Forecasting extreme trajectories using seminorm representations

Gilles De Truchis*
Department of Economics, Université d'Orléans
and
Sébastien Fries
Department of Econometrics and Data Science, Vrije Universiteit Amsterdam
and
Arthur Thomas
Department of Economics, Université Paris Dauphine - PSL

October 29, 2025

Abstract

We develop a new forecasting theory based on a novel spectral representation of α-stable random vectors on unit cylinders defined by seminorms. We prove necessary and sufficient conditions for such representations to exist. For two-sided α-stable moving averages $X_t = \sum_{k\in\mathbb{Z}} d_k \varepsilon_{t+k}$, we characterize when trajectory vectors of the form $\boldsymbol{X}_t = (X_{t-m}, \dots, X_{t+h})$ are representable on $\{\boldsymbol{s} \in \mathbb{R}^{m+h+1} : \quad \|\boldsymbol{s}\| = 1\}$ for $\|\cdot\|$ an appropriate seminorm. The key finding is that purely causal processes are inherently non-representable and that anticipativeness is required. For such non-causal and mixed causal processes, the tail behavior of normalized future paths has an explicit limit characterized by discrete spectral measures, thereby revealing that $\boldsymbol{X}_t/\|\boldsymbol{X}_t\|$ is necessarily collinear to some specific MA(∞) coefficients. For anticipative AR($p \geq 2$) and fractionally integrated processes, we further show that the tail conditional distribution is $\{0, 1\}$-valued, yielding deterministic future path prediction. We develop new procedures to forecast crash dates and reversal probabilities, and demonstrate through Monte Carlo simulations their finite-sample performances. As an empirical illustration, we estimate the reversal dates of El Niño and La Niña occurrences.

Keywords: Prediction, Stable random vectors, Spectral representation, Pattern identification

*We thank Guillaume Chevillon, Marie Bessec, Elena Dumistrescu, Christian Francq, Christian Gourieroux, Alain Hecq, Joann Jasiak, Frederik Krabbe, Sebastien Laurent, Yannick Le Pen, Gabriele Mingoli, Aryan Manafi Neyazi, Gilles Stupfler, Daniel Velasquez-Gavarira and Jean-Michel Zakoian for many useful suggestions and detailed comments, as well as Dauphine LEDa's Seminar, LEO Economic Seminar, Maastricht Econometric Seminars, York University Econometric Seminar and conference participants at the 39th Canadian Econometric Study Group Meeting (CESG 2024), 6th Quantitative Finance and Financial Econometrics (QFFE 2024) and 44th International Symposium on Forecasting (ISF), 18th International Joint Conference on Computational and Financial Econometrics (CFE) and Computational and Methodological Statistics (CMStatistics) for their insights. Corresponding authors: Arthur Thomas, arthur.thomas@dauphine.psl.eu

1 Introduction

Stochastic processes that depend on future values of an independent and identically distributed (i.i.d.) sequence, called anticipative processes, have recently attracted considerable attention in statistics and econometrics (Hencic & Gouriéroux 2015, Fries & Zakoian 2019, Gouriéroux & Zakoian 2017, Hecq et al. 2017*a*, Cavaliere et al. 2020, Gouriéroux & Jasiak 2017, Bec et al. 2020, Hecq et al. 2020, Fries 2022, Hecq & Voisin 2021, Blasques et al. 2025, Hecq & Velasquez-Gaviria 2025, Giancaterini et al. 2025). These processes are particularly useful for modeling financial bubbles, climate extremes and more generally nonlinear dynamics that are exhibiting heavy-tailed marginals and conditional heteroskedasticity, while preserving the simplicity of strictly stationary linear models.[1] Interestingly, they arise as solutions to rational expectations models with infinite variance innovations (Gourieroux et al. 2020), with applications in macroeconomics (Lanne & Saikkonen 2013, Chen et al. 2017, Gouriéroux et al. 2019, Chahrour & Jurado 2021) and climate modeling (Giancaterini et al. 2022). While estimation methods are well-developed for both univariate and multivariate settings (Andrews et al. 2009, Lanne & Saikkonen 2011, 2013, Hecq et al. 2016, Fries & Zakoian 2019, Hecq et al. 2017*b*, Cavaliere et al. 2020, Velasco & Lobato 2018, Hecq et al. 2020, Gouriéroux & Jasiak 2016, 2017, Gourieroux & Jasiak 2023, Jasiak & Neyazi 2025, Hecq & Velasquez-Gaviria 2025, Gouriéroux et al. 2025), the conditional distribution of future paths given past observations remains poorly understood. Closed-form formulas are unavailable, leading researchers to propose numerical approaches (Lanne et al. 2012, Lanne & Luoto 2016, Gourieroux, Hencic & Jasiak 2021, Gourieroux, Jasiak & Tong 2021). Recently, Gourieroux & Jasiak (2025) derived closed-form one-step-ahead predictive densities and developed Sampling Importance Resampling (SIR) methods for multi-step forecasting and nonlinear impulse response functions.[2] However, these algorithms often lack theoretical

[1] Fries & Zakoian (2019) shows that despite the i.i.d. nature of the error term, the conditional distribution in direct time is lighter-tailed than the error distribution, revealing ARCH effects in causal representations.

[2] Gourieroux & Jasiak (2025) extends Gouriéroux & Jasiak (2016)'s forecasting approach to multivariate mixed-causal (S)VAR, but their kernel-based conditioning methods remain computationally demanding.

guarantees and are computationally expensive for long horizons and they fail to capture dynamics during explosive episodes accurately (Hecq & Voisin 2021). An exception is the anticipative α-stable AR(1) process, for which partial results were obtained in Gouriéroux & Zakoian (2017) and later completed in Fries (2022). Even in the simplest case within the family of anticipative processes, future realizations exhibit a complex dependence on the observed past. This is reflected in the functional forms of the conditional moments, which are generally nonlinear (see Fries 2022). Interestingly, the dynamics of the anticipative stable AR(1) process simplify during extreme events, appearing to follow an explosive exponential path with a determined killing probability (see Fries 2022, Gouriéroux et al. 2025). This naturally raises the question of whether, and in what form, such behavior might occur in more general stable processes. For X_t a two-sided moving average, defined by

