.91	

in the large majority of replications of the generic MAR(0,2) simulation for α close to the SOI's estimated tail index. The good performance of $m = 10$ on the real SOI episodes therefore reflects the combined effect of the pattern-matching step and the tie-breaking rule of Section 5.4 , not of an unambiguous, tie-free identification of k_0.

Repeating the same exercise with the AML-fitted anticipative AR(2) (Table 16, $\hat{\psi}_1 = 0.4476$, $\hat{\psi}_2 = 0.2750$) gives near-identical results, reported in Table 20: the mean absolute error is again minimised at $m = 10$ (0.64), followed by $\hat{m}$ (0.73), $m = 5$ (0.91) and $m = 1$ (1.36). Only one episode (La Niña, 11/2007) differs materially between the two estimators, in the $m = 1$ and $\hat{m}$ columns; every other forecast and error coincides. The robustness of the reversal-date forecasts to the window length m therefore also holds across the two estimation methods.

Table 20: Sensitivity of the out-of-sample reversal (peak) date forecast to the window length m, AML estimator, for the 11 El Niño/La Niña episodes of Table 1 (main text). Forecast error is measured in months relative to the true peak date. $\hat{m}$ is the data-driven choice in $\{2, \ldots, 10\}$ minimising the normalised pattern-match distance $e_m(\hat{k}_0)/(m+1)$.

Type	Start	True peak	$m = 1$		$m = 5$		$m = 10$		$\hat{m}$	
			Peak	Err.	Peak	Err.	Peak	Err.	Peak ($\hat{m}$)	Err.
El Nino	12/1991	01/1992	01/1992	0	01/1992	0	01/1992	0	01/1992 ($\hat{m} = 2$)	0
El Nino	07/1994	09/1994	08/1994	-1	11/1994	2	09/1994	0	08/1994 ($\hat{m} = 2$)	-1
La Nina	11/2007	02/2008	04/2008	2	02/2008	0	02/2008	0	02/2008 ($\hat{m} = 3$)	0
El Nino	12/2009	02/2010	04/2010	2	01/2010	-1	01/2010	-1	01/2010 ($\hat{m} = 7$)	-1
La Nina	07/2010	09/2010	08/2010	-1	08/2010	-1	08/2010	-1	08/2010 ($\hat{m} = 2$)	-1
La Nina	11/2010	12/2010	01/2011	1	01/2011	1	01/2011	1	01/2011 ($\hat{m} = 7$)	1
El Nino	07/2015	10/2015	08/2015	-2	08/2015	-2	09/2015	-1	10/2015 ($\hat{m} = 2$)	0
La Nina	11/2021	01/2022	02/2022	1	02/2022	1	12/2021	-1	03/2022 ($\hat{m} = 3$)	2
La Nina	02/2022	03/2022	03/2022	0	04/2022	1	04/2022	1	04/2022 ($\hat{m} = 10$)	1
La Nina	08/2022	10/2022	02/2023	4	10/2022	0	10/2022	0	10/2022 ($\hat{m} = 10$)	0
La Nina	11/2022	12/2022	01/2023	1	01/2023	1	01/2023	1	01/2023 ($\hat{m} = 4$)	1
Mean absolute error (months)			1.36		0.91		0.64		0.73	

E Proofs

E.1 Proof of Proposition 1

Consider first the case where either $\alpha \neq 1$ or $\boldsymbol{X}$ is S1S. We only provide the proof for $\alpha \neq 1$ as it is similar under both assumptions.

Assume that $\Gamma(K^{\|\cdot\|}) = 0$ and let us show that $\boldsymbol{X}$ admits a representation of the unit cylinder $C_d^{\|\cdot\|}$ relative to the seminorm $\|\cdot\|$. The characteristic function of $\boldsymbol{X}$ writes for any $\boldsymbol{u} \in \mathbb{R}^d$, with $a = \mathrm{tg}(\pi\alpha/2)$,

$$
\begin{aligned}
\varphi_{\boldsymbol{X}}(\boldsymbol{u}) &= \exp\left\{ -\int_{S_d} \Big(|\langle \boldsymbol{u}, \boldsymbol{s}\rangle|^\alpha - ia(\langle \boldsymbol{u}, \boldsymbol{s}\rangle)^{<\alpha>}\Big)\Gamma(d\boldsymbol{s}) + i\,\langle \boldsymbol{u}, \boldsymbol{\mu}^0\rangle \right\} \\
&= \exp\left\{ -\int_{S_d\setminus K^{\|\cdot\|}} \Big(|\langle \boldsymbol{u}, \boldsymbol{s}\rangle|^\alpha - ia(\langle \boldsymbol{u}, \boldsymbol{s}\rangle)^{<\alpha>}\Big)\Gamma(d\boldsymbol{s}) + i\,\langle \boldsymbol{u}, \boldsymbol{\mu}^0\rangle \right\} \\
&= \exp\left\{ -\int_{S_d\setminus K^{\|\cdot\|}} \Big(|\langle \boldsymbol{u}, \frac{\boldsymbol{s}}{\|\boldsymbol{s}\|}\rangle|^\alpha - ia(\langle \boldsymbol{u}, \frac{\boldsymbol{s}}{\|\boldsymbol{s}\|}\rangle)^{<\alpha>}\Big)\|\boldsymbol{s}\|^\alpha\Gamma(d\boldsymbol{s}) + i\,\langle \boldsymbol{u}, \boldsymbol{\mu}^0\rangle \right\} \\
&= \exp\left\{ -\int_{T_{\|\cdot\|}(S_d\setminus K^{\|\cdot\|})} \Big(|\langle \boldsymbol{u}, \boldsymbol{s}'\rangle|^\alpha - ia(\langle \boldsymbol{u}, \boldsymbol{s}'\rangle)^{<\alpha>}\Big)\left\|\frac{\boldsymbol{s}'}{\|\boldsymbol{s}'\|_e}\right\|^\alpha \Gamma\circ T_{\|\cdot\|}^{-1}(d\boldsymbol{s}') + i\,\langle \boldsymbol{u}, \boldsymbol{\mu}^0\rangle \right\} \\
&= \exp\left\{ -\int_{C_d^{\|\cdot\|}} \Big(|\langle \boldsymbol{u}, \boldsymbol{s}\rangle|^\alpha - ia(\langle \boldsymbol{u}, \boldsymbol{s}\rangle)^{<\alpha>}\Big) \underbrace{\|\boldsymbol{s}\|_e^{-\alpha}\Gamma\circ T_{\|\cdot\|}^{-1}(d\boldsymbol{s})}_{\Gamma^{\|\cdot\|}(d\boldsymbol{s})} + i\,\langle \boldsymbol{u}, \boldsymbol{\mu}^0\rangle \right\}
\end{aligned}
$$

where we used the change of variable $\boldsymbol{s}' = T_{\|\cdot\|}(\boldsymbol{s}) = \boldsymbol{s}/\|\boldsymbol{s}\|$ between the third and fourth lines, which yields the representation on $C_d^{\|\cdot\|}$.

Reciprocally, assume that $\boldsymbol{X}$ is representable on $C_d^{\|\cdot\|}$. By definition of the representability of $\boldsymbol{X}$ on $C_d^{\|\cdot\|}$, there exists a measure $\gamma^{\|\cdot\|}$ on $C_d^{\|\cdot\|}$ and a non-random vector $\boldsymbol{m}^0_{\|\cdot\|} \in \mathbb{R}^d$ such that

$$
\varphi_{\boldsymbol{X}}(\boldsymbol{u}) = \exp\left\{ -\int_{C_d^{\|\cdot\|}} \Big(|\langle \boldsymbol{u}, \boldsymbol{s}\rangle|^\alpha - ia(\langle \boldsymbol{u}, \boldsymbol{s}\rangle)^{<\alpha>}\Big)\gamma^{\|\cdot\|}(d\boldsymbol{s}) + i\,\langle \boldsymbol{u}, \boldsymbol{m}^0_{\|\cdot\|}\rangle \right\}.
$$

With the change of variable $\boldsymbol{s}' = T^{-1}_{\|\cdot\|}(\boldsymbol{s}) = \boldsymbol{s}/\|\boldsymbol{s}\|_e$,

$$
\begin{aligned}
\varphi_{\boldsymbol{X}}(\boldsymbol{u}) &= \exp\left\{ -\int_{C_d^{\|\cdot\|}} \Big(|\langle \boldsymbol{u}, \frac{\boldsymbol{s}}{\|\boldsymbol{s}\|_e}\rangle|^\alpha - ia(\langle \boldsymbol{u}, \frac{\boldsymbol{s}}{\|\boldsymbol{s}\|_e}\rangle)^{<\alpha>}\Big)\|\boldsymbol{s}\|_e^\alpha\gamma^{\|\cdot\|}(d\boldsymbol{s}) + i\,\langle \boldsymbol{u}, \boldsymbol{m}^0_{\|\cdot\|}\rangle \right\} \\
&= \exp\left\{ -\int_{T^{-1}_{\|\cdot\|}(C_d^{\|\cdot\|})} \Big(|\langle \boldsymbol{u}, \boldsymbol{s}'\rangle|^\alpha - ia(\langle \boldsymbol{u}, \boldsymbol{s}'\rangle)^{<\alpha>}\Big)\left\|\frac{\boldsymbol{s}'}{\|\boldsymbol{s}'\|}\right\|_e^\alpha \gamma^{\|\cdot\|}\circ T_{\|\cdot\|}(d\boldsymbol{s}') + i\,\langle \boldsymbol{u}, \boldsymbol{m}^0_{\|\cdot\|}\rangle \right\} \\
&= \exp\left\{ -\int_{S_d\setminus K^{\|\cdot\|}} \Big(|\langle \boldsymbol{u}, \boldsymbol{s}\rangle|^\alpha - ia(\langle \boldsymbol{u}, \boldsymbol{s}\rangle)^{<\alpha>}\Big)\|\boldsymbol{s}\|^{-\alpha}\gamma^{\|\cdot\|}\circ T_{\|\cdot\|}(d\boldsymbol{s}) + i\,\langle \boldsymbol{u}, \boldsymbol{m}^0_{\|\cdot\|}\rangle \right\} \\
&= \exp\left\{ -\int_{S_d\setminus K^{\|\cdot\|}} \Big(|\langle \boldsymbol{u}, \boldsymbol{s}\rangle|^\alpha - ia(\langle \boldsymbol{u}, \boldsymbol{s}\rangle)^{<\alpha>}\Big)\gamma(d\boldsymbol{s}) + i\,\langle \boldsymbol{u}, \boldsymbol{m}^0_{\|\cdot\|}\rangle \right\},
\end{aligned}
$$

where $\gamma(d\boldsymbol{s}) := \|\boldsymbol{s}\|^{-\alpha}\gamma^{\|\cdot\|} \circ T_{\|\cdot\|}(d\boldsymbol{s})$. Letting now $\overline{\gamma}(A) := \gamma(A \cap (S_d \setminus K^{\|\cdot\|}))$ for any Borel set A of S_d, we have

$$\varphi_{\boldsymbol{X}}(\boldsymbol{u}) = \exp\left\{ -\int_{S_d} \Big(|\langle \boldsymbol{u}, \boldsymbol{s}\rangle|^\alpha - ia(\langle \boldsymbol{u}, \boldsymbol{s}\rangle)^{<\alpha>}\Big)\overline{\gamma}(d\boldsymbol{s}) + i\,\langle \boldsymbol{u}, \boldsymbol{m}^0_{\|\cdot\|}\rangle \right\}.$$

By the unicity of the spectral representation of $\boldsymbol{X}$ on S_d, we necessarily have $(\Gamma, \boldsymbol{\mu}^0) = (\overline{\gamma}, \boldsymbol{m}^0_{\|\cdot\|})$. Thus, $\overline{\gamma}$ and Γ have to coincide, and in particular

$$\Gamma(K^{\|\cdot\|}) = \overline{\gamma}(K^{\|\cdot\|}) = \gamma(K^{\|\cdot\|} \cap (S_d \setminus K^{\|\cdot\|})) = \gamma(\emptyset) = 0.$$

Given that $\Gamma = \overline{\gamma}$ and $\Gamma(K^{\|\cdot\|}) = 0$, we can follow the initial steps of the proof to show that $\gamma^{\|\cdot\|} = \Gamma^{\|\cdot\|}$.

Consider now the case where $\alpha = 1$ and $\boldsymbol{X}$ is not symmetric. Assume first that $\int_{S_d} \Big|\ln\|\boldsymbol{s}\|\Big|\Gamma(d\boldsymbol{s}) < +\infty$, that is, $\Gamma(K^{\|\cdot\|}) = 0$ and $\int_{S_d\setminus K^{\|\cdot\|}} \Big|\ln\|\boldsymbol{s}\|\Big|\Gamma(d\boldsymbol{s}) < +\infty$. With $a = 2/\pi$,

$$\begin{aligned}\varphi_{\boldsymbol{X}}(\boldsymbol{u}) &= \exp\left\{ -\int_{S_d} \Big(|\langle \boldsymbol{u}, \boldsymbol{s}\rangle| + ia\langle \boldsymbol{u}, \boldsymbol{s}\rangle \ln|\langle \boldsymbol{u}, \boldsymbol{s}\rangle|\Big)\Gamma(d\boldsymbol{s}) + i\,\langle \boldsymbol{u}, \boldsymbol{\mu}^0\rangle \right\} \\ &= \exp\left\{ -\int_{S_d\setminus K^{\|\cdot\|}} \left(|\langle \boldsymbol{u}, \frac{\boldsymbol{s}}{\|\boldsymbol{s}\|}\rangle| + ia\langle \boldsymbol{u}, \frac{\boldsymbol{s}}{\|\boldsymbol{s}\|}\rangle \ln|\langle \boldsymbol{u}, \frac{\boldsymbol{s}}{\|\boldsymbol{s}\|}\rangle|\right)\|\boldsymbol{s}\|\Gamma(d\boldsymbol{s}) \right. \\ &\qquad \left. + i\,\langle \boldsymbol{u}, \boldsymbol{\mu}^0\rangle - ia\int_{S_d\setminus K^{\|\cdot\|}} \langle \boldsymbol{u}, \boldsymbol{s}\rangle \ln\|\boldsymbol{s}\|\Gamma(d\boldsymbol{s}) \right\}.\end{aligned}$$

We have $\int_{S_d\setminus K^{\|\cdot\|}} \langle \boldsymbol{u}, \boldsymbol{s}\rangle \ln\|\boldsymbol{s}\|\Gamma(d\boldsymbol{s}) = \sum_{i=1}^d u_i \int_{S_d\setminus K^{\|\cdot\|}} s_i \ln\|\boldsymbol{s}\|\Gamma(d\boldsymbol{s}) = \langle \boldsymbol{u}, \tilde{\boldsymbol{\mu}}\rangle$, and thus,

$$i\,\langle \boldsymbol{u}, \boldsymbol{\mu}^0\rangle - ia\int_{S_d\setminus K^{\|\cdot\|}} \langle \boldsymbol{u}, \boldsymbol{s}\rangle \ln\|\boldsymbol{s}\|\Gamma(d\boldsymbol{s}) = i\langle \boldsymbol{u}, \boldsymbol{\mu}^0_{\|\cdot\|}\rangle.$$

The condition $\int_{S_d\setminus K^{\|\cdot\|}} \Big|\ln\|\boldsymbol{s}\|\Big|\Gamma(d\boldsymbol{s}) < +\infty$, ensures that $|\boldsymbol{\mu}^0_{\|\cdot\|}| < +\infty$. Again with the change of variable $\boldsymbol{s}' = T_{\|\cdot\|}(\boldsymbol{s}) = \boldsymbol{s}/\|\boldsymbol{s}\|$, we get

$$\begin{aligned}\varphi_{\boldsymbol{X}}(\boldsymbol{u}) &= \exp\left\{ -\int_{T_{\|\cdot\|}(S_d\setminus K^{\|\cdot\|})} \Big(|\langle \boldsymbol{u}, \boldsymbol{s}'\rangle| + ia\langle \boldsymbol{u}, \boldsymbol{s}'\rangle \ln|\langle \boldsymbol{u}, \boldsymbol{s}'\rangle|\Big) \left\|\frac{\boldsymbol{s}'}{\|\boldsymbol{s}'\|_e}\right\|^\alpha \Gamma \circ T^{-1}_{\|\cdot\|}(d\boldsymbol{s}') + i\langle \boldsymbol{u}, \boldsymbol{\mu}^0_{\|\cdot\|}\rangle \right\} \\ &= \exp\left\{ -\int_{C_d^{\|\cdot\|}} \Big(|\langle \boldsymbol{u}, \boldsymbol{s}\rangle| + ia\langle \boldsymbol{u}, \boldsymbol{s}\rangle \ln|\langle \boldsymbol{u}, \boldsymbol{s}\rangle|\Big) \underbrace{\|\boldsymbol{s}\|_e^{-\alpha}\Gamma \circ T^{-1}_{\|\cdot\|}(d\boldsymbol{s})}_{\Gamma^{\|\cdot\|}(d\boldsymbol{s})} + i\langle \boldsymbol{u}, \boldsymbol{\mu}^0_{\|\cdot\|}\rangle \right\}\end{aligned}$$