$$X_t = \sum_{k\in\mathbb{Z}} d_k \varepsilon_{t+k} \tag{1}$$

where (ε_t) is an i.i.d. α-stable sequence and (d_k) is a sequence of non-random coefficients, this paper examines the conditional distribution of future paths—specifically, $(X_{t+1}, \ldots, X_{t+h})$ given the observed trajectory $(X_{t-m}, \ldots, X_t)$ with $m \geq 0$ and $h \geq 1$ —when the process hugely deviates from its central values. Only mild summability conditions are assumed on the sequence (d_k), and in particular, we do not make any assumptions upfront about the anticipativeness or non-anticipativeness of (X_t). Under this general framework, any vector of the form $\boldsymbol{X_t} = (X_{t-m}, \ldots, X_{t+h})$ is multivariate α-stable and its distribution is characterised by a unique finite measure Γ on the Euclidean unit sphere $S_{m+h+1} = \{\boldsymbol{s} \in \mathbb{R}^{m+h+1} : \ \|\boldsymbol{s}\|_e = 1\}$, where $\|\cdot\|_e$ denotes the Euclidean norm (Theorem 2.3.1 in Samorodnitsky & Taqqu (1994)). The measure Γ completely describes the conditional distribution of the normalized paths $\boldsymbol{X_t}/\|\boldsymbol{X_t}\|_e$, which represents the "shape" of the trajectory, when $\boldsymbol{X_t}$ is large according to the Euclidean norm. A straightforward application of Theorem 4.4.8 by Samorodnitsky & Taqqu (1994) indeed shows that

$$\mathbb{P}\Big(\boldsymbol{X}_t/\|\boldsymbol{X}_t\|_e \in A \ \Big| \ \|\boldsymbol{X}_t\|_e > x, \quad \boldsymbol{X}_t/\|\boldsymbol{X}_t\|_e \in B\Big) \underset{x\longrightarrow\infty}{\longrightarrow} \frac{\Gamma(A\cap B)}{\Gamma(B)}, \tag{2}$$

for any appropriately chosen Borel sets $A, B \subset S_{m+h+1}$. As such, however, (2) is of limited value for prediction purposes, where only $X_{t-m}, \ldots, X_t$ are assumed to be observed. This is because, through the Euclidean norm of $\boldsymbol{X}_t$, the conditioning also depends on the future realizations $X_{t+1}, \ldots, X_{t+h}$. To overcome this critical issue, we investigate the conditional tail properties of $\boldsymbol{X}_t$ when the Euclidean norm is replaced by a seminorm $\|\cdot\|$ satisfying

$$\|(x_{-m}, \ldots, x_0, x_1, \ldots, x_h)\| = \|(x_{-m}, \ldots, x_0, 0, \ldots, 0)\|, \tag{3}$$

for any $(x_{-m}, \ldots, x_h) \in \mathbb{R}^{m+h+1}$.[3] More precisely, we seek the asymptotic behavior as $x \to \infty$ of equation (2) using a seminorm instead of a norm, namely

$$\mathbb{P}\left(\frac{\boldsymbol{X_t}}{\|\boldsymbol{X_t}\|} \in A \,\middle|\, \|\boldsymbol{X_t}\| > x,\ \frac{\boldsymbol{X_t}}{\|\boldsymbol{X_t}\|} \in B(V)\right), \quad B(V) = V \times \mathbb{R}^h \tag{4}$$

where $\|\boldsymbol{X_t}\| := \|(X_{t-m}, \ldots, X_t, 0, \ldots, 0)\|$, V describes a pattern in the past trajectory and A describes the future path region of interest. In this context, a new representation of stable random vectors on the "unit cylinder" $C_{m+h+1}^{\|\cdot\|} := \{\boldsymbol{s} \in \mathbb{R}^{m+h+1} : \ \|\boldsymbol{s}\| = 1\}$ is explored, where $\|\cdot\|$ is such a seminorm. Contrary to representations involving norms (see Theorem 2.3.8 in Samorodnitsky & Taqqu (1994)), not all stable random vectors can be characterized on unit cylinders, and a representation theorem is provided. We propose a new representation Theorem and its shows that $\boldsymbol{X_t}$ admits a representation by a measure $\Gamma^{\|\cdot\|}$ on $C_{m+h+1}^{\|\cdot\|}$ if and only if (X_t) is "anticipative enough". The equation (2) is then shown to hold with an adequate seminorm and with Γ (resp. S_{m+h+1}) replaced by $\Gamma^{\|\cdot\|}$ (resp. $C_{m+h+1}^{\|\cdot\|}$). The problem finally comes down to choosing the appropriate Borels B in (2) reflecting that only the past "shape" $(X_{t-m}, \ldots, X_t)/\|\boldsymbol{X_t}\|$ is observed.

The forecasting theory developed herein is linked to the conditional extreme value (CEV) literature (Davis & Hsing 1995, Konstantinides & Mikosch 2005, Meinguet & Segers 2010,

[3] A seminorm on $\mathbb{R}^d$ satisfies the same axioms as a norm (non-negativity, absolute homogeneity, and triangle inequality) except for the definiteness axiom. Specifically, for a seminorm $\|\cdot\|$, we may have $\|x\| = 0$ for some $x \neq 0$. The set $K^{\|\cdot\|} := \{x \in \mathbb{R}^d : \|x\| = 0\}$ is called the kernel of the seminorm. When $K^{\|\cdot\|} = \{0\}$, the seminorm is actually a norm.

Buraczewski et al. 2013, Janßen & Segers 2014, Basrak et al. 2016, Dombry et al. 2017, Janßen 2019, Planinić & Soulier 2017, Dombry et al. 2018, 2022) and, in particular, the concept of the spectral tail process, (Θ_t), introduced for time series by Basrak & Segers (2009).[4] When $\|\boldsymbol{X}_t\|_e$ is large, the limit distribution of the normalized vector $\boldsymbol{X}_t/\|\boldsymbol{X}_t\|_e$ can often be characterized by $Y\boldsymbol{\Theta}$, where the radial part Y represents the magnitude (following a Pareto(α) distribution) and the angular part (or spectral tail process) $\boldsymbol{\Theta} = (\Theta_t)$ captures the shape. Crucially, these limiting components Y and $\boldsymbol{\Theta}$ are independent. For strictly stationary univariate linear processes defined in (1), and as a consequence of the Single Big Jump (SBJ) heuristic (see e.g. Kulik & Soulier 2020), the spectral tail process Θ_t can be represented using two independent random components: $\Theta_t = \Theta_\varepsilon d_{J+t}|d_J|^{-1}$ (see e.g. Mikosch & Wintenberger 2024). Here, Θ_ε reflects the sign of the extreme event, and its distribution is linked to the extremal skewness of the innovation ε_t. J is the random index (or stochastic drift) measuring the time lag to the location of the SBJ, whose distribution crucially depends on the (d_k) coefficients and the tail parameter α. Drawing on various results from this literature, Gouriéroux et al. (2025) investigate the limit properties of mixed causal-noncausal autoregressive (MAR) processes under regularly varying tail assumptions. Their general framework differs from our α-stable, semi-norm based approach in at least three aspect. First, while their general regularly varying framework encompasses our α-stable assumption, they highlight that practical implementation within the CEV framework necessitates estimating key tail parameters (tail index and asymmetry), mirroring the parametric nature inherent in specific distributional assumptions like α-stable laws. Our focus on the α-stable case allows for an exact characterization of the conditional tail behavior. Second, Gouriéroux et al. (2025) primarily condition on large magnitudes ($|X_t| > x$) and discuss forecasting via online updates as new information arrives (e.g., $X_{t+1} > X_t$). In