Reciprocally, assume there exists a measure $\gamma^{\|\cdot\|}$ on $C_d^{\|\cdot\|}$ satisfying (8) and a non-random vector $\boldsymbol{m}^0_{\|\cdot\|} \in \mathbb{R}^d$ such that

$$\varphi_{\boldsymbol{X}}(\boldsymbol{u}) = \exp\left\{ -\int_{C_d^{\|\cdot\|}} \Big(|\langle \boldsymbol{u}, \boldsymbol{s}\rangle| + ia\langle \boldsymbol{u}, \boldsymbol{s}\rangle \ln|\langle \boldsymbol{u}, \boldsymbol{s}\rangle|\Big)\gamma^{\|\cdot\|}(d\boldsymbol{s}) + i\,\langle \boldsymbol{u}, \boldsymbol{m}^0_{\|\cdot\|}\rangle \right\}.$$

First, we can see that

$$\varphi_{\boldsymbol{X}}(\boldsymbol{u}) = \exp\left\{ -\int_{C_d^{\|\cdot\|}} \left[\Big(|\langle \boldsymbol{u}, \frac{\boldsymbol{s}}{\|\boldsymbol{s}\|_e}\rangle| + ia\langle \boldsymbol{u}, \frac{\boldsymbol{s}}{\|\boldsymbol{s}\|_e}\rangle \ln|\langle \boldsymbol{u}, \frac{\boldsymbol{s}}{\|\boldsymbol{s}\|_e}\rangle|\Big)\|\boldsymbol{s}\|_e + ia\langle \boldsymbol{u}, \boldsymbol{s}\rangle \ln\|\boldsymbol{s}\|_e\right]\gamma^{\|\cdot\|}(d\boldsymbol{s}) \right. \\ \left. + i\,\langle \boldsymbol{u}, \boldsymbol{m}^0_{\|\cdot\|}\rangle \right\}.$$

We will later show the following result:

Lemma 2 *Let $\gamma^{\|\cdot\|}$ a Borel measure on $C_d^{\|\cdot\|}$ satisfying (8). Then,*

$$\int_{C_d^{\|\cdot\|}} \|\boldsymbol{s}\|_e \Big| \ln \|\boldsymbol{s}\|_e \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty. \tag{53}$$

Assuming Lemma 2 holds, then by the Cauchy-Schwarz inequality, we have

$$\int_{C_d^{\|\cdot\|}} |\langle \boldsymbol{u}, \boldsymbol{s}\rangle| \Big| \ln \|\boldsymbol{s}\|_e \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty$$

, and thus

$$\begin{aligned}
\varphi_{\boldsymbol{X}}(\boldsymbol{u}) = \exp\Bigg\{ & - \int_{C_d^{\|\cdot\|}} \Big(|\langle \boldsymbol{u}, \frac{\boldsymbol{s}}{\|\boldsymbol{s}\|_e}\rangle| + ia\langle \boldsymbol{u}, \frac{\boldsymbol{s}}{\|\boldsymbol{s}\|_e}\rangle \ln |\langle \boldsymbol{u}, \frac{\boldsymbol{s}}{\|\boldsymbol{s}\|_e}\rangle| \Big) \|\boldsymbol{s}\|_e \gamma^{\|\cdot\|}(d\boldsymbol{s}) \\
& + i \langle \boldsymbol{u}, \boldsymbol{m}^0_{\|\cdot\|}\rangle - ia \int_{C_d^{\|\cdot\|}} \langle \boldsymbol{u}, \boldsymbol{s}\rangle \ln \|\boldsymbol{s}\|_e \gamma^{\|\cdot\|}(d\boldsymbol{s}) \Bigg\}, \\
= \exp\Bigg\{ & - \int_{S_d \setminus K^{\|\cdot\|}} \Big(|\langle \boldsymbol{u}, \boldsymbol{s}'\rangle| + ia\langle \boldsymbol{u}, \boldsymbol{s}'\rangle \ln |\langle \boldsymbol{u}, \boldsymbol{s}'\rangle| \Big) \gamma(d\boldsymbol{s}') \\
& + i \langle \boldsymbol{u}, \boldsymbol{m}^0_{\|\cdot\|}\rangle - ia \int_{S_d \setminus K^{\|\cdot\|}} \langle \boldsymbol{u}, \boldsymbol{s}'\rangle \ln \|\boldsymbol{s}'\| \gamma(d\boldsymbol{s}') \Bigg\},
\end{aligned}$$

where we used the change of variable $\boldsymbol{s}' = T^{-1}_{\|\cdot\|}(\boldsymbol{s}) = \boldsymbol{s}/\|\boldsymbol{s}\|_e$, and $\gamma(d\boldsymbol{s}) := \|\boldsymbol{s}\|^{-1}\gamma^{\|\cdot\|} \circ T_{\|\cdot\|}(d\boldsymbol{s})$. Letting then $\overline{\gamma}(A) := \gamma(A \cap (S_d \setminus K^{\|\cdot\|}))$ for any Borel set A of S_d and $\tilde{\boldsymbol{m}m} := (\tilde{m}_i)$ with $\tilde{m}_i = \int_{S_d \setminus K^{\|\cdot\|}} s_i \ln \|\boldsymbol{s}\| \overline{\gamma}(d\boldsymbol{s})$, $j = 1, \ldots, d$, we get

$$\varphi_{\boldsymbol{X}}(\boldsymbol{u}) = \exp\Bigg\{ - \int_{S_d} \Big(|\langle \boldsymbol{u}, \boldsymbol{s}\rangle| + ia\langle \boldsymbol{u}, \boldsymbol{s}\rangle \ln |\langle \boldsymbol{u}, \boldsymbol{s}\rangle| \Big) \overline{\gamma}(d\boldsymbol{s}) + i \langle \boldsymbol{u}, \boldsymbol{m}^0_{\|\cdot\|} - a\tilde{\boldsymbol{m}m}\rangle \Bigg\},$$

and $\boldsymbol{X}$ admits the pair $(\overline{\gamma}, \boldsymbol{m}^0_{\|\cdot\|} - a\tilde{\boldsymbol{m}m})$ for spectral representation on the Euclidean unit sphere. The unicity of the spectral representation of $\boldsymbol{X}$ on S_d implies that $(\Gamma, \boldsymbol{\mu}^0) = (\overline{\gamma}, \boldsymbol{m}^0_{\|\cdot\|} - a\tilde{\boldsymbol{m}m})$. Thus, $\overline{\gamma}$ and Γ have to coincide, and in particular

$$\begin{aligned}
\Gamma(K^{\|\cdot\|}) = \overline{\gamma}(K^{\|\cdot\|}) = \gamma(K^{\|\cdot\|} \cap (S_d \setminus K^{\|\cdot\|})) = \gamma(\emptyset) = 0, \\
\tilde{m}_i = \int_{S_d \setminus K^{\|\cdot\|}} s_i \ln \|\boldsymbol{s}\| \Gamma(d\boldsymbol{s}), \quad i = 1, \ldots, d.
\end{aligned}$$

Last, as $\int_{C_d^{\|\cdot\|}} \|\boldsymbol{s}\|_e \Big| \ln \|\boldsymbol{s}\|_e \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty$ (Lemma 2) and $\Gamma(K^{\|\cdot\|}) = 0$, we have by a change of variable

$$\begin{aligned}
\int_{C_d^{\|\cdot\|}} \|\boldsymbol{s}\|_e \Big| \ln \|\boldsymbol{s}\|_e \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) &= \int_{S_d \setminus K^{\|\cdot\|}} \Big| \ln \|\boldsymbol{s}\| \Big| \|\boldsymbol{s}\|^{-1} \gamma^{\|\cdot\|} \circ T_{\|\cdot\|}(d\boldsymbol{s}) \\
&= \int_{S_d \setminus K^{\|\cdot\|}} \Big| \ln \|\boldsymbol{s}\| \Big| \gamma(d\boldsymbol{s}) \\
&= \int_{S_d} \Big| \ln \|\boldsymbol{s}\| \Big| \Gamma(d\boldsymbol{s}) \\
&< +\infty,
\end{aligned}$$

which concludes the proof of Proposition 1.

Proof of Lemma 2

Notice that there exists a positive real number b such that for all $\boldsymbol{s} \in C_d^{\|\cdot\|}$, $\|\boldsymbol{s}\|_e \geq b$ because $\|\boldsymbol{s}\| = 1$. Letting $M > 0$, we have for all $\boldsymbol{u} \in \mathbb{R}^d$

$$\int_{C_d^{\|\cdot\|}} \|\boldsymbol{s}\|_e \Big| \ln \|\boldsymbol{s}\|_e \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) = \int_{C_d^{\|\cdot\|} \cap \{b \leq \|\boldsymbol{s}\|_e \leq M\}} + \int_{C_d^{\|\cdot\|} \cap \{\|\boldsymbol{s}\|_e > M\}} := I_1 + I_2.$$

We will show that both I_1 and I_2 are finite. Focus first on I_2. From (8), we know that for all $\boldsymbol{u} \in \mathbb{R}^d$

$$\int_{C_d^{\|\cdot\|}} |\langle \boldsymbol{u}, \boldsymbol{s} \rangle| \Big| \ln |\langle \boldsymbol{u}, \boldsymbol{s} \rangle| \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) = \int_{C_d^{\|\cdot\|} \cap \{b \leq \|\boldsymbol{s}\|_e \leq M\}} + \int_{C_d^{\|\cdot\|} \cap \{\|\boldsymbol{s}\|_e > M\}} < +\infty. \tag{54}$$

and thus, in particular

$$\begin{aligned} &\int_{\{\boldsymbol{s}' \in C_d^{\|\cdot\|} \colon \|\boldsymbol{s}'\|_e > M\}} |\langle \boldsymbol{u}, \boldsymbol{s} \rangle| \Big| \ln |\langle \boldsymbol{u}, \boldsymbol{s} \rangle| \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) \\ &\qquad = \int_{\{\boldsymbol{s}' \in C_d^{\|\cdot\|} \colon \|\boldsymbol{s}'\|_e > M\}} |\langle \boldsymbol{u}, \boldsymbol{s} \rangle| \Big| \ln \|\boldsymbol{s}\|_e + \ln |\langle \boldsymbol{u}, \frac{\boldsymbol{s}}{\|\boldsymbol{s}\|_e} \rangle| \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty. \end{aligned} \tag{55}$$

By the triangular inequality, for all $\boldsymbol{u} \in \mathbb{R}^d$,

$$\begin{aligned} &\int_{\{\boldsymbol{s}' \in C_d^{\|\cdot\|} \colon \|\boldsymbol{s}'\|_e > M\}} |\langle \boldsymbol{u}, \boldsymbol{s} \rangle| \Big| \ln \|\boldsymbol{s}\|_e + \ln |\langle \boldsymbol{u}, \frac{\boldsymbol{s}}{\|\boldsymbol{s}\|_e} \rangle| \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) \\ &\qquad = \int_{\{\boldsymbol{s}' \in C_d^{\|\cdot\|} \colon \|\boldsymbol{s}'\|_e > M\}} |\langle \boldsymbol{u}, \boldsymbol{s} \rangle| \Big| \ln \|\boldsymbol{s}\|_e \Big| 1 + \frac{\ln |\langle \boldsymbol{u}, \boldsymbol{s}/\|\boldsymbol{s}\|_e \rangle|}{\ln \|\boldsymbol{s}\|_e} \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) \\ &\qquad \geq \int_{\{\boldsymbol{s}' \in C_d^{\|\cdot\|} \colon \|\boldsymbol{s}'\|_e > M\}} |\langle \boldsymbol{u}, \boldsymbol{s} \rangle| \Big| \ln \|\boldsymbol{s}\|_e \Big| 1 - \Big| \frac{\ln |\langle \boldsymbol{u}, \boldsymbol{s}/\|\boldsymbol{s}\|_e \rangle|}{\ln \|\boldsymbol{s}\|_e} \Big| \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) \end{aligned} \tag{56}$$

Let us now partition the space $\mathbb{R}^d$ into subsets $R_1, \ldots, R_d$ such that, for any $i = 1, \ldots, d$ and any $\boldsymbol{s} = (s_1, \ldots, s_d) \in R_i$, $\sup_j |s_j| = |s_i|$.[9] We have by (55)-(56) that for any $i = 1, \ldots, d$, any $\boldsymbol{u} \in \mathbb{R}^d$,

$$\int_{\{\boldsymbol{s}' \in C_d^{\|\cdot\|} \colon \|\boldsymbol{s}'\|_e > M\} \cap R_i} |\langle \boldsymbol{u}, \boldsymbol{s} \rangle| \Big| \ln \|\boldsymbol{s}\|_e \Big| 1 - \Big| \frac{\ln |\langle \boldsymbol{u}, \boldsymbol{s}/\|\boldsymbol{s}\|_e \rangle|}{\ln \|\boldsymbol{s}\|_e} \Big| \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty.$$

Denoting $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d)$ the canonical orthonormal basis of $\mathbb{R}^d$, evaluate now the above at $\boldsymbol{u} = \boldsymbol{e}_i$. We get that

$$\int_{\{\boldsymbol{s}' \in C_d^{\|\cdot\|} \colon \|\boldsymbol{s}'\|_e > M\} \cap R_i} |\langle \boldsymbol{e}_i, \boldsymbol{s} \rangle| \Big| \ln \|\boldsymbol{s}\|_e \Big| 1 - \Big| \frac{\ln |\langle \boldsymbol{e}_i, \boldsymbol{s}/\|\boldsymbol{s}\|_e \rangle|}{\ln \|\boldsymbol{s}\|_e} \Big| \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty. \tag{57}$$

[9] Strictly speaking, $(R_1, \ldots, R_d)$ is not a partition of $\mathbb{R}^d$ as the R_i's may intersect because of ties in the components of vectors. This will not affect the proof.

Let us show that $\boldsymbol{s} \longmapsto \ln|\langle \boldsymbol{e}_i, \boldsymbol{s}/\|\boldsymbol{s}\|_e\rangle|$ is a bounded function for $\boldsymbol{s} \in \{\boldsymbol{s}' \in C_d^{\|\cdot\|} : \|\boldsymbol{s}'\|_e > M\} \cap R_i$. *Ad absurdum*, if it is not bounded, then for any $A > 0$, there exists $\boldsymbol{s} \in \{\boldsymbol{s}' \in C_d^{\|\cdot\|} : \|\boldsymbol{s}'\|_e > M\} \cap R_i$ such that

$$\Big|\ln|\langle \boldsymbol{e}_i, \boldsymbol{s}/\|\boldsymbol{s}\|_e\rangle|\Big| > A.$$

Taking the sequence $A_n = n$ for any $n \geq 1$, we get that there exists a sequence $(\boldsymbol{s}_n)$, $\boldsymbol{s}_n \in \{\boldsymbol{s}' \in C_d^{\|\cdot\|} : \|\boldsymbol{s}'\|_e > M\} \cap R_i$ such that

$$\Big|\ln|\langle \boldsymbol{e}_i, \boldsymbol{s}_n/\|\boldsymbol{s}_n\|_e\rangle|\Big| > n.$$

Thus, for all $n \geq 1$

$$0 \leq |\langle \boldsymbol{e}_i, \boldsymbol{s}_n/\|\boldsymbol{s}_n\|_e\rangle| \leq e^{-n}.$$

and

$$|\langle \boldsymbol{e}_i, \boldsymbol{s}_n/\|\boldsymbol{s}_n\|_e\rangle| \underset{n\to+\infty}{\longrightarrow} 0.$$

Consider now the decomposition of $\boldsymbol{s}_n/\|\boldsymbol{s}_n\|_e$ in the orthonormal basis $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d)$,

$$\boldsymbol{s}_n/\|\boldsymbol{s}_n\|_e = \sum_{j=1}^{d} \langle \boldsymbol{e}_j, \boldsymbol{s}_n/\|\boldsymbol{s}_n\|_e\rangle \boldsymbol{e}_j.$$

As $\boldsymbol{s}_n \in R_i$ for all $n \geq 1$, we also have that $\boldsymbol{s}_n/\|\boldsymbol{s}_n\|_e \in R_i$ for all $n \geq 1$, and thus, for any $j = 1, \ldots, d$

$$0 \leq |\langle \boldsymbol{e}_j, \boldsymbol{s}_n/\|\boldsymbol{s}_n\|_e\rangle| \leq |\langle \boldsymbol{e}_i, \boldsymbol{s}_n/\|\boldsymbol{s}_n\|_e\rangle| \underset{n\to+\infty}{\longrightarrow} 0.$$

Hence, $\boldsymbol{s}_n/\|\boldsymbol{s}_n\|_e \underset{n\to+\infty}{\longrightarrow} 0$, which is impossible since $\Big\|\boldsymbol{s}_n/\|\boldsymbol{s}_n\|_e\Big\|_e = 1$ for all $n \geq 1$. The function $\boldsymbol{s} \longmapsto \ln|\langle \boldsymbol{e}_i, \boldsymbol{s}/\|\boldsymbol{s}\|_e\rangle|$ is thus bounded on $\{\boldsymbol{s} \in C_d^{\|\cdot\|} : \|\boldsymbol{s}\|_e > M\} \cap R_i$, say $\Big|\ln|\langle \boldsymbol{e}_i, \boldsymbol{s}/\|\boldsymbol{s}\|_e\rangle|\Big| \leq A$ for some $A > 0$. Provided M is taken large enough (e.g., $M > 2A$), we will have in (57)

$$\left|1 - \left|\frac{\ln|\langle \boldsymbol{e}_i, \boldsymbol{s}/\|\boldsymbol{s}\|_e\rangle|}{\ln\|\boldsymbol{s}\|_e}\right|\right| = 1 - \left|\frac{\ln|\langle \boldsymbol{e}_i, \boldsymbol{s}/\|\boldsymbol{s}\|_e\rangle|}{\ln\|\boldsymbol{s}\|_e}\right| \geq 1 - \frac{A}{M} > 0,$$

which thus yields for all $i = 1, \ldots, d$

$$\int_{\{\boldsymbol{s}' \in C_d^{\|\cdot\|} :\ \|\boldsymbol{s}'\|_e > M\} \cap R_i} |\langle \boldsymbol{e}_i, \boldsymbol{s}\rangle| \Big|\ln\|\boldsymbol{s}\|_e\Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty.$$

As $|\langle \boldsymbol{e}_i, \boldsymbol{s}\rangle| \geq \|\boldsymbol{s}\|_e e^{-A}$, we further get that

$$\int_{\{\boldsymbol{s}' \in C_d^{\|\cdot\|} :\ \|\boldsymbol{s}'\|_e > M\} \cap R_i} \|\boldsymbol{s}\|_e \Big|\ln\|\boldsymbol{s}\|_e\Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty,$$

and because $\bigcup_{i=1,\dots,d} R_i = \mathbb{R}^d$,

$$I_2 = \int_{\{\boldsymbol{s}' \in C_d^{\|\cdot\|}:\ \|\boldsymbol{s}'\|_e > M\}} \|\boldsymbol{s}\|_e \Big| \ln \|\boldsymbol{s}\|_e \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s})$$
$$\leq \sum_{i=1}^{d} \int_{\{\boldsymbol{s}' \in C_d^{\|\cdot\|}:\ \|\boldsymbol{s}'\|_e > M\} \cap R_i} \|\boldsymbol{s}\|_e \Big| \ln \|\boldsymbol{s}\|_e \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty.$$

Let us now show that I_1 is finite. Assuming for a moment that

$$\gamma^{\|\cdot\|}\Big(\{\boldsymbol{s}' \in C_d^{\|\cdot\|} :\ b \leq \|\boldsymbol{s}'\|_e \leq M\}\Big) < +\infty,$$

we get

$$I_1 = \int_{\{\boldsymbol{s}' \in C_d^{\|\cdot\|}:\ b \leq \|\boldsymbol{s}'\|_e \leq M\}} \|\boldsymbol{s}\|_e \Big| \ln \|\boldsymbol{s}\|_e \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s})$$
$$\leq \Big(\max_{x \in [b,M]} x |\ln x| \Big)\ \gamma^{\|\cdot\|}\Big(\{\boldsymbol{s}' \in C_d^{\|\cdot\|} :\ b \leq \|\boldsymbol{s}'\|_e \leq M\}\Big),$$

because $x \longmapsto x|\ln x|$ is a bounded function on $[b, M]$, and thus $I_1 < +\infty$. We now show that $\gamma^{\|\cdot\|}$ is indeed finite on the set $\{\boldsymbol{s}' \in C_d^{\|\cdot\|} :\ b \leq \|\boldsymbol{s}'\|_e \leq M\}$.

Proceeding as in the case of I_2, it can be obtained that for $i = 1, \dots, d$, the function $\boldsymbol{s} \longmapsto \ln |\langle \boldsymbol{e}_i, \boldsymbol{s}/\|\boldsymbol{s}\|_e \rangle|$ is bounded on the set $\{\boldsymbol{s}' \in C_d^{\|\cdot\|} :\ b \leq \|\boldsymbol{s}'\|_e \leq M\} \cap R_i$. Say, again, that $\Big| \ln |\langle \boldsymbol{e}_i, \boldsymbol{s}/\|\boldsymbol{s}\|_e \rangle| \Big| \leq A$ for some $A > 0$. Then, $|\langle \boldsymbol{e}_i, \boldsymbol{s} \rangle| \geq \|\boldsymbol{s}\|_e e^{-A}$, and for any $\lambda > 2b^{-1}e^A$, we have

$$|\langle \lambda \boldsymbol{e}_i, \boldsymbol{s} \rangle| \geq 2,$$

for any $i = 1, \dots, d$, $\boldsymbol{s} \in \{\boldsymbol{s}' \in C_d^{\|\cdot\|} :\ b \leq \|\boldsymbol{s}'\|_e \leq M\} \cap R_i$. From (54), we have for any $\boldsymbol{u} \in \mathbb{R}^d$

$$\int_{\{\boldsymbol{s}' \in C_d^{\|\cdot\|}:\ b \leq \|\boldsymbol{s}'\|_e \leq M\}} |\langle \boldsymbol{u}, \boldsymbol{s} \rangle| \Big| \ln |\langle \boldsymbol{u}, \boldsymbol{s} \rangle| \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty,$$

and thus, for any $\boldsymbol{u} \in \mathbb{R}^d$,

$$\int_{\{\boldsymbol{s}' \in C_d^{\|\cdot\|}:\ b \leq \|\boldsymbol{s}'\|_e \leq M\} \cap R_i} |\langle \boldsymbol{u}, \boldsymbol{s} \rangle| \Big| \ln |\langle \boldsymbol{u}, \boldsymbol{s} \rangle| \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty,$$

for any $i = 1, \dots, d$. Evaluating the above in particular at $\boldsymbol{u} = \lambda \boldsymbol{e}_i$, for any $\lambda > 2b^{-1}e^A$, we get

$$\int_{\{\boldsymbol{s}' \in C_d^{\|\cdot\|}:\ b \leq \|\boldsymbol{s}'\|_e \leq M\} \cap R_i} |\langle \lambda \boldsymbol{e}_i, \boldsymbol{s} \rangle| \Big| \ln |\langle \lambda \boldsymbol{e}_i, \boldsymbol{s} \rangle| \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty.$$

Noticing that $x \longmapsto x|\ln x|$ is increasing on $[1, +\infty)$ and that $|\langle \lambda \boldsymbol{e}_i, \boldsymbol{s} \rangle| \geq 2$ for any $\boldsymbol{s}$ in the domain of integration, we have $|\langle \boldsymbol{u}, \boldsymbol{s} \rangle| \Big| \ln |\langle \boldsymbol{u}, \boldsymbol{s} \rangle| \Big| \geq 2 \ln 2$, and

$$\int_{\{\boldsymbol{s}' \in C_d^{\|\cdot\|}:\ b \leq \|\boldsymbol{s}'\|_e \leq M\} \cap R_i} \gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty,$$

for any $i = 1, \dots, d$. Hence,

$$\int_{\{\boldsymbol{s}' \in C_d^{\|\cdot\|}:\ b \leq \|\boldsymbol{s}'\|_e \leq M\}} \gamma^{\|\cdot\|}(d\boldsymbol{s}) \leq \sum_{i=1}^{d} \int_{\{\boldsymbol{s}' \in C_d^{\|\cdot\|}:\ b \leq \|\boldsymbol{s}'\|_e \leq M\} \cap R_i} \gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty,$$

and $\gamma^{\|\cdot\|}\Big(\{\boldsymbol{s}' \in C_d^{\|\cdot\|} :\ b \leq \|\boldsymbol{s}'\|_e \leq M\}\Big)$ is finite. □

E.2 Proof of Proposition 2

The proposition is an immediate consequence of Bayes formula and of the following result, which is an adaptation of Theorem 4.4.8 by Samorodnitsky and Taqqu (1994) Samorodnitsky and Taqqu (1994) to seminorms.

Proposition 9 *Let $\boldsymbol{X} = (X_1, \ldots, X_d)$ be an α-stable random vector and let $\|\cdot\|$ be a seminorm on $\mathbb{R}^d$ such that $\boldsymbol{X}$ is representable on $C_d^{\|\cdot\|}$. Then, for every Borel set $A \subseteq C_d^{\|\cdot\|}$ with $\Gamma^{\|\cdot\|}(\partial A) = 0$,*

$$\lim_{x\to+\infty} x^\alpha \mathbb{P}\Big(\|\boldsymbol{X}\| > x, \frac{\boldsymbol{X}}{\|\boldsymbol{X}\|} \in A\Big) = C_\alpha \Gamma^{\|\cdot\|}(A), \tag{58}$$

with $C_\alpha = \dfrac{1-\alpha}{\Gamma(2-\alpha)\cos(\pi\alpha/2)}$ if $\alpha \neq 1$, and $C_1 = 2/\pi$.

Proof.

We follow the proof of Theorem 4.4.8 by Samorodnitsky and Taqqu (1994). The main hurdle is to show that, with $\|\cdot\|$ a seminorm, $K^{\|\cdot\|} = \{\boldsymbol{s} \in S_d : \|\boldsymbol{s}\| = 0\}$, and $\Gamma^{\|\cdot\|}(K^{\|\cdot\|}) = 0$, we have the series representation of $\boldsymbol{X}$, $(X_1, \ldots, X_d) \stackrel{d}{=} (Z_1, \ldots, Z_d)$ where

$$Z_k = (C_\alpha \Gamma^{\|\cdot\|}(C_d^{\|\cdot\|}))^{1/\alpha} \sum_{i=1}^{\infty} [\Gamma_i^{-1/\alpha} S_i^{(k)} - b_{i,k}(\alpha)], \quad k = 1, \ldots, d, \tag{59}$$

with $\boldsymbol{S}_i = (S_i^{(1)}, \ldots, S_i^{(d)})$, $i \geq 1$, are i.i.d. $C_d^{\|\cdot\|}$-valued random vectors with common law $\Gamma^{\|\cdot\|}/\Gamma^{\|\cdot\|}(C_d^{\|\cdot\|})$ and the $b_{i,k(\alpha)}$'s are constants. By Proposition 1, we know that $\boldsymbol{X}$ admits a characteristic function of the form (5). This allows to restate the integral representation Theorem 3.5.6 in Samorodnitsky and Taqqu (1994) on the seminorm unit cylinder as follows: with the measurable space $(E, \mathcal{E}) = (C_d^{\|\cdot\|}, \text{Borel } \sigma\text{-algebra on } C_d^{\|\cdot\|})$, let M be an α-stable random measure on $(E, \mathcal{E})$ with control measure $m = \Gamma^{\|\cdot\|}$, skewness intensity $\beta(\cdot) \equiv 1$ (see Definition 3.3.1 in Samorodnitsky and Taqqu (1994) for details). Letting also $f_j : C_d^{\|\cdot\|} \longrightarrow \mathbb{R}$ defined by $f_j\Big((s_1, \ldots, s_d)\Big) = s_j$, $j = 1, \ldots, d$, then

$$\boldsymbol{X} \stackrel{d}{=} \Big(\int_{C_d^{\|\cdot\|}} f_1(\boldsymbol{s}) M(d\boldsymbol{s}), \ldots, \int_{C_d^{\|\cdot\|}} f_d(\boldsymbol{s}) M(d\boldsymbol{s}) \Big) + \boldsymbol{\mu}^{\|\cdot\|}.$$

This representation can be checked directly by comparing the characteristic functions of the left-hand and right-hand sides. We can now apply Theorem 3.10.1 in Samorodnitsky and Taqqu (1994) to the above integral representation with $(E, \mathcal{E}, m)$ the measure space as described before, and $\hat{m} = \Gamma^{\|\cdot\|}/\Gamma^{\|\cdot\|}(C_d^{\|\cdot\|})$. This establishes (59). The rest of the proof is similar to that of Theorem 4.4.8 in Samorodnitsky and Taqqu (1994). We rely on the triangle inequality property of seminorms and the fact that any norm is finer than any seminorm in finite dimension.[10] □

[10] We say that a norm N is finer than a seminorm N_s if there is a positive constant C such that $N_s(x) \leq CN(x)$ for any $x \in \mathbb{R}^d$.

E.3 Proof of Theorem 1

From Proposition 1, we know that a necessary condition for the representability of $\boldsymbol{X_t}$ on $C^{\|\cdot\|}_{m+h+1}$ is $\Gamma(K^{\|\cdot\|}) = 0$, where $K^{\|\cdot\|} = \{\boldsymbol{s} \in S_{m+h+1} : \|\boldsymbol{s}\| = 0\}$. This condition is also sufficient when either $\alpha \neq 1$ or $\alpha = 1$, $\beta = 0$. Using the fact that Γ only charges discrete atoms on S_{m+h+1},

$$\begin{aligned}
\Gamma(K^{\|\cdot\|}) = 0 &\iff \{\boldsymbol{s} \in S_{m+h+1} : \Gamma(\{\boldsymbol{s}\}) > 0\} \cap K^{\|\cdot\|} = \emptyset \\
&\iff \forall \boldsymbol{s} \in S_{m+h+1}, \quad \Big[\Gamma(\{\boldsymbol{s}\}) > 0 \Longrightarrow \|\boldsymbol{s}\| > 0\Big] \\
&\iff \forall k \in \mathbb{Z}, \quad \Big[\|\boldsymbol{d}_k\|_e > 0 \Longrightarrow \|\boldsymbol{d}_k\| > 0\Big] \\
&\iff \forall k \in \mathbb{Z}, \quad \Big[\|\boldsymbol{d}_k\| = 0 \Longrightarrow \|\boldsymbol{d}_k\|_e = 0\Big] \\
&\iff \forall k \in \mathbb{Z}, \quad \Big[\|\boldsymbol{d}_k\| = 0 \Longrightarrow \boldsymbol{d}_k = 0\Big] \\
&\iff \forall k \in \mathbb{Z}, \quad \Big[(d_{k+m}, \dots, d_k) = \mathbf{0} \Longrightarrow (d_{k+m}, \dots, d_{k-h}) = \mathbf{0}\Big],
\end{aligned}$$

by (10). Now assume that the following holds:

$$\forall k \in \mathbb{Z}, \quad \Big[(d_{k+m}, \dots, d_k) = \mathbf{0} \Longrightarrow (d_{k+m}, \dots, d_{k-h}) = \mathbf{0}\Big]. \tag{60}$$

Then, if for some particular $k_0 \in \mathbb{Z}$, we have

$$(d_{k_0+m}, \dots, d_{k_0}) = \mathbf{0}.$$

It implies that

$$(d_{k_0+m}, \dots, d_{k_0-h}) = \mathbf{0},$$

and especially, as we assume $h \geq 1$,

$$(d_{(k_0-1)+m}, \dots, d_{k_0-1}) = \mathbf{0}.$$

Invoking (60), we deduce by recurrence that for any $n \geq 0$,

$$(d_{(k_0-n)+m}, \dots, d_{k_0-n}) = \mathbf{0}.$$

Therefore, (60) implies

$$\forall k \in \mathbb{Z}, \quad \Big[(d_{k+m}, \dots, d_k) = \mathbf{0} \Longrightarrow \forall \ell \leq k-1, \quad d_\ell = 0\Big]$$

The reciprocal is clearly true. This establishes that (16) is a necessary and sufficient condition for $\boldsymbol{X}_t$ to be representable on $C^{\|\cdot\|}_d$ in the cases where either $\alpha \neq 1$, or $\alpha = 1$, $\beta = 0$.