[4]Another strand of this vast literature relates to the extremogram introduced by Davis & Mikosch (2009), defined as $\rho_E(h) = \lim_{u\to\infty} \mathbb{P}(X_{t+h} > u \mid X_t > u)$ for $h \geq 1$. However, this approach is arguably more suited for causal processes where extremes arise unpredictably from past innovations, whereas anticipative processes often exhibit extremes emerging through predictable trajectory patterns, enabling turning point forecasts.

contrast, our vector-based semi-norm approach conditions directly on the observed shape of the past trajectory $(X_{t-m}, \ldots, X_t)$ being close to a specific pattern, given its magnitude (measured by the semi-norm $\|\cdot\|$) is large. This richer initial conditioning, focused on trajectory patterns, potentially avoids the need for sequential updating for most forecasting goals. Third, our approach leverages the specific properties of the α-stable distribution combined with the semi-norm cylinder representation. This framework allows us to explicitly characterize the spectral measure $\Gamma^{\|\cdot\|}$ on the unit cylinder $C^{\|\cdot\|}_{m+h+1}$ as being inherently discrete, consisting of a weighted sum of Dirac masses located at points directly related to the normalized MA(∞) coefficient vectors (see Section 4). While the discreteness of the stochastic drift distribution is known in the general spectral tail process literature, our key finding is that for sufficiently anticipative processes (such as AR($p \geq 2$) or fractionally integrated processes), the conditional limit distribution under our semi-norm representation (i.e., given a specific past trajectory pattern) degenerates to a single Dirac mass (Proposition 6). This leads to asymptotically deterministic forecasts for the entire future path shape based solely on matching the observed past pattern.

Section 2 characterizes the representation of general α-stable vectors on seminorm unit cylinders and shows that (2) can be restated under this new representation. Focusing on α-stable processes defined by (1), Section 3 studies under which condition the vector $(X_{t-m}, \ldots, X_{t+h})$ admits a representation on the unit cylinder $C^{\|\cdot\|}_{m+h+1}$. The anticipativeness of (X_t) interestingly emerges as a necessary condition for such a representation to exist. Section 4 leverages (2) to analyze the tail conditional distribution of specific processes: the anticipative AR(1), AR(2), and the anticipative fractionally integrated process. For the last two processes, it appears possible to determine, in theory with certainty, the dates of extreme events, provided the current pattern is properly identified. Section 5 demonstrates the empirical relevance of our theoretical results through a climate forecasting exercise. Specifically, we predict the occurrences of El Niño and La Niña, as well as their reversal dates,

using the Southern Oscillation Index (SOI) data.[5] To replicate the numerical and empirical results presented in the paper and to demonstrate the generality of our approach, we have developed a web application. This application enables users to replicate our findings, explore examples of simulated trajectories, and apply our methods to other time series, particularly in the fields of macroeconomics, finance, and climate science.[6] Section 6 concludes the paper. Proofs, additional Monte Carlo simulations and complementary empirical results are provided in the Supplementary Material.

2 Representation of stable random vectors on unit cylinders

We start with a review of the characterization of stable random vectors on the Euclidean unit sphere. First, we recall that a random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ is defined to be a stable random vector in $\mathbb{R}^d$ if and only if, for any positive numbers A and B there is a positive number C and a deterministic vector $\boldsymbol{D} \in \mathbb{R}^d$ such that

$$A\boldsymbol{X}^{(1)} + B\boldsymbol{X}^{(2)} \stackrel{d}{=} C\boldsymbol{X} + \boldsymbol{D},$$

where $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$ are independent copies of $\boldsymbol{X}$. Furthermore, if $\boldsymbol{X}$ is stable, then there exists a real number $\alpha \in (0, 2]$ such that the above relation holds with $C = (A^\alpha + B^\alpha)^{1/\alpha}$, and $\boldsymbol{X}$ is referred to as α-stable. The Gaussian case ($\alpha = 2$) is henceforth excluded.

For, $\boldsymbol{X}$ an α-stable random vector with $0 < \alpha < 2$, one can define a unique pair $(\Gamma, \boldsymbol{\mu}^0)$, where Γ is a finite measure on S_d and $\boldsymbol{\mu}^0$ a non-random vector in $\mathbb{R}^d$, such that,

$$\mathbb{E}\left[e^{i\langle \boldsymbol{u}, \boldsymbol{X}\rangle}\right] = \exp\left\{ -\int_{S_d} |\langle \boldsymbol{u}, \boldsymbol{s}\rangle|^\alpha \Big(1 - i\,\mathrm{sign}(\langle \boldsymbol{u}, \boldsymbol{s}\rangle) w(\alpha, \langle \boldsymbol{u}, \boldsymbol{s}\rangle)\Big)\Gamma(d\boldsymbol{s}) + i\,\langle \boldsymbol{u}, \boldsymbol{\mu}^0\rangle \right\}, \quad \forall \boldsymbol{u} \in \mathbb{R}^d, \tag{5}$$

where $\langle \cdot, \cdot \rangle$ denotes the canonical scalar product, $w(\alpha, s) = \mathrm{tg}\left(\frac{\pi\alpha}{2}\right)$, if $\alpha \neq 1$, and $w(1, s) = -\frac{2}{\pi}\ln|s|$ otherwise, for $s \in \mathbb{R}$. The pair $(\Gamma, \boldsymbol{\mu}^0)$ is called the spectral representation of

[5] Data and methodology are available at: https://www.ncei.noaa.gov/access/monitoring/enso/soi