In the case $\alpha = 1$, $\beta \neq 0$, Proposition 1 states that the necessary and sufficient condition for representability reads $\int_{S_d} \left| \ln \|\boldsymbol{s}\| \right| \Gamma(d\boldsymbol{s}) < +\infty$. That is

$$\Gamma(K^{\|\cdot\|}) = 0 \qquad \text{and} \qquad \int_{S_d \setminus K^{\|\cdot\|}} \left| \ln \|\boldsymbol{s}\| \right| \Gamma(d\boldsymbol{s}) < +\infty.$$

Substituting Γ by its expression in (15), the above condition holds if and only if (16) is true and

$$\sigma \sum_{\vartheta \in S_1} \sum_{k \in \mathbb{Z}} w_\vartheta \|\boldsymbol{d}_k\|_e \left| \ln \left\| \frac{\vartheta \boldsymbol{d}_k}{\|\boldsymbol{d}_k\|_e} \right\| \right| < +\infty,$$

the latter being equivalent to

$$\sum_{k \in \mathbb{Z}} \|\boldsymbol{d}_k\|_e \left| \ln \frac{\|\boldsymbol{d}_k\|}{\|\boldsymbol{d}_k\|_e} \right| < +\infty.$$

E.4 Proof of Proposition 3

By Definition 2, (X_t) is past-representable if and only if there exists $m \geq 0$, $h \geq 1$ such that the vector $(X_{t-m}, \ldots, X_t, X_{t+1}, \ldots, X_{t+h})$ is representable on $C^{\|\cdot\|}_{m+h+1}$. Consider first point $(\iota)(a)$, that is, the case $\alpha \neq 1$, $(\alpha, \beta) = (1, 0)$. By Theorem 1,

$$\begin{aligned}
(X_t) \text{ is past-representable} &\iff \text{There exist } m \geq 0,\ h \geq 1, \text{ such that (16) holds} \\
&\iff \exists m \geq 0, \forall k \in \mathbb{Z}, \Big[d_{k+m} = \ldots = d_k = 0 \implies \forall \ell \leq k-1,\ d_\ell = 0 \Big].
\end{aligned}$$

Thus,

$$\begin{aligned}
(X_t) \text{ not past-representable} &\iff \forall m \geq 0, \exists k \in \mathbb{Z}, d_{k+m} = \ldots = d_k = 0 \text{ and } \exists \ell \leq k-1,\ d_\ell \neq 0 \\
&\iff \forall m \geq 0, \exists k \in \mathbb{Z}, d_{k+m} = \ldots = d_k = 0 \text{ and } d_{k-1} \neq 0 \\
&\iff \forall m \geq 1, \exists k \in \mathbb{Z}, d_{k+m} = \ldots = d_{k+1} = 0 \text{ and } d_k \neq 0 \\
&\iff \sup\{m \geq 1 :\ \exists k \in \mathbb{Z},\ d_{k+m} = \ldots = d_{k+1} = 0,\ d_k \neq 0\} = +\infty,
\end{aligned}$$

hence (18).

Regarding the last statement of point $(\iota)(a)$, assume first that $m_0 < +\infty$ and $m \geq m_0$. Property (16) necessarily holds with m_0. Indeed, if it did not, there would exist $k \in \mathbb{Z}$ such that

$$d_{k+m_0} = \ldots = d_k = 0, \quad \text{and} \quad d_\ell \neq 0, \quad \text{for some} \quad \ell \leq k-1,$$

and we would have found a sequence of consecutive zero values of length at least m_0+1 preceded by a non-zero value, contradicting the fact that

$$m_0 = \sup\{m \geq 1 :\ \exists\, k \in \mathbb{Z},\ d_{k+m} = \ldots = d_{k+1} = 0, \quad \text{and} \quad d_k \neq 0\}.$$

As (16) holds with m_0, it holds *a fortiori* for any $m' \geq m_0$. Thus, $\boldsymbol{X}_t = (X_{t-m}, \ldots, X_t, X_{t+1}, \ldots, X_{t+h})$ is representable for any $m' \geq m_0$, $h \geq 1$ by Theorem 1, and

(X_t) is in particular (m,h)-past-representable.

Reciprocally let $m \geq 0$, $h \geq 1$ and assume that (X_t) is (m,h)-past-representable. The process (X_t) is thus in particular past-representable, which as we have shown previously, implies that $m_0 < +\infty$. *Ad absurdum*, suppose now that $0 \leq m < m_0 < +\infty$. If $m_0 = 0$, there is nothing to do. Otherwise if $m_0 \geq 1$, by definition, there exists a $k \in \mathbb{Z}$ such that

$$d_{k+m_0} = \ldots = d_{k+1} = 0, \quad \text{and} \quad d_k \neq 0. \tag{61}$$

Because (X_t) is (m,h)-past-representable, we have by Theorem 1 that (16) holds with m. As $m < m_0$ and $d_{k+m_0} = \ldots = d_{k+1} = 0$, we thus have that $d_\ell = 0$ for all $\ell \leq k+1$, and in particular $d_k = 0$, hence the contradiction. We conclude that $m \geq m_0$.

Consider now point $(\iota)(b)$, i.e., the case $\alpha = 1$ and $\beta \neq 0$. From Theorem 1,

$$(X_t) \text{ is past-representable} \iff \text{There exist } m \geq 0,\ h \geq 1, \text{ such that (16) and (17) hold}$$

From the previous proof, we moreover have that

$$\exists\, m \geq 0, \text{ such that (16) holds} \iff m_0 < +\infty \iff \begin{cases} m_0 < +\infty \\ \forall\, m' \geq m_0, \text{ (16) holds} \\ \forall\, m' < m_0, \text{ (16) does not hold} \end{cases}$$

Hence

$$\exists\, m \geq 0,\ h \geq 1, \text{ such that (16) and (17) hold} \iff \begin{cases} m_0 < +\infty \\ \forall\, m' \geq m_0, \text{ (16) holds} \\ \forall\, m' < m_0, \text{ (16) does not hold} \\ \exists\, m \geq 0,\ h \geq 1, \text{ such that (16) and (17) hold.} \end{cases}$$

The latter in particular implies $m_0 < +\infty$ and the existence of $m \geq m_0$, $h \geq 1$ such that (17) holds. Reciprocally,

$$\begin{cases} m_0 < +\infty \\ \exists\, m \geq m_0,\ h \geq 1, \text{ such that (17) holds} \end{cases} \implies \begin{cases} m_0 < +\infty \\ \forall\, m' \geq m_0, \text{ (16) holds} \\ \exists\, m \geq m_0,\ h \geq 1, \text{ such that (17) holds,} \end{cases}$$

which in particular implies that there exists $m \geq m_0$, $h \geq 1$ such that both (16) and (17) hold. Hence the past-representability of (X_t).

In view of Definition 2, point $(\iota\iota)$ is a direct consequence of the second part of Proposition 1.

E.5 Proof of Corollary 1

Letting k_0 be the greatest integer such that $d_{k_0} \neq 0$ (such an index exists by (11)), then immediately, for any $m \geq 1$, $d_{k_0+m} = \ldots = d_{k_0+1} = 0$ and therefore $m_0 = +\infty$.

E.6 Proof of Corollary 2

We first show that $\deg(\Psi) \geq 1$ if and only if $m_0 < +\infty$.

Clearly, if $\deg(\Psi) = 0$, then $X_t = \sum_{k=-\infty}^{k_0} d_k \varepsilon_{t+k}$ for some k_0 in $\mathbb{Z}$ and $m_0 = +\infty$.

Reciprocally, assume $\deg(\Psi) = p \geq 1$. Let us first show that (19) holds.

Denote $\Psi(F)\Phi(B) = \sum_{i=-q}^{p} \varphi_i F^i$ and $\Theta(F)H(B) = \sum_{k=-r}^{s} \theta_i F^i$, for any non-negative degrees $q = \deg(\Phi)$, $r = \deg(H)$, $s = \deg(\Theta)$. From the recursive equation satisfied by (X_t), we have that

$$\begin{aligned}
& \sum_{i=-q}^{p} \varphi_i X_{t+i} = \sum_{k=-r}^{s} \theta_k \varepsilon_{t+k} \\
\iff \quad & \sum_{i=-q}^{p} \varphi_i \sum_{k\in\mathbb{Z}} d_k \varepsilon_{t+k+i} = \sum_{k=-r}^{s} \theta_k \varepsilon_{t+k} \\
\iff \quad & \sum_{k\in\mathbb{Z}} \Big(\sum_{i=-q}^{p} \varphi_i d_{k-i} \Big) \varepsilon_{t+k} = \sum_{k=-r}^{s} \theta_k \varepsilon_{t+k}. \qquad (62)
\end{aligned}$$

Proceeding by identification using the uniqueness of representation of heavy-tailed moving averages (see Gouriéroux and Zakoian (2015)), we get that for $|k| > \max(r, s)$,

$$\sum_{i=-q}^{p} \varphi_i d_{k-i} = 0. \qquad (63)$$

Ad absurdum, if (X_t) is not past-representable, then by Proposition 3

$$\sup\{m \geq 1 : \ \exists k \in \mathbb{Z}, \ d_{k+m} = \ldots = d_{k+1} = 0, \ d_k \neq 0\} = +\infty.$$

Thus, there exists a sequence $\{m_n : \ n \geq 0\}$, $m_n \geq 1$, $\lim_{n\to+\infty} = +\infty$, satisfying: for any $n \geq 0$, there is an index $k \in \mathbb{Z}$ such that

$$d_{k-p} \neq 0 \quad \text{and} \quad d_{k-p+1} = d_{k-p+2} = \ldots = d_{k+m_n} = 0.$$

We can therefore construct a sequence (k_n) such that the above relation holds for all $n \geq 0$. This sequence of integers in $\mathbb{Z}$ is either bounded or unbounded. We will show that both cases lead to a contradiction.

First case: $\sup\{|k_n| : n \geq 0\} = +\infty$

There are two subsequences such that $m_{g(n)} \longrightarrow +\infty$ and $|k_{g(n)}| \longrightarrow +\infty$. For some n large enough such that (63) holds and $m_{g(n)} \geq p + q$, we have both

$$\sum_{i=-q}^{p} \varphi_i d_{k_{g(n)}-i} = 0.$$

and

$$d_{k_{g(n)}-p} \neq 0, \quad d_{k_{g(n)}-p+1} = \ldots = d_{k_{g(n)}+q} = 0.$$

Hence,

$$\varphi_p d_{k_{g(n)}-p} = 0,$$

which is impossible given that $d_{k_{g(n)}-p} \neq 0$ and $\varphi_p \neq 0$. Indeed, denoting $\Psi(z) = 1 + \psi_1 z + \ldots + \psi_p z^p$, $\psi_p \neq 0$ because $\deg(\Psi) = p$, it can be shown that $\varphi_p = \psi_p$.

Second case: $\sup\{|k_n| : n \geq 0\} < +\infty$

Given that (k_n) is a bounded sequence, there exists by the Bolzano-Weierstrass theorem a convergent subsquence $(k_{g(n)})$. As $(k_{g(n)})$ takes only discrete values, it necessarily holds that $(k_{g(n)})$ reaches its limit at a finite integer $n_0 \geq 1$, that is, for all $n \geq n_0$, $k_{g(n)} = \lim_{n\to+\infty} k_{g(n)} := \bar{k} \in \mathbb{Z}$. Thus, for all $n \geq n_0$

$$d_{\bar{k}} \neq 0, \quad \text{and} \quad d_{\bar{k}+m_{g(n)}} = 0,$$

and as $m_{g(n)} \to +\infty$, we deduce that

$$d_{\bar{k}} \neq 0, \quad \text{and} \quad d_{\bar{k}+\ell} = 0, \quad \text{for all} \quad \ell \geq 1.$$

The process (X_t) hence admit a moving average representation of the form

$$X_t = \sum_{k=-\infty}^{\bar{k}} d_k \varepsilon_{t+k}, \qquad t \in \mathbb{Z}. \tag{64}$$

However, we also have by partial fraction decomposition

$$\begin{aligned}
X_t &= \frac{\Theta(F)H(B)}{\Psi(F)\Phi(B)}\varepsilon_t \\
&= \Theta(F)H(B)\frac{B^p}{B^p\Psi(F)\Phi(B)}\varepsilon_t \\
&= \Theta(F)H(B)B^p\left[\frac{b_1(B)}{B^p\Psi(F)} + \frac{b_2(B)}{\Phi(B)}\right]\varepsilon_t \\
&= \Theta(F)H(B)\left[\frac{b_1(B)}{\Psi(F)} + \frac{B^p b_2(B)}{\Phi(B)}\right]\varepsilon_t,
\end{aligned}$$

for some polynomials $b1$ and b_2 such that $0 \leq \deg(b_1) \leq p-1$, $0 \leq \deg(b_2) \leq q-1$ and $\Phi(B)b_1(B) + B^p b_2(B)\Psi(F) = 1$. We can write in general

$$\begin{aligned}
\frac{\Theta(F)H(B)b_1(B)}{\Psi(F)} &= \sum_{k=-\ell_1}^{+\infty} c_k \varepsilon_{t+k}, \\
\frac{\Theta(F)H(B)B^p b_2(B)}{\Phi(B)} &= \sum_{k=-\infty}^{\ell_2} e_k \varepsilon_{t+k},
\end{aligned}$$

for some sequences of coefficients (c_k), (e_k), and where ℓ_1 is the degree of the largest order monomial in B of $\Theta(F)H(B)b_1(B)$ (recall that $F = B^{-1}$) and ℓ_2 is the degree of the largest

monomial in F of $B^p\Theta(F)H(B)b_2(B)$. By (64), we deduce by identification that there is some $\bar{\ell} \in \mathbb{Z}$ such that $c_k = 0$ for all $k \geq \bar{\ell} + 1$ and

$$\frac{\Theta(F)H(B)b_1(B)}{\Psi(F)} = \sum_{k=-\ell_1}^{\bar{\ell}} c_k F^k.$$

Necessarily, $\bar{\ell} \geq \ell_1$, otherwise $\Theta(F)H(B)b_1(B)\Psi^{-1}(F) = 0$ which is impossible as all the polynomials involved have non-negative degrees. Thus, we deduce that there exist two polynomials P and Q of non-negative degrees such that

$$\frac{\Theta(z^{-1})H(z)b_1(z)}{\Psi(z^{-1})} = \sum_{k=-\ell_1}^{\bar{\ell}} c_k z^k := P(z^{-1}) + Q(z), \qquad z \in \mathbb{C}.,$$

which yields

$$\Theta(z^{-1})H(z)b_1(z) = \Psi(z^{-1})(P(z^{-1}) + Q(z)), \qquad z \in \mathbb{C}. \tag{65}$$

As $\deg(\Psi) = p$ and $\Psi(z) = 0$ if and only if $|z| > 1$, we know that there are p complex numbers $z_1, \ldots, z_p$ such that $0 < |z_i| < 1$ and $\Psi(z_i^{-1}) = 0$ for $i = 1, \ldots, p$. Evaluating (65) at the z_i's, we get that

$$\Theta(z_i^{-1})b_1(z_i) = 0, \qquad \text{for} \quad i = 1, \ldots, p,$$

because H has no roots inside the unit circle and P and Q are of finite degrees. From the fact that $\deg(b_1) \leq p - 1$, we also know that for some z_{i_0}, $b(z_{i_0}) \neq 0$ which finally yields

$$\Theta(z_{i_0}^{-1}) = 0.$$

We therefore obtain that Ψ and Θ have a common root, which is ruled out by assumption, hence the contradiction. The sequence (k_n) can thus be neither bounded nor unbounded, which is absurd. We conclude that

$$m_0 = \sup\{m \geq 1 : \ \exists k \in \mathbb{Z}, \ d_{k+m} = \ldots = d_{k+1} = 0, \ d_k \neq 0\} < +\infty.$$

Hence the equivalence between $(\iota\iota)$ and $(\iota\iota\iota)$.

Let us now show that whenever $m_0 < +\infty$, then (16) holds for any $m \geq m_0$.

As $m_0 < +\infty$, we have that for any $m \geq m_0$ and $h \geq 1$, $\|\boldsymbol{d}_k\| > 0$ as soon as $\boldsymbol{d}_k \neq \mathbf{0}$, for all $k \in \mathbb{Z}$ (recall $\boldsymbol{d}_k = (d_{k+m}, \ldots, d_k, d_{k+1}, \ldots, d_{k-h})$). For ARMA processes, the non-zero coefficients d_k of the moving average necessarily decay geometrically (times a monomial) as $k \to \pm\infty$. To fix ideas, say $d_k \underset{k\to\pm\infty}{\sim} ak^b\lambda^k$, for constants $a \neq 0$, b a non-negative integer, and $0 < |\lambda| < 1$, which may change according to whether $k \to +\infty$ or $k \to -\infty$ (if $\deg(\Phi) = 0$, then

$d_{-k} = 0$ for $k \geq 0$ large enough, however, since we assume $\deg(\Psi) \geq 1$, it always holds that $|d_k| \underset{k\to+\infty}{\sim} ak^b\lambda^k$, for the non-zero terms d_k). Hence,

$$\boldsymbol{d}_k \underset{k\to\pm\infty}{\sim} ak^b\lambda^k\boldsymbol{d}_*,$$

for some constant vector $\boldsymbol{d}_*$ such that $\|\boldsymbol{d}_*\| > 0$ (which may change according to whether $k \to +\infty$ or $k \to -\infty$). We then have that

$$\frac{\|\boldsymbol{d}_k\|}{\|\boldsymbol{d}_k\|_e} \underset{k\to\pm\infty}{\longrightarrow} \frac{\|\boldsymbol{d}_*\|}{\|\boldsymbol{d}_*\|_e} > 0,$$

and

$$\|\boldsymbol{d}_k\|_e \left| \ln\Big(\|\boldsymbol{d}_k\|/\|\boldsymbol{d}_k\|_e\Big)\right| \leq C \;\; k^b|\lambda|^k, \quad k \to \pm\infty,$$

where C is a constant that depends on a, λ, m and h. Therefore, for any $m \geq m_0$, $h \geq 1$,

$$\sum_{k\in\mathbb{Z}} \|\boldsymbol{d}_k\|_e \left| \ln\Big(\|\boldsymbol{d}_k\|/\|\boldsymbol{d}_k\|_e\Big)\right| < +\infty$$

The equivalence between (ι) and $(\iota\iota)$ is now clear: on the one hand, if $m_0 < +\infty$, then (16) holds for all $m \geq m_0$, $h \geq 1$, which yields the (m,h)-past-representability of $(X_{t-m},\ldots,X_t,X_{t+1},\ldots,X_{t+h})$ for any $m \geq m_0$, $h \geq 1$, by Theorem 1. In particular, (X_t) is past-representable. On the other hand, assuming (X_t) is past-representable, then necessarily $m_0 < +\infty$.

Regarding the last statement, it follows from the above proof that the condition $m_0 < +\infty$ and $m \geq m_0$ is sufficient for (m,h)-past-representability. It is also necessary, as (16) never holds with $m < m_0$ (*a fortiori*, with $m < m_0 = +\infty$), concluding the proof.

E.7 Proof of Proposition 4

By Proposition 2

$$\mathbb{P}_x^{\|\cdot\|}(\boldsymbol{X_t}, A|B) \underset{x\to+\infty}{\longrightarrow} \frac{\Gamma^{\|\cdot\|}(A\cap B(V))}{\Gamma^{\|\cdot\|}(B(V))}.$$

The conclusion follows by considering the points of $B(V)$ and $A \cap B(V)$ that are charged by the spectral measure $\Gamma^{\|\cdot\|}$ in (23).

E.8 Proof of Lemma 1

By Proposition 3, we have

$$\Gamma^{\|\cdot\|} = \sum_{\vartheta\in S_1}\sum_{k\in\mathbb{Z}} \|\boldsymbol{d}_k\|^\alpha \delta_{\left\{\frac{\vartheta \boldsymbol{d}_k}{\|\boldsymbol{d}_k\|}\right\}},$$

with $\boldsymbol{d}_k = (\rho^{k+m}\mathbb{1}_{\{k+m\geq 0\}}, \ldots, \rho^{k-h}\mathbb{1}_{\{k-h\geq 0\}})$ and $k \in \mathbb{Z}$. Thus,

$$\boldsymbol{d}_k = \begin{cases} \mathbf{0}, & \text{if} \quad k \leq -m-1, \\ (\rho^{k+m}, \ldots, \rho, 1, 0, \ldots, 0), & \text{if} \quad -m \leq k \leq h, \\ \rho^{k-h}\boldsymbol{d}_h, & \text{if} \quad k \geq h. \end{cases}$$

Therefore,

$$\Gamma^{\|\cdot\|} = \sum_{\vartheta\in S_1}\left[\sum_{k=-m}^{h-1}\|\boldsymbol{d}_k\|^\alpha \delta_{\left\{\frac{\vartheta \boldsymbol{d}_k}{\|\boldsymbol{d}_k\|}\right\}} + \sum_{k=h}^{+\infty}|\rho|^{\alpha(k-h)}\|\boldsymbol{d}_h\|^\alpha \delta_{\left\{\frac{\vartheta\rho^{k-h}\boldsymbol{d}_h}{|\rho|^{k-h}\|\boldsymbol{d}_h\|}\right\}}\right].$$

Moreover,

$$\begin{aligned} &\sum_{\vartheta\in S_1}\sum_{k=h}^{+\infty}|\rho|^{\alpha(k-h)}\|\boldsymbol{d}_h\|^\alpha \delta_{\left\{\mathrm{sign}(\rho)^{k-h}\frac{\vartheta \boldsymbol{d}_h}{\|\boldsymbol{d}_h\|}\right\}} \\ &\qquad = \sum_{\vartheta\in S_1}\|\boldsymbol{d}_h\|^\alpha \frac{1}{2}\left[\sum_{k=h}^{+\infty}|\rho|^{\alpha(k-h)} + \vartheta\beta\sum_{k=h}^{+\infty}(\rho^{<\alpha>})^{k-h}\right]\delta_{\left\{\frac{\vartheta \boldsymbol{d}_h}{\|\boldsymbol{d}_h\|}\right\}} \\ &\qquad = \sum_{\vartheta\in S_1}\frac{1}{1-|\rho|^\alpha}\|\boldsymbol{d}_h\|^\alpha \bar{w}_\vartheta \delta_{\left\{\frac{\vartheta \boldsymbol{d}_h}{\|\boldsymbol{d}_h\|}\right\}}. \end{aligned}$$

Finally, noticing that for $k=-m$ and $\boldsymbol{d}_k = (1,0,\ldots,0)$,

$$\begin{aligned} \Gamma^{\|\cdot\|} &= \sum_{\vartheta\in S_1}\left[w_\vartheta\sum_{k=-m}^{h-1}\|\boldsymbol{d}_k\|^\alpha\delta_{\left\{\frac{\vartheta\boldsymbol{d}_k}{\|\boldsymbol{d}_k\|}\right\}} + \frac{\bar{w}_\vartheta}{1-|\rho|^\alpha}\|\boldsymbol{d}_h\|^\alpha\delta_{\left\{\frac{\vartheta\,\boldsymbol{d}_h}{\|\boldsymbol{d}_h\|}\right\}}\right] \\ &= \sum_{\vartheta\in S_1}\left[w_\vartheta\left(\delta_{\{(\vartheta,0,\ldots,0)\}} + \sum_{k=-m+1}^{h-1}\|\boldsymbol{d}_k\|^\alpha\delta_{\left\{\frac{\vartheta\boldsymbol{d}_k}{\|\boldsymbol{d}_k\|}\right\}}\right) + \frac{\bar{w}_\vartheta}{1-|\rho|^\alpha}\|\boldsymbol{d}_h\|^\alpha\delta_{\left\{\frac{\vartheta\,\boldsymbol{d}_h}{\|\boldsymbol{d}_h\|}\right\}}\right] \\ &= \sum_{\vartheta\in S_1}\left[w_\vartheta\delta_{\{(\vartheta,0,\ldots,0)\}} + \left(w_\vartheta\sum_{k=-m+1}^{h-1}\|\boldsymbol{d}_k\|^\alpha\delta_{\left\{\frac{\vartheta\boldsymbol{d}_k}{\|\boldsymbol{d}_k\|}\right\}} + \frac{\bar{w}_\vartheta}{1-|\rho|^\alpha}\|\boldsymbol{d}_h\|^\alpha\delta_{\left\{\frac{\vartheta\,\boldsymbol{d}_h}{\|\boldsymbol{d}_h\|}\right\}}\right)\right]. \end{aligned}$$

E.9 Proof of Proposition 5

Lemma 3 *Let $\Gamma^{\|\cdot\|}$ be the spectral measure given in Lemma 1 and assume that the ρ is positive. Letting $(\vartheta_0, k_0) \in \mathcal{I}$, consider*

$$I_0 := \left\{\frac{\vartheta'\boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|} : \quad \frac{\vartheta' f(\boldsymbol{d}_{k'})}{\|\boldsymbol{d}_{k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{k_0})}{\|\boldsymbol{d}_{k_0}\|} \ \textit{ for } \ (\vartheta', k') \in \mathcal{I}\right\}.$$

For $m \geq 1$, and $0 \leq k_0 \leq h$, then

$$I_0 = \left\{\frac{\vartheta_0\boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|} : \quad 0 \leq k' \leq h\right\}.$$

For $m \geq 1$, and $-m \leq k_0 \leq -1$, then

$$I_0 = \begin{cases} \left\{\frac{\vartheta_0\boldsymbol{d}_{k_0}}{\|\boldsymbol{d}_{k_0}\|}\right\}, & \textit{if} \quad -m+1 \leq k_0 \leq -1 \\ \\ \left\{\frac{\vartheta_0\boldsymbol{d}_{0,k_0}}{\|\boldsymbol{d}_{0,k_0}\|}\right\} = \{(\vartheta_0, 0, \ldots, 0)\}, & \textit{if} \quad k_0 = -m. \end{cases}$$

For $m = 0$, then

$$I_0 = \left\{ \frac{\vartheta_0 \boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|} : \quad k' \in \{1, \ldots, h\} \cup \{(0,0)\} \right\}.$$

Proof.

Case $m \geq 1$ and $k_0 \in \{0, \ldots, h\}$

If $k' \in \{-m, \ldots, -1\}$, the $(m{+}1)^{\text{th}}$ component of $f(\boldsymbol{d}_{k'})$ is zero, whereas the $(m{+}1)^{\text{th}}$ component of $f(\boldsymbol{d}_{k_0})$ is $\rho^{k_0} \neq 0$. Necessarily, $\vartheta' f(\boldsymbol{d}_{k'})/\|\boldsymbol{d}_{k'}\| \neq \vartheta_0 f(\boldsymbol{d}_{k_0})/\|\boldsymbol{d}_{k_0}\|$ and

$$I_0 = \left\{ \frac{\vartheta' \boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|} : \quad \frac{\vartheta' f(\boldsymbol{d}_{k'})}{\|\boldsymbol{d}_{k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{k_0})}{\|\boldsymbol{d}_{k_0}\|} \text{ for } (\vartheta', k') \in \{-1, +1\} \times \{0, \ldots, h\} \right\}.$$

Now, with $k' \in \{0, \ldots, h\}$, we have that

$$\begin{aligned} f(\boldsymbol{d}_{k'}) &= (\rho^{k'+m}, \ldots, \rho^{k'+1}, \rho^{k'}), \\ f(\boldsymbol{d}_{k_0}) &= (\rho^{k_0+m}, \ldots, \rho^{k_0+1}, \rho^{k_0}), \end{aligned}$$

and by Lemma 1 we also have that

$$\begin{aligned} \|\boldsymbol{d}_{k'}\| &= \|(\rho^{k'+m}, \ldots, \rho^{k'+1}, \rho^{k'}, \overbrace{0, \ldots, 0}^{h})\|, \\ \|\boldsymbol{d}_{k_0}\| &= \|(\rho^{k_0+m}, \ldots, \rho^{k_0+1}, \rho^{k_0}, \underbrace{0, \ldots, 0}_{h})\|. \end{aligned}$$

Thus,

$$\begin{aligned} \frac{\vartheta' f(\boldsymbol{d}_{k'})}{\|\boldsymbol{d}_{k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{k_0})}{\|\boldsymbol{d}_{k_0}\|} & \\ \iff \frac{\vartheta' \rho^{k'} f(\boldsymbol{d}_0)}{|\rho|^{k'} \|\boldsymbol{d}_0\|} &= \frac{\vartheta_0 \rho^{k_0} f(\boldsymbol{d}_0)}{|\rho|^{k_0} \|\boldsymbol{d}_0\|} \\ \iff \frac{\vartheta' \rho^{\ell}}{\|\boldsymbol{d}_0\|} &= \frac{\vartheta_0 \rho^{\ell}}{\|\boldsymbol{d}_0\|}, \qquad \ell = 0, \ldots, m \\ \iff \vartheta' \vartheta_0 \frac{\|\boldsymbol{d}_0\|}{\|\boldsymbol{d}_0\|} &= \left(\frac{\rho}{\rho}\right)^{\ell}, \quad \ell = 0, \ldots, m \\ \iff \vartheta' \vartheta_0 &= 1 \\ \iff \vartheta' &= \vartheta_0, \end{aligned}$$

because $\rho \neq 0$ is assumed.

Case $m \geq 1$ and $k_0 \in \{-m, \ldots, -1\}$

By comparing the place of the first zero component, it is easy to see that

$$\frac{\vartheta' f(\boldsymbol{d}_{k'})}{\|\boldsymbol{d}_{k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{k_0})}{\|\boldsymbol{d}_{k_0}\|} \Longrightarrow k' = k_0.$$

$$\begin{aligned} f(\boldsymbol{d}_{k'}) &= (\overbrace{\rho^{k'+m}, \ldots, \rho, 1, 0, \ldots, 0}^{m+1}, \overbrace{0, \ldots, 0}^{h}), \\ f(\boldsymbol{d}_{k_0}) &= (\underbrace{\rho^{k_0+m}, \ldots, \rho, 1, 0, \ldots, 0}_{m+1}, \underbrace{0, \ldots, 0}_{h}), \end{aligned}$$

and we also have that

$$\|\boldsymbol{d}_{k'}\| = \|(\overbrace{\rho^{k'+m},\dots,\rho,1,0,\dots,0}^{m+1},\overbrace{0,\dots,0}^{h})\|,$$
$$\|\boldsymbol{d}_{k_0}\| = \|(\underbrace{\rho^{k_0+m},\dots,\rho,1,0,\dots,0}_{m+1},\underbrace{0,\dots,0}_{h})\|.$$

As $k' = k_0 \leq -1$,

$$\begin{aligned}\frac{\vartheta' f(\boldsymbol{d}_{k'})}{\|\boldsymbol{d}_{k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{k_0})}{\|\boldsymbol{d}_{k_0}\|}&\\ \iff \frac{\vartheta'\rho^\ell}{\|\boldsymbol{d}_{k_0}\|} = \frac{\vartheta_0\rho^\ell}{\|\boldsymbol{d}_{k_0}\|},&\quad \ell = 0,\dots,m+k_0,\ \text{and}\ k' = k_0\\ \iff \vartheta'\vartheta_0\frac{\|\boldsymbol{d}_{k_0}\|}{\|\boldsymbol{d}_{k_0}\|} = \left(\frac{\rho}{\rho}\right)^\ell,&\quad \ell = 0,\dots,m+k_0,\ \text{and}\ k' = k_0.\end{aligned}$$

Now if $-m+1 \leq k_0 \leq -1$,

$$\begin{aligned}\vartheta'\vartheta_0\frac{\|\boldsymbol{d}_{k_0}\|}{\|\boldsymbol{d}_{k_0}\|} = \left(\frac{\rho}{\rho}\right)^\ell,&\quad \ell = 0,1,\dots,m+k_0,\ \text{and}\ k' = k_0\\ &\iff \vartheta' = \vartheta_0\ \text{ and }\ k' = k_0.\end{aligned}$$

If $k_0 = -m$, given that $(\vartheta_0, k_0) \in \mathcal{I} = S_1 \times \Big(\{-m,\dots,-1,0,1,\dots,h\} \cup \{(0,-m)\}\Big)$, and as $k' = k_0 = -m$, we have that $\boldsymbol{d}_{k_0} = \boldsymbol{d}_{0,-m} = (1,0,\dots,0)$. Hence

$$\begin{aligned}\vartheta'\vartheta_0\frac{\|\boldsymbol{d}_{k_0}\|}{\|\boldsymbol{d}_{k_0}\|} = \left(\frac{\rho}{\rho}\right)^\ell,&\quad \ell = 0,\ \text{and}\ k' = k_0 = -m\ ,\\ &\iff \vartheta' = \vartheta_0\ \text{and}\ k' = k_0 = -m\end{aligned}$$

Case $m = 0$

If $k_0 \in \{1,\dots,h\}$ then $f(\boldsymbol{d}_{k_0}) = \rho^{k_0}$ and by Lemma 1, $\|\boldsymbol{d}_{k_0}\| = |\rho|^{k_0}$. Thus, $\vartheta_0 f(\boldsymbol{d}_{k_0})/\|\boldsymbol{d}_{k_0}\| = \vartheta_0$. If $k_0 = -m = 0$, then $f(\boldsymbol{d}_{k_0}) = 1$ and $\vartheta_0 f(\boldsymbol{d}_{k_0})/\|\boldsymbol{d}_{k_0}\| = \vartheta_0$. The same holds for $(\vartheta', k') \in \mathcal{I}$ and we obtain that

$$\frac{\vartheta' f(\boldsymbol{d}_{k'})}{\|\boldsymbol{d}_{k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{k_0})}{\|\boldsymbol{d}_{k_0}\|} \iff \vartheta' = \vartheta_0.$$

□

Let us now prove Proposition 5. By Proposition 4,

$$\mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t}, A_{\vartheta,k}\Big|B(V_0)\Big) \underset{x\to\infty}{\longrightarrow} \frac{\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta'\boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|} \in A_{\vartheta,k} :\ \frac{\vartheta' f(\boldsymbol{d}_{,k'})}{\|\boldsymbol{d}_{k'}\|} \in V_0\right\}\right)}{\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta'\boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|} \in C_{m+h+1}^{\|\cdot\|} :\ \frac{\vartheta' f(\boldsymbol{d}_{k'})}{\|\boldsymbol{d}_{k'}\|} \in V_0\right\}\right)}. \tag{66}$$

Focusing on the denominator, we have by (27)

$$\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta'\boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|} \in C_{m+h+1}^{\|\cdot\|} :\ \frac{\vartheta' f(\boldsymbol{d}_{k'})}{\|\boldsymbol{d}_{k'}\|} \in V_0\right\}\right) = \Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta'\boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|} \in C_{m+h+1}^{\|\cdot\|} :\ \frac{\vartheta' f(\boldsymbol{d}_{k'})}{\|\boldsymbol{d}_{k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{k_0})}{\|\boldsymbol{d}_{k_0}\|}\right\}\right)$$

We will now distinguish the cases arising from the application of Lemma 3. Recall that we assume for this proposition that the ρ is positive. Thus, $\text{sign}(\rho) = 1$ and $\bar{\beta} = \beta \frac{1-|\rho|^\alpha}{1-\rho^{<\alpha>}} = \beta$ and $\bar{w}_\vartheta = w_\vartheta$ in Lemma 1 for $\vartheta \in \{-1, +1\}$.