[6] The web application is available at the following link: https://marforecast.streamlit.app/?utm_medium=oembed

the stable vector $\boldsymbol{X}$, Γ is its spectral measure and $\boldsymbol{\mu}^0$ its shift vector. In particular, $\boldsymbol{X}$ is symmetric if and only if $\boldsymbol{\mu}^0 = 0$ and $\Gamma(A) = \Gamma(-A)$ for any Borel set A in S_d (Theorem 2.4.3 in Samorodnitsky & Taqqu (1994)), and in that case

$$\mathbb{E}\Big[e^{i\langle \boldsymbol{u}, \boldsymbol{X}\rangle}\Big] = \exp\left\{ -\int_{S_d} |\langle \boldsymbol{u}, \boldsymbol{s}\rangle|^{\alpha} \Gamma(d\boldsymbol{s}) \right\}, \quad \forall \boldsymbol{u} \in \mathbb{R}^d. \tag{6}$$

The representations (5) and (6) of a stable random vector involves integration over all directions of $\mathbb{R}^d$,[7] parameterized here by the unit sphere with respect to the Euclidean norm. Proposition 2.3.8 in Samorodnitsky & Taqqu (1994) shows that the unit sphere with respect to any norm can be used instead, provided a change of spectral measure and shift vector. We explore alternative representations where integration is carried out over a unit cylinder relative to a seminorm. However, for a given seminorm, not all stable vectors admit such a representation, motivating the following definition.

Definition 1 *Let $\|\cdot\|$ be a seminorm on $\mathbb{R}^d$, $C_d^{\|\cdot\|} := \{\boldsymbol{s} \in \mathbb{R}^d : \ \|\boldsymbol{s}\| = 1\}$ be the corresponding unit cylinder, and let $\boldsymbol{X} = (X_1, \ldots, X_d)$ be an α-stable random vector.*

(ι) *Asymmetric case:*

In the case where $\boldsymbol{X}$ is not symmetric, we say that $\boldsymbol{X}$ is representable on $C_d^{\|\cdot\|}$ if there exists a non-random vector $\boldsymbol{\mu}^0_{\|\cdot\|} \in \mathbb{R}^d$ and a Borel measure $\Gamma^{\|\cdot\|}$ on $C_d^{\|\cdot\|}$ satisfying for all $\boldsymbol{u} \in \mathbb{R}^d$

$$\int_{C_d^{\|\cdot\|}} |\langle \boldsymbol{u}, \boldsymbol{s}\rangle|^{\alpha} \Gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty, \tag{7}$$

if $\alpha \neq 1$, and if $\alpha = 1$,

$$\int_{C_d^{\|\cdot\|}} |\langle \boldsymbol{u}, \boldsymbol{s}\rangle| \Big| \ln |\langle \boldsymbol{u}, \boldsymbol{s}\rangle| \Big| \Gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty, \tag{8}$$

such that the joint characteristic function of $\boldsymbol{X}$ can be written as in (5) with $(S_d, \Gamma, \boldsymbol{\mu}^0)$ replaced by $(C_d^{\|\cdot\|}, \Gamma^{\|\cdot\|}, \boldsymbol{\mu}^0_{\|\cdot\|})$.

[7] By "direction" of $\mathbb{R}^d$, we mean the equivalence classes of the relation "$\equiv$" defined by: $\boldsymbol{u} \equiv \boldsymbol{v}$ if and only if there exists $\lambda > 0$ such that $\boldsymbol{u} = \lambda \boldsymbol{v}$, for $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^d$.

($\iota\iota$) *Symmetric case:*

In the case where $\boldsymbol{X}$ *is symmetric* α*-stable (S*α*S),* $0 < \alpha < 2$*, we say that* $\boldsymbol{X}$ *is representable on* $C_d^{\|\cdot\|}$ *if there exists a symmetric Borel measure* $\Gamma^{\|\cdot\|}$ *on* $C_d^{\|\cdot\|}$ *satisfying* (7) *such that the joint characteristic function of* $\boldsymbol{X}$ *can be written as in* (6) *with* (S_d, Γ) *replaced by* $(C_d^{\|\cdot\|}, \Gamma^{\|\cdot\|})$.

The integrability conditions (7)-(8) ensure the validity of the above definition. When $K^{\|\cdot\|} := \{x \in \mathbb{R}^d : \|x\| = 0\}$ satisfies $K^{\|\cdot\|} \neq \{0\}$, i.e. $\|\cdot\|$ is an appropriate seminorm, unit cylinders are unbounded sets, which necessitates these integrability conditions.[8] If $K^{\|\cdot\|} = \{0\}$, i.e. $\|\cdot\|$ comes down to a norm, the unit cylinder $C_d^{\|\cdot\|}$ coincides with the classical unit sphere, which is bounded, and conditions (7)-(8) are automatically satisfied for any finite measure $\Gamma^{\|\cdot\|}$ and the theory presented here reduces to standard representations of stable vectors on unit spheres (see Proposition 2.3.8 in Samorodnitsky & Taqqu (1994)). The genuine interest of seminorm representations arises when $K^{\|\cdot\|} \neq \{0\}$, as they allow for representations that encode information about only a subset of all possible directions in $\mathbb{R}^d$. This is precisely the framework needed for studying conditional extremes of time series, where we condition on observed past values (which determine a direction in a subspace) and wish to make inferences about future values. In this setting, the unit cylinder $C_d^{\|\cdot\|}$ captures all directions consistent with a given "shape" of the past trajectory, while remaining agnostic about the magnitude or direction of future components. For instance, in the example discussed in Remark 1, the cylinder $C_2^{\|\cdot\|} = \{(s_1, s_2) \in \mathbb{R}^2 : |s_1| = 1\}$ spans all directions in $\mathbb{R}^2$ except those where $s_1 = 0$, reflecting that we condition on the first coordinate being extreme while the second coordinate is unrestricted.

Proposition 1 *Let* $\|\cdot\|$ *be a seminorm on* $\mathbb{R}^d$ *and* $C_d^{\|\cdot\|}$ *be the corresponding unit cylinder. Denote* $K^{\|\cdot\|} = \{x \in S_d : \|x\| = 0\}$*. Let also* $\boldsymbol{X}$ *be an* α*-stable random vector on* $\mathbb{R}^d$ *with spectral representation* $(\Gamma, \boldsymbol{\mu}^0)$ *on the Euclidean unit sphere (with* $\boldsymbol{\mu}^0 = 0$ *if* $\boldsymbol{X}$ *is S*α*S). If*

[8] To see that $C_d^{\|\cdot\|}$ is unbounded when $K^{\|\cdot\|} \neq \{0\}$, observe that for any $x \in C_d^{\|\cdot\|}$ and any $v \in K^{\|\cdot\|} \setminus \{0\}$, we have $\|x + \lambda v\| = \|x\| = 1$ for all $\lambda \in \mathbb{R}$, so that $x + \lambda v \in C_d^{\|\cdot\|}$ for all $\lambda \in \mathbb{R}$.

$\alpha \neq 1$ or if $\boldsymbol{X}$ is S1S, then

$$\boldsymbol{X} \text{ is representable on } C_d^{\|\cdot\|} \iff \Gamma(K^{\|\cdot\|}) = 0.$$

If $\alpha = 1$ and $\boldsymbol{X}$ is not symmetric, then

$$\boldsymbol{X} \text{ is representable on } C_d^{\|\cdot\|} \iff \int_{S_d} \left| \ln \|\boldsymbol{s}\| \right| \Gamma(d\boldsymbol{s}) < +\infty.$$

Moreover, if $\boldsymbol{X}$ is representable on $C_d^{\|\cdot\|}$, its spectral representation is then given by $(\Gamma^{\|\cdot\|}, \boldsymbol{\mu}_{\|\cdot\|}^0)$ where

$$\Gamma^{\|\cdot\|}(d\boldsymbol{s}) = \|\boldsymbol{s}\|_e^{-\alpha}\, \Gamma \circ T_{\|\cdot\|}^{-1}(d\boldsymbol{s})$$

with $T_{\|\cdot\|} : S_d \setminus K^{\|\cdot\|} \longrightarrow C_d^{\|\cdot\|}$ defined by $T_{\|\cdot\|}(\boldsymbol{s}) = \boldsymbol{s}/\|\boldsymbol{s}\|$, and

$$\boldsymbol{\mu}_{\|\cdot\|}^0 = \begin{cases} \boldsymbol{\mu}^0, & \text{if } \alpha \neq 1 \text{ or if } \boldsymbol{X} \text{ is } S1S, \\ \boldsymbol{\mu}^0 + \tilde{\boldsymbol{\mu}}, & \text{if } \alpha = 1 \text{ and } \boldsymbol{X} \text{ is not symmetric,} \end{cases}$$

$$\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_j), \qquad \text{and} \qquad \tilde{\mu}_j = -\frac{2}{\pi} \int_{S_d \setminus K^{\|\cdot\|}} s_j \ln \|\boldsymbol{s}\| \Gamma(d\boldsymbol{s}), \quad j = 1, \ldots, d.$$

It can be noticed from Proposition 1 that the representability condition in the case $\alpha = 1$ and $\boldsymbol{X}$ is not symmetric, is slightly stronger than that in the other cases. Indeed, $\int_{K^{\|\cdot\|}} \left| \ln \|\boldsymbol{s}\| \right| \Gamma(d\boldsymbol{s}) \leq \int_{S_d} \left| \ln \|\boldsymbol{s}\| \right| \Gamma(d\boldsymbol{s}) < +\infty$ necessarily implies that $\Gamma(K^{\|\cdot\|}) = 0$ since $\left| \ln \|\boldsymbol{s}\| \right| = +\infty$ for $\boldsymbol{s} \in K^{\|\cdot\|}$.

Remark 1 The case $d = 2$ is particularly insightful. In light of (2), the spectral measure of the α-stable vector (X_1, X_2) characterizes its likelihood of being in any particular direction of $\mathbb{R}^2$ when it is large in norm. Since unit spheres relative to norms encompass all directions in $\mathbb{R}^2$, spectral measures on such spheres can capture any potential tail dependence of (X_1, X_2). However, unit cylinders do not span all directions in $\mathbb{R}^2$ and therefore, spectral measures on such cylinders encode less information. Consider for instance the unit cylinder $C_2^{\|\cdot\|} = \{(s_1, s_2) \in \mathbb{R}^2 : \ |s_1| = 1\}$ associated to the seminorm such that $\|(x_1, x_2)\| = |x_1|$ for all $(x_1, x_2) \in \mathbb{R}^2$. It is straightforward to observe that $C_2^{\|\cdot\|}$ spans all directions of $\mathbb{R}^2$ except for those of $(0, -1)$ and $(0, +1)$. A stable vector (X_1, X_2) will admit a representation

on $C_2^{\|\cdot\|}$ if these directions are irrelevant for characterizing its distribution, that is, if $\Gamma\Big(\{(0,-1),(0,+1)\}\Big) = 0$. In terms of tail dependence, the latter condition intuitively implies that realizations (X_1, X_2) where X_2 is extreme and X_1 is not, almost never occur (i.e., they occur with probability zero). The conditions $\Gamma\big(\{(0,-1),(0,+1)\}\big) = 0$ and $\int_{S_2} |\ln \|\boldsymbol{s}\||\Gamma(d\boldsymbol{s}) < +\infty$ can also be related to the stronger condition ensuring the existence of conditional moments of X_2 given X_1 as discussed in Cioczek-Georges & Taqqu (1994, 1998) - see also Theorem 5.1.3 in Samorodnitsky & Taqqu (1994). This stronger condition requires that Γ is not too concentrated around the points $(0, \pm 1)$. Specifically, assuming $\int_{S_2} |s_1|^{-\nu}\Gamma(ds) < +\infty$ for some $\nu \geq 0$, then $\mathbb{E}[|X_2|^\gamma | X_1] < +\infty$ for $\gamma < \min(\alpha + \nu, 2\alpha + 1)$, even though $\mathbb{E}[|X_2|^\alpha] = +\infty$. If this condition holds for some $\nu > 0$, then both of the aforementioned conditions must necessarily be satisfied.