Case $m \geq 1$ and $0 \leq k_0 \leq h$

By Lemma 3,

$$\begin{aligned}\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta' \boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|} \in C^{\|\cdot\|}_{m+h+1} : \quad \frac{\vartheta' f(\boldsymbol{d}_{k'})}{\|\boldsymbol{d}_{k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{k_0})}{\|\boldsymbol{d}_{k_0}\|}\right\}\right) \\ = \Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta_0 \boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|} : \quad 0 \leq k' \leq h\right\}\right) \\ = \left[w_{\vartheta_0} \sum_{k'=0}^{h-1} \|\boldsymbol{d}_{k'}\|^\alpha + \frac{\bar{w}_{\vartheta_0}}{1-|\rho|^\alpha} \|\boldsymbol{d}_h\|^\alpha\right]\end{aligned}$$

By Lemma 1, for $k' \in \{0, 1, \ldots, h\}$

$$\begin{aligned}\|\boldsymbol{d}_{k'}\| &= \|(\rho^{k'+m}, \ldots, \rho^{k'+1}, \rho^{k'}, \underbrace{0, \ldots, 0}_{h})\| \\ &= |\rho|^{k'-h} \|(\rho^{m+h}, \ldots, \rho^{h+1}, \rho^h, \underbrace{0, \ldots, 0}_{h})\| \\ &= |\rho|^{k'-h} \|\boldsymbol{d}_h\|.\end{aligned}$$

Thus,

$$\begin{aligned}\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta' \boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|} \in C^{\|\cdot\|}_{m+h+1} : \quad \frac{\vartheta' f(\boldsymbol{d}_{k'})}{\|\boldsymbol{d}_{k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{k_0})}{\|\boldsymbol{d}_{k_0}\|}\right\}\right) &= w_{\vartheta_0} \|\boldsymbol{d}_h\|^\alpha \left[\sum_{k'=0}^{h-1} \rho^{\alpha(k'-h)} + \frac{1}{1-|\rho|^\alpha}\right] \\ &= w_{\vartheta_0} \|\boldsymbol{d}_h\|^\alpha \frac{|\rho|^{-\alpha h}}{1-|\rho|^\alpha}.\end{aligned}$$

Similarly for the numerator in (66), by (28),

$$\begin{aligned}\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta' \boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|} \in A_{\vartheta,k} : \quad \frac{\vartheta' f(\boldsymbol{d}_{k'})}{\|\boldsymbol{d}_{k'}\|} \in V_0\right\}\right) \\ = \Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta_0 \boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|} \in A_{\vartheta,k} : \quad 0 \leq k' \leq h\right\}\right)\end{aligned}$$

$$= \begin{cases} \Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta_0 \boldsymbol{d}_k}{\|\boldsymbol{d}_k\|}\right\}\right), & \text{if} \quad \vartheta = \vartheta_0, \\ \Gamma^{\|\cdot\|}(\emptyset), & \text{if} \quad \vartheta \neq \vartheta_0, \end{cases}$$

$$= \begin{cases} w_{\vartheta_0} \|\boldsymbol{d}_h\|^\alpha |\rho|^{\alpha(k-h)} \delta_{\{\vartheta_0\}}(\vartheta), & \text{if} \quad 0 \leq k \leq h-1, \\ w_{\vartheta_0} \|\boldsymbol{d}_h\|^\alpha \frac{1}{1-|\rho|^\alpha} \delta_{\{\vartheta_0\}}(\vartheta), & \text{if} \quad k = h. \end{cases}$$

The conclusion follows.

Case $m \geq 1$ and $-m \leq k_0 \leq -1$

We have by Lemma 3

$$\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta' \boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|} \in C_{m+h+1}^{\|\cdot\|} : \quad \frac{\vartheta' f(\boldsymbol{d}_{k'})}{\|\boldsymbol{d}_{k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{k_0})}{\|\boldsymbol{d}_{k_0}\|}\right\}\right) = \Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta_0 \boldsymbol{d}_{k_0}}{\|\boldsymbol{d}_{k_0}\|}\right\}\right).$$

If $-m+1 \leq k_0 \leq -1$,

$$\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta_0 \boldsymbol{d}_{k_0}}{\|\boldsymbol{d}_{k_0}\|}\right\}\right) = w_{\vartheta_0}\|\boldsymbol{d}_{k_0}\|^{\alpha},$$

and

$$\begin{aligned}
&\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta' \boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|} \in A_{\vartheta,k} : \quad \frac{\vartheta' f(\boldsymbol{d}_{k'})}{\|\boldsymbol{d}_{k'}\|} \in V_0\right\}\right) \\
&\qquad = \Gamma^{\|\cdot\|}\left(A_{\vartheta,k} \cap \left\{\frac{\vartheta_0 \boldsymbol{d}_{k_0}}{\|\boldsymbol{d}_{k_0}\|}\right\}\right) \\
&\qquad = \begin{cases} \Gamma^{\|\cdot\|}\left(\left\{\dfrac{\vartheta_0 \boldsymbol{d}_{k_0}}{\|\boldsymbol{d}_{k_0}\|}\right\}\right), & \text{if } \vartheta = \vartheta_0, \text{ and } k = k_0, \\ \Gamma^{\|\cdot\|}(\emptyset), & \text{if } \vartheta \neq \vartheta_0 \text{ or } k \neq k_0, \end{cases} \\
&\qquad = w_{\vartheta_0}\|\boldsymbol{d}_{k_0}\|^{\alpha}\delta_{\{\vartheta_0\}}(\vartheta)\delta_{\{k_0\}}(k).
\end{aligned}$$

If $k_0 = -m$, then $\boldsymbol{d}_{k_0} = \boldsymbol{d}_{0,-m} = (1, 0, \ldots, 0)$, and

$$\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta_0 \boldsymbol{d}_{k_0}}{\|\boldsymbol{d}_{k_0}\|}\right\}\right) = \Gamma^{\|\cdot\|}\Big(\{\vartheta_0(1, 0, \ldots, 0)\}\Big) = w_{\vartheta_0},$$

and

$$\begin{aligned}
&\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta' \boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|} \in A_{\vartheta,k} : \quad \frac{\vartheta' f(\boldsymbol{d}_{k'})}{\|\boldsymbol{d}_{k'}\|} \in V_0\right\}\right) \\
&\qquad = \Gamma^{\|\cdot\|}\left(A_{\vartheta,k} \cap \left\{\frac{\vartheta_0 \boldsymbol{d}_{k_0}}{\|\boldsymbol{d}_{k_0}\|}\right\}\right) \\
&\qquad = \begin{cases} \Gamma^{\|\cdot\|}\Big(A_{\vartheta,k} \cap \{\vartheta_0(1, 0, \ldots, 0)\}\Big), & \text{if } \vartheta = \vartheta_0, \text{ and } k = k_0 = -m, \\ \Gamma^{\|\cdot\|}(\emptyset), & \text{if } \vartheta \neq \vartheta_0 \text{ or } k \neq k_0, \end{cases} \\
&\qquad = w_{\vartheta_0}\delta_{\{\vartheta_0\}}(\vartheta)\delta_{\{k_0\}}(k).
\end{aligned}$$

The conclusion follows as previously.

Case $m = 0$

By Lemma 3, as the ρ is positive

$$\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta' \boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|} \in C^{\|\cdot\|}_{m+h+1} : \ \frac{\vartheta' f(\boldsymbol{d}_{k'})}{\|\boldsymbol{d}_{k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{k_0})}{\|\boldsymbol{d}_{k_0}\|}\right\}\right)$$
$$= \Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta_0 \boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|} \in C^{\|\cdot\|}_{m+h+1} : \ k' \in \{0,\dots,h\} \cup \{(0,0)\}\right\}\right)$$

Given that $\|\boldsymbol{d}_{k'}\| = |\rho|^{k'}$, for any $1 \le k' \le h$,

$$\begin{aligned}
\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta' \boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|} \in C^{\|\cdot\|}_{m+h+1} : \ \frac{\vartheta' f(\boldsymbol{d}_{k'})}{\|\boldsymbol{d}_{k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{k_0})}{\|\boldsymbol{d}_{k_0}\|}\right\}\right) & \\
= w_{\vartheta_0} + w_{\vartheta_0}\left[\sum_{k'=1}^{h-1} \|\boldsymbol{d}_{k'}\|^{\alpha} + \frac{\|\boldsymbol{d}_h\|^{\alpha}}{1-|\rho|^{\alpha}}\right] & \\
= w_{\vartheta_0}\left[1 + \sum_{k'=1}^{h-1} |\rho|^{\alpha k'} + \frac{|\rho|^{\alpha h}}{1-|\rho|^{\alpha}}\right] & \\
= w_{\vartheta_0}\left[\frac{1-|\rho|^{\alpha h}}{1-|\rho|^{\alpha}} + \frac{|\rho|^{\alpha h}}{1-|\rho|^{\alpha}}\right] & \\
= w_{\vartheta_0}\frac{1}{1-|\rho|^{\alpha}}. &
\end{aligned}$$

Similarly, by (28),

$$\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta' \boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|} \in A_{\vartheta,k} : \ \frac{\vartheta' f(\boldsymbol{d}_{k'})}{\|\boldsymbol{d}_{k'}\|} \in V_0\right\}\right)$$
$$= \Gamma^{\|\cdot\|}\left(A_{\vartheta,k} \cap \left\{\frac{\vartheta_0 \boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|} \in C^{\|\cdot\|}_{m+h+1} : \ k' \in \{0,\dots,h\} \cup \{(0,0)\}\right\}\right)$$

$$= \begin{cases} \Gamma^{\|\cdot\|}\left(\left\{\dfrac{\vartheta_0 \boldsymbol{d}_k}{\|\boldsymbol{d}_k\|}\right\}\right), & \text{if } \vartheta = \vartheta_0, \\ \Gamma^{\|\cdot\|}(\emptyset), & \text{if } \vartheta \neq \vartheta_0, \end{cases}$$

$$= \begin{cases} w_{j\vartheta_0}\delta_{\{\vartheta_0\}}(\vartheta), & \text{if } \ k = 0, \\ w_{\vartheta_0}|\rho|^{\alpha k}\delta_{\{\vartheta_0\}}(\vartheta), & \text{if } \ 1 \le k \le h-1, \\ w_{\vartheta_0}\dfrac{|\rho|^{\alpha h}}{1-|\rho|^{\alpha}}\delta_{\{\vartheta_0\}}(\vartheta), & \text{if } \ k = h. \end{cases}$$

The conclusion follows.

E.10 Proof of Proposition 6

Lemma 4 *Let X_t be the α-stable anticipative AR(2) as in (29) with real characteristic roots (resp. the anticipative fractionally integrated AR as in (31)). With f as in (22), and for any*

$m \geq 1$, $h \geq 0$,

$$\forall k, \ell \geq -m, \ \forall \vartheta_1, \vartheta_2 \in S_1, \quad \left[\frac{f(\vartheta_1 \boldsymbol{d}_k)}{\|\boldsymbol{d}_k\|} = \frac{f(\vartheta_2 \boldsymbol{d}_\ell)}{\|\boldsymbol{d}_\ell\|} \implies k = \ell \ \textit{and} \ \vartheta_1 = \vartheta_2 \right].$$

Proof.

The result is clear for both processes for $-m \leq k, \ell \leq -1$. For both processes, $d_j \neq 0$ for every $j \geq 0$. For the AR(2) with real characteristic roots and $\lambda_1 \neq \lambda_2$, $d_j = 0$ would force $(\lambda_1/\lambda_2)^{j+1} = 1$, which is impossible since λ_1/λ_2 is real and differs from 1 and -1 by $\lambda_1 \neq \lambda_2$ and $\lambda_1 + \lambda_2 \neq 0$; when $\lambda_1 = \lambda_2 = \lambda$, $d_j = (j+1)\lambda^j \neq 0$. For the fractionally integrated AR, d_j is a ratio of Gamma values, none of which vanishes. Two consequences follow. First, for $k \geq 0$ and $\ell \in \{-m, \ldots, -1\}$ (or vice versa), the last entry of $f(\boldsymbol{d}_\ell)$, namely d_ℓ, vanishes while every entry of $f(\boldsymbol{d}_k)$ is nonzero, so the two normalized shapes differ. Second, the ratio chain below is well defined. It thus remains to treat the case $k, \ell \geq 0$:

$$\begin{aligned} \frac{f(\vartheta_1 \boldsymbol{d}_k)}{\|\boldsymbol{d}_k\|} = \frac{f(\vartheta_2 \boldsymbol{d}_\ell)}{\|\boldsymbol{d}_\ell\|} &\iff \left[\forall\, i = 0, \ldots, m, \quad \frac{\vartheta_1 d_{k+i}}{\|\boldsymbol{d}_k\|} = \frac{\vartheta_2 d_{\ell+i}}{\|\boldsymbol{d}_\ell\|} \right] \\ &\iff \frac{d_k}{d_\ell} = \frac{d_{k+1}}{d_{\ell+1}} = \ldots = \vartheta_1 \vartheta_2 \frac{\|\boldsymbol{d}_k\|}{\|\boldsymbol{d}_\ell\|}. \end{aligned} \tag{67}$$

The last statement in particular implies that $\dfrac{d_k}{d_\ell} = \dfrac{d_{k+1}}{d_{\ell+1}}$.

For the anticipative AR(2), if $\lambda_1 \neq \lambda_2$, we then have

$$\begin{aligned} \frac{d_k}{d_\ell} = \frac{d_{k+1}}{d_{\ell+1}} &\iff \frac{\lambda_1^{k+1} - \lambda_2^{k+1}}{\lambda_1^{\ell+1} - \lambda_2^{\ell+1}} = \frac{\lambda_1^{k+2} - \lambda_2^{k+2}}{\lambda_1^{\ell+2} - \lambda_2^{\ell+2}} \\ &\iff \left(\lambda_1/\lambda_2\right)^{k-\ell} = 1 \iff k = \ell, \end{aligned}$$

the last equivalence holding since λ_1/λ_2 is real and, by $\lambda_1 \neq \lambda_2$ and $\lambda_1 + \lambda_2 \neq 0$, differs from 1 and -1, hence has modulus different from 1. The case $\lambda_1 = \lambda_2 = \lambda$ is similar. For the anticipative fractionally integrated AR, given that $\Gamma(z+1) = z\Gamma(z)$ for any $z \in \mathbb{C}$, we have

$$\begin{aligned} \frac{d_k}{d_\ell} = \frac{d_{k+1}}{d_{\ell+1}} &\iff \frac{\Gamma(k+d)\Gamma(\ell+1)}{\Gamma(\ell+d)\Gamma(k+1)} = \frac{\Gamma(k+d+1)\Gamma(\ell+2)}{\Gamma(\ell+d+1)\Gamma(k+2)} \\ &\iff \frac{\Gamma(\ell+d+1)\Gamma(k+2)}{\Gamma(\ell+d)\Gamma(k+1)} = \frac{\Gamma(k+d+1)\Gamma(\ell+2)}{\Gamma(k+d)\Gamma(\ell+1)} \\ &\iff (k-\ell)(d-1) = 0 \\ &\iff k = \ell. \end{aligned}$$

Therefore, in all cases,

$$\frac{d_k}{d_\ell} = \frac{d_{k+1}}{d_{\ell+1}} = \ldots = \vartheta_1 \vartheta_2 \frac{\|\boldsymbol{d}_k\|}{\|\boldsymbol{d}_\ell\|} \implies k = \ell \ \text{and} \ \vartheta_1 \vartheta_2 = 1.$$

□

Let us now prove Proposition 6. The spectral measure of $\boldsymbol{X}_t$ writes

$$\Gamma^{\|\cdot\|} = \sigma^\alpha \sum_{\vartheta \in S_1} \sum_{k \in \mathbb{Z}} w_\vartheta \|\boldsymbol{d}_k\|^\alpha \delta_{\left\{\frac{\vartheta \boldsymbol{d}_k}{\|\boldsymbol{d}_k\|}\right\}},$$

where the sequences (d_k) are given respectively by (30) and (32) for the anticipative AR(2) and fractionally integrated processes. By Proposition 2,

$$\mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t}, A \Big| B(V_0)\Big) \underset{x\to\infty}{\longrightarrow} \frac{\Gamma^{\|\cdot\|}(A \cap B(V_0))}{\Gamma^{\|\cdot\|}(B(V_0))}.$$

On the one hand, we have by definition of $B(V_0)$, V_0 and Lemma 4,

$$\begin{aligned}
\Gamma^{\|\cdot\|}(B(V_0)) &= \Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta \boldsymbol{d}_k}{\|\boldsymbol{d}_k\|} \in B(V_0) :\ (\vartheta, k) \in \{-1,+1\} \times \mathbb{Z}\right\}\right) \\
&= \Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta \boldsymbol{d}_k}{\|\boldsymbol{d}_k\|} \in C_{m+h+1}^{\|\cdot\|} :\ \frac{\vartheta f(\boldsymbol{d}_k)}{\|\boldsymbol{d}_k\|} \in V_0,\ (\vartheta, k) \in \{-1,+1\} \times \mathbb{Z}\right\}\right) \\
&= \Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta \boldsymbol{d}_k}{\|\boldsymbol{d}_k\|} \in C_{m+h+1}^{\|\cdot\|} :\ \frac{\vartheta f(\boldsymbol{d}_k)}{\|\boldsymbol{d}_k\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{k_0})}{\|\boldsymbol{d}_{k_0}\|},\ (\vartheta, k) \in \{-1,+1\} \times \mathbb{Z}\right\}\right) \\
&= \Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta_0 \boldsymbol{d}_{k_0}}{\|\boldsymbol{d}_{k_0}\|}\right\}\right).
\end{aligned}$$

Similarly, it is easily shown that

$$\Gamma^{\|\cdot\|}(A \cap B(V_0)) = \Gamma^{\|\cdot\|}\left(A \cap \left\{\frac{\vartheta_0 \boldsymbol{d}_{k_0}}{\|\boldsymbol{d}_{k_0}\|}\right\}\right).$$

The conclusion follows.