Provided that the appropriate representation exists, (2) holds with seminorms instead of norms. This forms the foundation for studying the tail conditional distribution of stable processes and leads to the following proposition:

Proposition 2 *Let $\boldsymbol{X} = (X_1, \ldots, X_d)$ be an α-stable random vector and let $\|\cdot\|$ be a seminorm on $\mathbb{R}^d$. If $\boldsymbol{X}$ is representable on $C_d^{\|\cdot\|}$, then for every Borel sets $A, B \subset C_d^{\|\cdot\|}$ with $\Gamma^{\|\cdot\|}\Big(\partial(A \cap B)\Big) = \Gamma^{\|\cdot\|}(\partial B) = 0$, and $\Gamma^{\|\cdot\|}(B) > 0$,*

$$\mathbb{P}_x^{\|\cdot\|}(\boldsymbol{X}, A|B) \underset{x\to+\infty}{\longrightarrow} \frac{\Gamma^{\|\cdot\|}(A \cap B)}{\Gamma^{\|\cdot\|}(B)}, \tag{9}$$

where ∂B (resp. $\partial(A \cap B)$) denotes the boundary of B (resp. $A \cap B$), and

$$\mathbb{P}_x^{\|\cdot\|}(\boldsymbol{X}, A|B) := \mathbb{P}\Big(\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|} \in A \Big| \|\boldsymbol{X}\| > x, \frac{\boldsymbol{X}}{\|\boldsymbol{X}\|} \in B\Big).$$

3 Unit cylinder representation for the paths of stable linear processes

Given a seminorm, Proposition 2 is only applicable to stable vectors that are representable on the corresponding unit cylinder. This section investigates under which condition on an

stable moving average (X_t) vectors of the form $(X_{t-m}, \ldots, X_t, X_{t+1}, \ldots, X_{t+h})$ admit such representations. A characterization is proposed and extended to linear combinations of stable moving averages. Any seminorm satisfying (3) could be relevant for the prediction framework mentioned in the introduction. However, to simplify the discussion and avoid considering numerous cases for all possible kernels, we restrict attention to seminorms such that

$$\|(x_{-m}, \ldots, x_0, x_1, \ldots, x_h)\| = 0 \iff x_{-m} = \ldots = x_0 = 0, \tag{10}$$

for any $(x_{-m}, \ldots, x_h) \in \mathbb{R}^{m+h+1}$, which in particular satisfy (3). As an example, seminorms on $\mathbb{R}^{m+h+1}$ satisfying (10) can be naturally derived from norms on the $m+1$ first components of vectors. For any $p \in [1, +\infty]$, one can, for instance, consider seminorms of the form $\|\cdot\|$ defined by

$$\|(x_{-m}, \ldots, x_0, x_1, \ldots, x_h)\| = \Big(\sum_{i=-m}^{0} |x_i|^p \Big)^{1/p},$$

for any $(x_{-m}, \ldots, x_0, x_1, \ldots, x_h) \in \mathbb{R}^{m+h+1}$ with by convention $\big(\sum_{i=-m}^{0} |x_i|^p\big)^{1/p} = \sup_{-m \le i \le 0} |x_i|$ for $p = +\infty$. Now, assuming any seminorm satisfying (3) and (10), we consider (X_t) the α-stable moving average defined by

$$X_t = \sum_{k \in \mathbb{Z}} d_k \varepsilon_{t+k}, \qquad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, \beta, \sigma, 0) \tag{11}$$

with (d_k) a real deterministic sequence such that

$$\text{if} \quad \alpha \neq 1 \quad \text{or} \quad (\alpha, \beta) = (1, 0), \qquad 0 < \sum_{k \in \mathbb{Z}} |d_k|^s < +\infty, \qquad \text{for some} \quad s \in (0, \alpha) \cap [0, 1], \tag{12}$$

and

$$\text{if} \quad \alpha = 1 \quad \text{and} \quad \beta \neq 0, \qquad 0 < \sum_{k \in \mathbb{Z}} |d_k| \Big| \ln |d_k| \Big| < +\infty. \tag{13}$$

Letting for $m \ge 0$, $h \ge 1$,

$$\boldsymbol{X_t} = (X_{t-m}, \ldots, X_t, X_{t+1}, \ldots, X_{t+h}), \tag{14}$$

it follows from Proposition 13.3.1 in Brockwell & Davis (1991) that the infinite series converge almost surely and both (X_t) and $\boldsymbol{X}_t$ are well-defined and the random vector $\boldsymbol{X_t}$ is multivariate α-stable. Denoting $\boldsymbol{d}_k := (d_{k+m}, \ldots, d_k, d_{k-1}, \ldots, d_{k-h})$ for $k \in \mathbb{Z}$, the spectral representation of $\boldsymbol{X_t}$ on the Euclidean sphere is given by $(\Gamma, \boldsymbol{\mu}^0)$ with

$$\Gamma = \sigma^\alpha \sum_{\vartheta \in S_1} \sum_{k \in \mathbb{Z}} w_\vartheta \|\boldsymbol{d}_k\|_e^\alpha \delta_{\left\{\frac{\vartheta \boldsymbol{d}_k}{\|\boldsymbol{d}_k\|_e}\right\}}, \tag{15}$$

$$\boldsymbol{\mu}^0 = -\mathbb{1}_{\{\alpha=1\}} \frac{2}{\pi} \beta\sigma \sum_{k \in \mathbb{Z}} \boldsymbol{d}_k \ln \|\boldsymbol{d}_k\|_e,$$

where $w_\vartheta = (1 + \vartheta\beta)/2$, $S_1 = \{-1, +1\}$, δ is the Dirac mass and by convention, if for some $k \in \mathbb{Z}$, $\boldsymbol{d}_k = \boldsymbol{0}$, i.e. $\|\boldsymbol{d}_k\|_e = 0$, then the kth term vanishes from the sums. Equation (15) reveals that the spectral measure Γ is a weighted sum of Dirac masses concentrated on the points $\{\vartheta \boldsymbol{d}_k / \|\boldsymbol{d}_k\|_e : k \in \mathbb{Z}, \vartheta \in S_1, \boldsymbol{d}_k \neq \boldsymbol{0}\}$. This specific discrete structure, where the spectral measure is intrinsically linked to the MA coefficients, distinguishes stable moving averages and is new in the heavy-tailed time series literature (Mikosch & Wintenberger 2024). It has important implications for extreme event prediction, as extreme trajectories concentrate on a countable (often finite) collection of deterministic patterns rather than a continuum of possible directions, which allows for inference of these deterministic paths. Notice in particular that for $\beta = 0$, it holds that $w_{-1} = w_{+1} = 1/2$, $\boldsymbol{\mu}^0 = \boldsymbol{0}$, and both the measure Γ and the random vector $\boldsymbol{X}_t$ are symmetric. The following representation theorem characterizes the representability of $\boldsymbol{X_t}$ on a unit cylinder for fixed m and h.