E.11 Proof of Proposition 7

We prove the three parts in order: part (ι) establishes the representability gap (38); part ($\iota\iota$) identifies the excluded directions $\mathcal{D}$ as post-peak patterns and shows they are invisible to coordinate-magnitude conditioning; part ($\iota\iota\iota$) turns this gap into the asymptotic missed-mass formula (40).

Part (ι). Under the hypotheses $d_k = 0$ for $k \le -1$ and $d_k \neq 0$ for $k \ge 0$, one has $m_0 = 0$ in the sense of (18), so that any $m \ge 1$ satisfies $m \ge m_0$. The vanishing $(d_{k+m}, \dots, d_k) = \mathbf{0}$ then forces every index $k+m, \dots, k$ to be ≤ -1, i.e. $k+m \le -1$, so that $k \le -m-1$; in that case $d_\ell = 0$ is automatic for all $\ell \le k-1$, and (16) holds. Condition (17) concerns only the case $\alpha = 1$ and is not needed here, since $\alpha \neq 1$ by assumption. Hence $\boldsymbol{X_t}$ is representable on $C_{m+h+1}^{\|\cdot\|}$ for any seminorm $\|\cdot\|$ satisfying (10) (Proposition 3(ι)). For non-representability on $C_{m+h+1}^{\|\cdot\|_\star}$ we apply Proposition 1. By (15), the Euclidean spectral measure Γ of $\boldsymbol{X_t}$ charges the countable set

$\{\vartheta \boldsymbol{d}_k/\|\boldsymbol{d}_k\|_e : \vartheta \in S_1,\, k \in \mathbb{Z},\, \boldsymbol{d}_k \neq \boldsymbol{0}\}$ with weights $\sigma^\alpha w_\vartheta \|\boldsymbol{d}_k\|_e^\alpha$. The coordinate of index 0, in the seminorm (37) of $\boldsymbol{d}_k$ is d_k, so $\vartheta \boldsymbol{d}_k/\|\boldsymbol{d}_k\|_e \in K^{\|\cdot\|_\star}$ if and only if $d_k = 0$. Since (X_t) is purely anticipative, $d_k = 0$ for $k \leq -1$. For $k \in \{-m, \ldots, -1\}$ the window $\boldsymbol{d}_k = (d_{k+m}, \ldots, d_0, 0, \ldots, 0)$ contains $d_0 \neq 0$ and is therefore non-zero. Summing the contributions yields (38), which is strictly positive since $d_0 \neq 0$ and $m \geq 1$. By Proposition 1, $\boldsymbol{X_t}$ is not representable on $C_{m+h+1}^{\|\cdot\|_\star}$.

Part ($\iota\iota$). The characterization of $\mathcal{D}$ as post-peak patterns is immediate from the form of $\boldsymbol{d}_k$: for $k \in \{-m, \ldots, -1\}$, the entries of $\boldsymbol{d}_k$ at positions $i \in \{k+1, \ldots, h\}$ vanish because $(\boldsymbol{d}_k)_i = d_{k-i}$ with $k - i \leq -1$, hence zero by the purely anticipative assumption. The only potentially nonzero entries are in positions $-m, \ldots, k$, encoding past observations.

To establish (39), we use the single-big-jump representation (36) of the spectral tail process, valid under the regular variation induced by (11)–(13) (Mikosch and Wintenberger, 2024). Under conditioning on $|X_t| > x$, the rescaled vector $\boldsymbol{X_t}/|X_t|$ converges in distribution to

$$(\Theta_{-m}, \ldots, \Theta_h) \overset{d}{=} \Theta_\varepsilon \frac{\boldsymbol{d}_J}{|d_J|},$$

with $\Theta_\varepsilon \in \{-1, +1\}$ independent of J, and J is supported on $\{j \in \mathbb{Z} : d_j \neq 0\}$ by (36). For a purely anticipative process, this support reduces to $\{j \in \mathbb{Z} : j \geq 0\}$. Consequently, the limit in (35) is concentrated on the countable set

$$\Big\{\vartheta\, \boldsymbol{d}_j/|d_j| : \vartheta \in S_1,\, j \geq 0,\, d_j \neq 0\Big\}. \tag{68}$$

Normalizing by the Euclidean norm, the distributional limit of $\boldsymbol{X_t}/\|\boldsymbol{X_t}\|_e$ given $|X_t| > x$ is concentrated on $\{\vartheta \boldsymbol{d}_j/\|\boldsymbol{d}_j\|_e : \vartheta \in S_1,\, j \geq 0,\, d_j \neq 0\}$, none of whose points belongs to the closed set $\widetilde{A}_{\vartheta_0,k_0}$ by assumption; the Portmanteau theorem then yields (39).

Part ($\iota\iota\iota$). By Theorem 4.4.8 of Samorodnitsky and Taqqu (1994), $\mathbb{P}(\boldsymbol{X_t}/x \in \cdot)/\mathbb{P}(\|\boldsymbol{X_t}\|_e > x)$ converges vaguely on $\mathbb{R}^{m+h+1} \setminus \{\boldsymbol{0}\}$, as $x \to \infty$, to the measure μ given in polar coordinates $\boldsymbol{s} = r\boldsymbol{\theta}$, $r > 0$, $\boldsymbol{\theta} = (\theta_{-m}, \ldots, \theta_h) \in S_{m+h+1}$, by $\mu(dr\, d\boldsymbol{\theta}) = \alpha r^{-\alpha-1} dr\, \Gamma(d\boldsymbol{\theta})/\Gamma(S_{m+h+1})$. Fix $\epsilon > 0$ and let $E_\epsilon := \{\boldsymbol{s} \in \mathbb{R}^{m+h+1} : \|\boldsymbol{s}\|_e > 1,\ |s_0| \leq \epsilon\}$. This set is bounded away from $\boldsymbol{0}$, and $\partial E_\epsilon \subset \{\|\boldsymbol{s}\|_e = 1\} \cup \{|s_0| = \epsilon\}$ is μ-null: on each ray $\{r\boldsymbol{\theta} : r > 0\}$, either $\theta_0 = 0$ and the condition $|s_0| = \epsilon$ never holds, or $\theta_0 \neq 0$ and each of the two conditions selects a single value of r. Since $\{|X_t| \leq \epsilon x\} \cap \{\|\boldsymbol{X_t}\|_e > x\} = \{\boldsymbol{X_t}/x \in E_\epsilon\}$, the Portmanteau theorem yields

$$\begin{aligned}
\lim_{x\to\infty} \mathbb{P}\big(|X_t| \leq \epsilon x \,\big|\, \|\boldsymbol{X_t}\|_e > x\big) &= \mu(E_\epsilon) \\
&= \frac{1}{\Gamma(S_{m+h+1})} \int_{S_{m+h+1}} \int_1^\infty \mathbf{1}\{r|\theta_0| \leq \epsilon\}\, \alpha r^{-\alpha-1} dr\, \Gamma(d\boldsymbol{\theta}) \\
&= \frac{1}{\Gamma(S_{m+h+1})} \int_{S_{m+h+1}} \Big(1 - (|\theta_0|/\epsilon)^\alpha\Big)_+ \Gamma(d\boldsymbol{\theta}),
\end{aligned}$$

which is the first equality in (40). For $\boldsymbol{\theta}$ with $\theta_0 = 0$ the integrand equals 1 for every $\epsilon > 0$, whereas for $\theta_0 \neq 0$ the map $\epsilon \mapsto (1 - (|\theta_0|/\epsilon)^\alpha)_+$ is nondecreasing and vanishes as soon as

$\epsilon \le |\theta_0|$; as $\epsilon \to 0$, the integrand therefore decreases pointwise to $\mathbf{1}\{\theta_0 = 0\}$. It is bounded by 1 and Γ is finite, so dominated convergence yields

$$\lim_{\epsilon\to 0}\lim_{x\to\infty} \mathbb{P}(|X_t| \le \epsilon x \,|\, \|\boldsymbol{X_t}\|_e > x) = \frac{\Gamma(\{\boldsymbol{\theta} \in S_{m+h+1} : \theta_0 = 0\})}{\Gamma(S_{m+h+1})} = \frac{\Gamma(K^{\|\cdot\|_\star})}{\Gamma(S_{m+h+1})},$$

the last equality holding by (37), which gives $\|\boldsymbol{\theta}\|_\star = |\theta_0|$. It remains to evaluate this ratio. By (15), $\Gamma(S_{m+h+1}) = \sigma^\alpha \sum_{\vartheta\in S_1}\sum_{k\in\mathbb{Z}} w_\vartheta \|\boldsymbol{d}_k\|_e^\alpha$, while $\Gamma(K^{\|\cdot\|_\star})$ is given by (38). Since $w_{-1}+w_{+1} = 1$, the sum over ϑ reduces to $\sum_{\vartheta\in S_1} w_\vartheta \|\boldsymbol{d}_k\|_e^\alpha = \|\boldsymbol{d}_k\|_e^\alpha$ in both expressions, and the common factor σ^α cancels, which gives the closed form in (40). Finally, the numerator $\sum_{k=-m}^{-1} \|\boldsymbol{d}_k\|_e^\alpha$ is positive by part (ι), and the denominator exceeds it strictly since it also contains the term $k = 0$, with $\|\boldsymbol{d}_0\|_e > 0$ because $d_0 \neq 0$; the ratio therefore lies in $(0, 1)$.

E.12 Proof of Proposition 8

Throughout, N is a fixed norm on $\mathbb{R}^{m+1}$, f is the past projection (22), and we write $\|\boldsymbol{s}\|_N := N(f(\boldsymbol{s}))$ for $\boldsymbol{s} \in \mathbb{R}^{m+h+1}$; by equivalence of norms on $\mathbb{R}^{m+1}$, there is a constant $c_N > 0$ such that $\|\boldsymbol{s}\|_N \le c_N\|\boldsymbol{s}\|_e$ for all $\boldsymbol{s}$. We first check that the identities of Kulik and Soulier (2020) apply to (X_t) and identify the resulting tail measure of the sequence, in the representation (69)–(70) below; we then project this measure onto the observation window $\{-m, \dots, h\}$, inject the single-big-jump representation to make the random normalization cancel, and finally reindex and integrate to identify the limit as Γ^N.

Under (11)–(12), $\sum_{k\in\mathbb{Z}} |d_k|^s < \infty$ for some $s \in (0, \alpha) \cap (0, 1]$, hence $d_k \to 0$ as $|k| \to \infty$ and $\sum_k |d_k|^\alpha < \infty$. By Mikosch and Wintenberger (2024) and (36), the tail process Y of (X_t) admits the representation

$$Y_i \;=\; |Y_0|\,\Theta_\varepsilon\,\frac{d_{J-i}}{|d_J|},\quad i \in \mathbb{Z}, \tag{69}$$

where $|Y_0| \sim \text{Pareto}(\alpha)$ on $(1, \infty)$, $\Theta_\varepsilon \in S_1$ satisfies $\mathbb{P}(\Theta_\varepsilon = \vartheta) = w_\vartheta$, $J \in \mathbb{Z}$ has distribution $\mathbb{P}(J = j) = |d_j|^\alpha / \sum_{\ell\in\mathbb{Z}} |d_\ell|^\alpha$ (supported by $\{j : d_j \neq 0\}$), and $|Y_0|, \Theta_\varepsilon, J$ are mutually independent. Since $d_{J-i} \to 0$ as $|i| \to \infty$, $Y \in \ell_0$ almost surely, so that $\mathbb{P}(\mathcal{E}(Y) < \infty) = 1$, where $\mathcal{E}(Y) = \#\{j : |Y_j| > 1\}$; as $|Y_0| > 1$, the first-exceedance anchor $A_1(Y) = \inf\{j : |Y_j| > 1\}$ belongs to $\mathbb{Z}$ almost surely, which implies condition (5.5.2) of Kulik and Soulier (2020) for A_1. Corollary 5.5.4 of Kulik and Soulier (2020) therefore applies, so that $\theta_e := \mathbb{P}(A_1(Y) = 0)$ is the candidate extremal index of (X_t), and identity (5.5.7) there yields

$$\nu \;=\; \theta_e \sum_{j\in\mathbb{Z}} \int_0^\infty \mathbb{E}\big[\delta_{r(Y^*)^{-1}B^jY} \,\big|\, A_1(Y) = 0\big]\, \alpha r^{-\alpha-1}\, dr, \tag{70}$$

where ν is the tail measure of (X_t) on $\mathbb{R}^{\mathbb{Z}}$, $Y^* := \sup_{i\in\mathbb{Z}} |Y_i|$ and B is the backshift operator, $(By)_k = y_{k-1}$. Moreover $\theta_e > 0$: by (69), $|Y_j| = |Y_0||d_{J-j}|/|d_J|$ for $j < 0$, so that

$$\{A_1(Y) = 0\} = \Big\{|Y_0| \sup_{k>J} |d_k| \le |d_J|\Big\}, \tag{71}$$

and since $d_k \to 0$ there exist indices j with $d_j \neq 0$ and $|d_j| > \sup_{k>j} |d_k|$, it suffices to take the largest index at which $|d_k|$ attains its maximum, for which $\mathbb{P}(J = j,\ |Y_0| \le |d_j|/\sup_{k>j} |d_k|) > 0$, with the convention $|d_j|/0 := +\infty$, the event then being $\{J = j\}$. In fact, by (71) and a telescoping sum, $\theta_e = (D^*)^\alpha / \sum_{k\in\mathbb{Z}} |d_k|^\alpha$, with $D^* := \sup_k |d_k|$ as below, in accordance with Example 3.1 of Basrak and Planinić (2021); the constant $\theta_e (D^*)^{-\alpha}$ appearing below is therefore $1/\sum_k |d_k|^\alpha$.

Let $\pi : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{m+h+1}$, $\pi(y) = (y_{-m}, \dots, y_h)$, and $\widetilde{\nu} := \nu \circ \pi^{-1}$. Recall that ν is shift-invariant, α-homogeneous and normalized by $\nu(\{y : |y_0| > 1\}) = 1$, so that $\nu(\{y : |y_i| > \epsilon\}) = \epsilon^{-\alpha}$ for every $i \in \mathbb{Z}$ and every $\epsilon > 0$. Let $\widetilde{C} \subset \mathbb{R}^{m+h+1}$ be Borel and bounded away from $\mathbf{0}$, say $\|\boldsymbol{s}\|_e > \delta$ for every $\boldsymbol{s} \in \widetilde{C}$, and set $\epsilon := \delta/\sqrt{m+h+1}$. Every $\boldsymbol{s} \in \widetilde{C}$ then satisfies $\max_{-m\le i\le h} |s_i| \ge \|\boldsymbol{s}\|_e/\sqrt{m+h+1} > \epsilon$, so that $\pi^{-1}(\widetilde{C}) \subset \bigcup_{i=-m}^{h} \{y : |y_i| > \epsilon\}$ and, by subadditivity, $\widetilde{\nu}(\widetilde{C}) \le \sum_{i=-m}^{h} \nu(\{y : |y_i| > \epsilon\}) = (m+h+1)\epsilon^{-\alpha} < \infty$; and (70) yields

$$\widetilde{\nu}(\widetilde{C}) \;=\; \theta_e \sum_{j\in\mathbb{Z}} \int_0^\infty \mathbb{P}\Big(r\,(Y^*)^{-1}\pi(B^j Y) \in \widetilde{C} \,\Big|\, A_1(Y) = 0\Big)\, \alpha r^{-\alpha-1}\, dr. \tag{72}$$

By the multivariate regular variation of the finite-dimensional distributions of (X_t), $\widetilde{\nu}$ is the vague limit of $\mathbb{P}(\boldsymbol{X_t}/x \in \cdot)/\mathbb{P}(|X_t| > x)$ on $\mathbb{R}^{m+h+1} \setminus \{\mathbf{0}\}$ as $x \to \infty$.