Theorem 1 *Let $\boldsymbol{X_t}$ satisfy (11)-(14) and let $\|\cdot\|$ be a seminorm on $\mathbb{R}^{m+h+1}$ satisfying (10). For $\alpha \neq 1$ or $(\alpha, \beta) = (1, 0)$, the vector $\boldsymbol{X_t}$ is representable on $C^{\|\cdot\|}_{m+h+1}$ if and only if*

$$\forall k \in \mathbb{Z}, \quad \Big[(d_{k+m}, \ldots, d_k) = \boldsymbol{0} \quad \Longrightarrow \quad \forall \ell \leq k-1, \quad d_\ell = 0\Big]. \tag{16}$$

For $\alpha = 1$ and $\beta \neq 0$, the vector $\boldsymbol{X_t}$ is representable on $C^{\|\cdot\|}_{m+h+1}$ if and only if in addition to (16), it holds that

$$\sum_{k \in \mathbb{Z}} \|\boldsymbol{d}_k\|_e \left| \ln\left(\|\boldsymbol{d}_k\| / \|\boldsymbol{d}_k\|_e\right) \right| < +\infty. \tag{17}$$

In the cases, $\alpha \neq 1$ and $(\alpha, \beta) = (1, 0)$, Theorem 1 shows that the representability of $\boldsymbol{X_t}$ on a $C^{\|\cdot\|}_{m+h+1}$ depends on the number of observation $m+1$ but not on the prediction horizon h.[9] This is particularly important for the applicability of Proposition 2 when studying the conditional dynamics of a given stable moving average. Indeed, this proposition relies on seminorms instead of norms, meaning that we use only a portion of size $m+1$ of an observed path. This leads to the notion of *past-representability*, stated by the following definition:

Definition 2 Let (X_t) be an α-stable moving average satisfying (11)-(13). We say that the stable process (X_t) is *past-representable* if there exists at least one pair (m, h), $m \geq 0$, $h \geq 1$, such that $\boldsymbol{X_t} = (X_{t-m}, \ldots, X_t, X_{t+1}, \ldots, X_{t+h})$ is representable on $C^{\|\cdot\|}_{m+h+1}$ for any seminorm satisfying (10). For any such pair (m, h), we will say that (X_t) is (m, h)-past-representable.

Definition 2, holds for any seminorm satisfying (10) because it ensures that all these seminorms have the same kernel. Moreover, it is straightforward to observe that this definition holds for any $m' \geq m$, because if (16) is true for some $m \geq 0$, it is true for any $m' \geq m$. Thus, the notion of past-representability can be defined independently of the particular choice of a seminorm but relies on the existence of at least one m, for which the considered process is (m, h)-past-representable.[10] The following proposition provides this characterization.

Proposition 3 *Let (X_t) be an α-stable moving average satisfying (11)-(13).*

(ι) *With the set $\mathcal{M} = \{m \geq 1 : \ \exists k \in \mathbb{Z}, \ d_{k+m} = \ldots = d_{k+1} = 0, \ d_k \neq 0\}$, define*

$$m_0 = \begin{cases} \sup \mathcal{M}, & \text{if} \quad \mathcal{M} \neq \emptyset, \\ 0, & \text{if} \quad \mathcal{M} = \emptyset. \end{cases} \tag{18}$$

[9] The case $\alpha = 1$, $\beta \neq 0$ is more intricate, as the roles of m and h in the validity of the additional requirement (17) are not as clear-cut.

[10] This will not be true in general under the weaker assumption (3) and different notions of representability of a process could emerge depending on the kernels of the seminorms.

(*a*) *For* $\alpha \neq 1$ *or* $(\alpha, \beta) = (1, 0)$*, the process* (X_t) *is past-representable if and only if*

$$m_0 < +\infty. \tag{19}$$

Moreover, letting $m \geq 0$*,* $h \geq 1$*, the process* (X_t) *is* (m, h)*-past-representable if and only if* (19) *holds and* $m \geq m_0$*.*

(*b*) *For* $\alpha = 1$ *and* $\beta \neq 0$*, the process* (X_t) *is past-representable if and only if in addition to* (19)*, there exist an* $m \geq m_0$ *and an* $h \geq 1$ *such that* (17) *holds. If such a pair* (m, h) *exists,* (X_t) *is then* (m, h)*-past-representable.*

($\iota\iota$) *Let* $\|\cdot\|$ *a seminorm satisfying* (10) *and assume that* (X_t) *is* (m, h)*-past-representable for some* $m \geq 0$*,* $h \geq 1$*. The spectral representation* $(\Gamma^{\|\cdot\|}, \boldsymbol{\mu}^{\|\cdot\|})$ *of the vector* $\boldsymbol{X_t} = (X_{t-m}, \ldots, X_t, X_{t+1}, \ldots, X_{t+h})$ *on* $C^{\|\cdot\|}_{m+h+1}$ *is then given by* (15) *with the Euclidean norm* $\|\cdot\|_e$ *replaced by the seminorm* $\|\cdot\|$*.*

It can be noted, that $m_0 = 0$ if and only if for some $k_0 \in \mathbb{Z} \cup \{-\infty\}$, $d_k \neq 0$ for all $k \geq k_0$ and $d_k = 0$ for all $k < k_0$. Proposition 3 shows that for an α-stable moving average to be past-representable, sequences of consecutive zero values in the coefficients (d_k) must be either of finite length or infinite to the right. This surprisingly makes the anticipativeness of a stable moving average a necessary condition, and sufficient for $\alpha \neq 1$ and $(\alpha, \beta) = (1, 0)$, for its past-representability. The less anticipative a moving average is (i.e., the larger the gaps of zeros on its forward-looking side), the higher m must be chosen to ensure the representability of $(X_{t-m}, \ldots, X_t, X_{t+1}, \ldots, X_{t+h})$ on the appropriate unit cylinder. Purely non-anticipative moving averages are, in particular, immediately ruled out. This property is demonstrated in the following corollary.

Corollary 1 Let (X_t) an α-stable moving average satisfying (11)-(13). If (X_t) is purely non-anticipative, i.e., $d_k = 0$ for all $k \geq 1$, then (X_t) is not past-representable.