Set $D^* := \sup_{k\in\mathbb{Z}} |d_k|$, which is finite since $d_k \to 0$ and strictly positive since $(d_k) \neq 0$. With the convention $(\boldsymbol{d}_k)_i = d_{k-i}$ for $i \in \{-m, \dots, h\}$, for every $j \in \mathbb{Z}$,

$$\big(\pi(B^j Y)\big)_i \;=\; Y_{i-j} \;=\; |Y_0|\,\Theta_\varepsilon \frac{d_{J-(i-j)}}{|d_J|} \;=\; \frac{|Y_0|\,\Theta_\varepsilon}{|d_J|}\,(\boldsymbol{d}_{J+j})_i,$$

and $Y^* = |Y_0|\,D^*/|d_J|$. Therefore $|Y_0|$ and $|d_J|$ cancel:

$$r\,(Y^*)^{-1}\pi(B^j Y) \;=\; \frac{r\,\Theta_\varepsilon}{D^*}\,\boldsymbol{d}_{J+j},\ r > 0,\ j \in \mathbb{Z}. \tag{73}$$

Define

$$\widetilde{A} := \{\boldsymbol{s} \in \mathbb{R}^{m+h+1} : \|\boldsymbol{s}\|_N > 1,\ \boldsymbol{s}/\|\boldsymbol{s}\|_N \in A \cap B(V)\},\ \widetilde{B} := \{\boldsymbol{s} \in \mathbb{R}^{m+h+1} : \|\boldsymbol{s}\|_N > 1,\ \boldsymbol{s}/\|\boldsymbol{s}\|_N \in B(V)\},$$

both bounded away from $\mathbf{0}$ since $\|\boldsymbol{s}\|_N \le c_N \|\boldsymbol{s}\|_e$. For fixed $(\Theta_\varepsilon, J, j)$ with $\overleftarrow{\boldsymbol{d}_{J+j}} \neq \mathbf{0}$, by (73) and the homogeneity of $\boldsymbol{s} \mapsto \boldsymbol{s}/\|\boldsymbol{s}\|_N$, the event $\{r\Theta_\varepsilon \boldsymbol{d}_{J+j}/D^* \in \widetilde{A}\}$ equals

$$\Big\{r > \frac{D^*}{N(\overleftarrow{\boldsymbol{d}_{J+j}})}\Big\} \cap \Big\{\frac{\Theta_\varepsilon \boldsymbol{d}_{J+j}}{N(\overleftarrow{\boldsymbol{d}_{J+j}})} \in A \cap B(V)\Big\},$$

since $\|r\Theta_\varepsilon \boldsymbol{d}_{J+j}/D^*\|_N = r\,N(\overleftarrow{\boldsymbol{d}_{J+j}})/D^*$; it is empty when $\overleftarrow{\boldsymbol{d}_{J+j}} = \boldsymbol{0}$. Integration with respect to $\alpha r^{-\alpha-1}dr$ gives $\big(N(\overleftarrow{\boldsymbol{d}_{J+j}})/D^*\big)^\alpha$. All the quantities involved being nonnegative, Tonelli's theorem allows this radial integration to be carried out inside the conditional expectation, so that the summand of index j in (72), taken at $\widetilde{C} = \widetilde{A}$, equals

$$(D^*)^{-\alpha}\,\mathbb{E}\left[N(\overleftarrow{\boldsymbol{d}_{J+j}})^\alpha\,\mathbf{1}\Big\{\frac{\Theta_\varepsilon \boldsymbol{d}_{J+j}}{N(\overleftarrow{\boldsymbol{d}_{J+j}})} \in A\cap B(V)\Big\}\,\Big|\,A_1(Y)=0\right].$$

Summing over $j\in\mathbb{Z}$ and exchanging, again by nonnegativity, the sum with the conditional expectation, it remains to observe that for each fixed value of J the map $j\mapsto \ell := J+j$ is a bijection of $\mathbb{Z}$ onto itself, so that the summation range becomes $\ell\in\mathbb{Z}$ and no longer depends on J. Therefore (72) becomes

$$\widetilde{\nu}(\widetilde{A}) \;=\; \theta_e\,(D^*)^{-\alpha}\,\mathbb{E}\left[\sum_{\ell\in\mathbb{Z}} N(\overleftarrow{\boldsymbol{d}_\ell})^\alpha\,\mathbf{1}\Big\{\frac{\Theta_\varepsilon \boldsymbol{d}_\ell}{N(\overleftarrow{\boldsymbol{d}_\ell})} \in A\cap B(V)\Big\}\,\Big|\,A_1(Y)=0\right], \tag{74}$$

with the convention that the summand vanishes when $\overleftarrow{\boldsymbol{d}_\ell} = \boldsymbol{0}$. By (71), the conditioning event $\{A_1(Y)=0\}$ is a function of $(|Y_0|, J)$ alone, hence independent of Θ_ε, while the sum in (74) is a deterministic function of Θ_ε alone. The conditional expectation therefore reduces to the expectation with respect to the law of Θ_ε, that is, to $\sum_{\vartheta\in S_1} w_\vartheta \sum_{\ell\in\mathbb{Z}} N(\overleftarrow{\boldsymbol{d}_\ell})^\alpha \mathbf{1}\{\vartheta \boldsymbol{d}_\ell/N(\overleftarrow{\boldsymbol{d}_\ell}) \in A\cap B(V)\}$, which is $\Gamma^N(A\cap B(V))$ by the definition (44). Therefore

$$\widetilde{\nu}(\widetilde{A}) = \theta_e (D^*)^{-\alpha}\,\Gamma^N(A\cap B(V)), \qquad \widetilde{\nu}(\widetilde{B}) = \theta_e (D^*)^{-\alpha}\,\Gamma^N(B(V)),$$

with Γ^N as in (44), the second identity being obtained by the same computation with $\widetilde{B}$ in place of $\widetilde{A}$. The measure Γ^N is finite. Indeed, by equivalence of norms on $\mathbb{R}^{m+1}$ there is $c'_N > 0$ such that $N(\boldsymbol{u}) \le c'_N \sum_{i=0}^m |u_i|$, while $\big(\sum_{i=0}^m |u_i|\big)^\alpha \le \sum_{i=0}^m |u_i|^\alpha$ when $\alpha\le 1$, by subadditivity of $x\mapsto x^\alpha$ on $[0,\infty)$, and $\big(\sum_{i=0}^m |u_i|\big)^\alpha \le (m+1)^{\alpha-1}\sum_{i=0}^m |u_i|^\alpha$ when $\alpha>1$, by the power mean inequality. In both cases $N(\overleftarrow{\boldsymbol{d}_\ell})^\alpha \le c' \sum_{i=0}^m |d_{\ell+i}|^\alpha$ with $c' = c'(N,m,\alpha)$, so that $\sum_\ell N(\overleftarrow{\boldsymbol{d}_\ell})^\alpha \le c'(m+1)\sum_k |d_k|^\alpha < \infty$. Moreover, by (72) and (73), $\widetilde{\nu}$ is concentrated on the countable union of rays $\{r\vartheta \boldsymbol{d}_\ell : r>0\}$, $\boldsymbol{d}_\ell \neq \boldsymbol{0}$, with radial density $\alpha r^{-\alpha-1}dr$. The rays with $\overleftarrow{\boldsymbol{d}_\ell} = \boldsymbol{0}$ do not meet $\{\|\boldsymbol{s}\|_N = 1\}$, since $\|\cdot\|_N$ vanishes identically on them, while on the remaining ones $\{\|\boldsymbol{s}\|_N = 1\}$ selects a single radius; hence $\widetilde{\nu}(\{\|\boldsymbol{s}\|_N = 1\}) = 0$, and since $\partial\widetilde{A} \subset \{\|\boldsymbol{s}\|_N = 1\} \cup \{\boldsymbol{s} : \|\boldsymbol{s}\|_N \ge 1,\ \boldsymbol{s}/\|\boldsymbol{s}\|_N \in \partial(A\cap B(V))\}$ by continuity of $\boldsymbol{s}\mapsto \boldsymbol{s}/\|\boldsymbol{s}\|_N$ from $\{\|\boldsymbol{s}\|_N>0\}$ onto C^N, with the analogous inclusion for $\widetilde{B}$, the assumptions $\Gamma^N(\partial(A\cap B(V))) = \Gamma^N(\partial B(V)) = 0$ imply $\widetilde{\nu}(\partial\widetilde{A}) = \widetilde{\nu}(\partial\widetilde{B}) = 0$. By homogeneity, the conditional probability in (43) equals $\mathbb{P}(\boldsymbol{X_t}/x\in\widetilde{A})/\mathbb{P}(\boldsymbol{X_t}/x\in\widetilde{B})$; the vague convergence stated after (72) applied to these two continuity sets, with $\widetilde{\nu}(\widetilde{B})>0$ by the assumption $\Gamma^N(B(V))>0$, yields

$$\lim_{x\to\infty} \mathbb{P}\Big(\frac{\boldsymbol{X_t}}{N(\overleftarrow{\boldsymbol{X_t}})} \in A\,\Big|\,\frac{\overleftarrow{\boldsymbol{X_t}}}{N(\overleftarrow{\boldsymbol{X_t}})} \in V,\ N(\overleftarrow{\boldsymbol{X_t}}) > x\Big) = \frac{\widetilde{\nu}(\widetilde{A})}{\widetilde{\nu}(\widetilde{B})} = \frac{\Gamma^N(A\cap B(V))}{\Gamma^N(B(V))},$$

the factors $\theta_e(D^*)^{-\alpha}$ cancelling. This proves claims (ι) and $(\iota\iota)$; the set $\{\boldsymbol{s} : f(\boldsymbol{s}) = \mathbf{0}\}$ has played no role, since $\widetilde{A}$ and $\widetilde{B}$ are contained in $\{\|\boldsymbol{s}\|_N > 1\}$.

For claim $(\iota\iota\iota)$, assume that $\|\boldsymbol{s}\| = N(f(\boldsymbol{s}))$ satisfies (10) and that (X_t) is (m,h)-past-representable. Then $\|\boldsymbol{d}_\ell\| = N(\overleftarrow{\boldsymbol{d}_\ell})$ for every ℓ, and by (16) the indices with $\boldsymbol{d}_\ell \neq \mathbf{0}$ coincide with those with $\overleftarrow{\boldsymbol{d}_\ell} \neq \mathbf{0}$, so that (23) reads $\Gamma^{\|\cdot\|} = \sigma^\alpha \Gamma^N$; the ratio above is then the right-hand side of (24). The Dirac collapse for the anticipative AR(2) with real characteristic roots and for the anticipative FWN follows from Lemma 4 and Proposition 6: with $V = V_0$ as in (27), the set of charged indices (ϑ', ℓ) with $\vartheta' \overleftarrow{\boldsymbol{d}_\ell} / N(\overleftarrow{\boldsymbol{d}_\ell}) \in V_0$ reduces to the singleton $\{(\vartheta_0, k_0)\}$, hence $\Gamma^N(B(V_0)) = w_{\vartheta_0} N(\overleftarrow{\boldsymbol{d}_{k_0}})^\alpha$ and, by (28), $\Gamma^N(A_{\vartheta,k} \cap B(V_0)) = w_{\vartheta_0} N(\overleftarrow{\boldsymbol{d}_{k_0}})^\alpha \, \mathbf{1}\{(\vartheta, k) = (\vartheta_0, k_0)\}$, which is the asserted collapse.

References

Andrews, B., Calder, M., Davis, R., 2009. Maximum likelihood estimation for α-stable autoregressive process. The Annals of Statistics 37, 1946–1982.

Ausloos, M., Ivanova, K., 2001. Power-law correlations in the Southern-Oscillation-Index fluctuations characterizing El Niño. Physical Review E 63, 047201.

Basrak, B., Planinić, H., 2021. Compound poisson approximation for regularly varying fields with application to sequence alignment. Bernoulli 27, 1371 – 1408.

Basrak, B., Segers, J., 2009. Regularly varying multivariate time series. Stochastic processes and their applications 119, 1055–1080.

Cavaliere, G., Nielsen, H.B., Rahbek, A., 2020. Bootstrapping noncausal autoregressions: With applications to explosive bubble modeling. Journal of Business & Economic Statistics 38, 55–67.

Davis, R.A., Resnick, S., 1985. Limit theory for moving averages of random variables with regularly varying tail probabilities. The Annals of Probability 13, 179–195.

Davis, R.A., Resnick, S., 1986. Limit theory for the sample covariance and correlation functions of moving averages. The Annals of Statistics 14, 533–558.

Dumitrescu, E., Peignon, J., Thomas, A., 2026. Tail-Aware Density Forecasting of Locally Explosive Time Series: A Neural Network Approach. arXiv preprint arXiv:2601.14049 .

Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events for Insurance and Finance. Springer, Berlin.

Fries, S., 2022. Conditional moments of noncausal alpha-stable processes and the prediction of bubble crash odds. Journal of Business & Economic Statistics 40, 1596–1616.

Fries, S., Zakoian, J.M., 2019. Mixed causal-noncausal ar processes and the modelling of explosive bubbles. Econometric Theory 35, 1234–1270.

Giancaterini, F., 2023. Essays on univariate and multivariate noncausal processes. Ph.D. thesis. Maastricht University. Netherlands. doi:10.26481/dis.20231204fg.

Gourieroux, C., Jasiak, J., 2023. Generalized covariance estimator. Journal of Business & Economic Statistics 41, 1315–1327.

Gourieroux, C., Jasiak, J., 2026. Nonlinear fore(back)casting and innovation filtering for causal–noncausal var models. Journal of Financial Econometrics 24, nbag005. doi:10.1093/jjfinec/nbag005.

Gouriéroux, C., Jasiak, J., 2016. Filtering, prediction and simulation methods for noncausal processes. Journal of Time Series Analysis 37, 405–430.

Gouriéroux, C., Lu, Y., Robert, C.Y., 2025. The causal-noncausal tail processes: An introduction. arXiv:2506.04046.

Gouriéroux, C., Zakoian, J.M., 2015. On uniqueness of moving average representations of heavy-tailed stationary processes. Journal of Time Series Analysis 36, 876–887.

Hecq, A., Issler, J.V., Telg, S., 2020. Mixed causal–noncausal autoregressions with exogenous regressors. Journal of Applied Econometrics 35, 328–343.

Hecq, A., Telg, S., Lieb, L., 2017. Simulation, estimation and selection of mixed causal-noncausal autoregressive models: The marx package. SSRN .

Hecq, A., Voisin, E., 2021. Forecasting bubbles with mixed causal-noncausal autoregressive models. Econometrics and Statistics 20, 29–45.

Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. The Annals of Statistics 3, 1163–1174.

Jasiak, J., Manafi Neyazi, A., 2026. GCov-based portmanteau test. Econometric Reviews 45, 996–1024. doi:10.1080/07474938.2026.2643749.

Knight, K., 1989. Consistency of Akaike's information criterion for infinite variance autoregressive processes. The Annals of Statistics 17, 824–840.

Kulik, R., Soulier, P., 2020. Heavy-tailed time series. Springer.

Lanne, M., Luoto, J., Saikkonen, P., 2012. Optimal forecasting of noncausal autoregressive time series. International Journal of Forecasting 28, 623–631.

Lanne, M., Saikkonen, P., 2011. Noncausal autogressions for economic time series. Journal of Time Series Econometrics 3.

Last, G., 2023. Tail processes and tail measures: An approach via Palm calculus. Extremes 26, 715–746.

Mikosch, T., Wintenberger, O., 2024. Extreme Value Theory for Time Series: Models with Power-Law Tails. Springer Series in Operations Research and Financial Engineering, Springer, Cham. doi:10.1007/978-3-031-59156-3.

Nolan, J., 2021. Univariate Stable Distributions: Models for Heavy Tailed Data. Springer Series in Operations Research and Financial Engineering, Springer International Publishing.

Samorodnitsky, G., Taqqu, M.S., 1994. Stable non-Gaussian random processes. Chapman & Hall, London.

de Truchis, G., Thomas, A., 2026. Laurent series expansion for MA(∞) representation of mixed causal-noncausal autoregressive processes. Journal of Time Series Analysis , 1–7doi:10.1111/jtsa.70070.