Corollary 1, sheds light on the predictability of extreme events in linear processes. For illustration, we consider the following two α-stable AR(1) processes, defined as the stationary

solutions of:

$$X_t = \rho X_{t+1} + \varepsilon_t, \qquad \forall t \in \mathbb{Z}, \tag{20}$$

$$Y_t = \rho Y_{t-1} + \eta_t, \qquad \forall t \in \mathbb{Z}, \tag{21}$$

where $0 < |\rho| < 1$, and (ε_t), (η_t) are independent i.i.d. stable sequences. While (X_t) generates bubble-like trajectories –explosive exponential paths eventually followed by sharp returns to central values–, the trajectories of (Y_t) feature sudden jumps followed by exponential decays. In both processes, an extreme event results from a large realisation of an underlying error ε_τ or η_τ, at some time τ. On the one hand, for the non-anticipative AR(1) process (21), a jump does not manifest any early visible signs before its date of occurrence, as it is independent of the past trajectory. Jumps in the trajectory of (Y_t) are unpredictable, and one only has information about their unconditional likelihood of occurrence. On the other hand, for the anticipative AR(1) process (20), extremes do manifest early visible signs and are gradually reached as their occurrence dates approach. The past trajectory is informative about future extreme events, and, in particular, it is more informative than their plain unconditional likelihood of occurrence. Building on the "information encoding" interpretation of spectral measures given in Remark 1, the fact that (X_t) (resp. (Y_t)) is past-representable (resp. not past-representable) can be seen as a consequence of the dependence (resp. independence) of future extreme events on past ones. The following Corollary 2 states the condition for past-representability for ARMA.

Corollary 2 Let (X_t) be the strictly stationary solution of

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

where Ψ, Φ, Θ, H are polynomials of arbitrary finite degrees with roots located outside the unit disk and F (resp. B) is the forward (resp. backward) operator: $FX_t := X_{t+1}$ (resp. $BX_t := X_{t-1}$). We suppose furthermore that Ψ and Θ (resp. Φ and H) have no common roots. Then, for any $\alpha \in (0, 2)$ and $\beta \in [-1, 1]$, the following statements are equivalent:

(ι) (X_t) is past-representable,

$(\iota\iota)$ $\deg(\Psi) \geq 1$,

$(\iota\iota\iota)$ $m_0 < +\infty$,

with m_0 as in (18). Moreover, letting $m \geq 0$, $h \geq 1$, the process (X_t) is (m, h)-past-representable if and only if $m \geq m_0$ with $m_0 < +\infty$.

Corollary 2 reveals that for ARMA processes, the past-representability condition simplifies and is equivalent to the autoregressive polynomial having at least one root inside the unit circle. Also, only the roots of the AR polynomial matter for past-representability, with the MA part playing no role. Numerical illustrations of the representation theorem for ARMA processes and the contrasting angular concentration patterns between causal and anticipative processes in extreme regimes are provided in Appendices A.3 and A.4 of the Supplementary Material, respectively. In the noncausal literature, when $\Theta = H = 1$, this is called a mixed-causal $MAR(p, q)$ model, first introduced by Lanne & Saikkonen (2011), where p corresponds to $\deg(\Phi)$ and q corresponds to $\deg(\Psi)$.

4 Tail conditional distribution of stable anticipative processes

In this section, we will derive the tail conditional distribution of linear stable processes for which Proposition 2 will be applicable. The case of a general past-representable stable process is considered, along with particular examples. To be relevant for the prediction framework, the Borel set B appearing in Proposition 2 must be chosen such that the conditioning event $\{\|\boldsymbol{X_t}\| > x\} \cap \{\boldsymbol{X_t}/\|\boldsymbol{X_t}\| \in B\}$ is independent of the future realizations $X_{t+1}, \ldots, X_{t+h}$. For $\|\cdot\|$ a seminorm on $\mathbb{R}^{m+h+1}$ satisfying (10), denote $S_{m+1}^{\|\cdot\|} = \{(s_{-m}, \ldots, s_0) \in \mathbb{R}^{m+1} : \|(s_{-m}, \ldots, s_0, 0, \ldots, 0)\| = 1\}$.[11] Then, for any Borel set $V \subset S_{m+1}^{\|\cdot\|}$, define the Borel set $B(V) \subset C_{m+h+1}^{\|\cdot\|}$ as

$$B(V) = V \times \mathbb{R}^h.$$

[11] The set $S_{m+1}^{\|\cdot\|}$ corresponds to the unit sphere of $\mathbb{R}^{m+1}$ relative to the restriction of $\|\cdot\|$ to the first $m+1$ dimensions.

Notice in particular that for $V = S_{m+1}^{\|\cdot\|}$, we have $B(V) = C_{m+1}^{\|\cdot\|} = S_{m+1}^{\|\cdot\|}$. In the following, we will use Borel sets of the above form to condition the distribution of the entire vector $\boldsymbol{X_t}/\|\boldsymbol{X_t}\|$ on the observed "shape" of the past trajectory. The latter information is contained in the Borel set V, which we will typically assume to be a small neighborhood on $S_{m+1}^{\|\cdot\|}$. It will be useful in the following to note that

$$V \times \mathbb{R}^h = \left\{ \boldsymbol{s} \in C_{m+h+1}^{\|\cdot\|} : \quad f(\boldsymbol{s}) \in V \right\},$$

where f is the function defined by

$$f : \begin{array}{ccc} \mathbb{R}^{m+h+1} & \longrightarrow & \mathbb{R}^{m+1} \\ (x_{-m}, \dots, x_0, x_1, \dots, x_h) & \longmapsto & (x_{-m}, \dots, x_0) \end{array}. \tag{22}$$

4.1 Stable past-representable processes: general case

Let (X_t) an α-stable process satisfying Definition 2. This states that (X_t) is (m, h)-past-representable, for some $m \geq 0$, $h \geq 1$ and let $\boldsymbol{X_t}$ as in (14). Denoting $\Gamma^{\|\cdot\|}$ the spectral measure of $\boldsymbol{X}_t$ on the unit cylinder $C_{m+h+1}^{\|\cdot\|}$ for some seminorm satisfying (10), we know by Proposition 3 ($\iota\iota$), that $\Gamma^{\|\cdot\|}$ is of the form

$$\Gamma^{\|\cdot\|} = \sum_{\vartheta \in S_1} \sum_{k \in \mathbb{Z}} w_\vartheta \|\boldsymbol{d}_k\|^\alpha \delta_{\left\{ \frac{\vartheta \boldsymbol{d}_k}{\|\boldsymbol{d}_k\|} \right\}}. \tag{23}$$

leading to the following proposition:

Proposition 4 *Under the above assumptions, we have*

$$\mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t}, A \Big| B(V)\Big) \underset{x \to +\infty}{\longrightarrow} \frac{\Gamma^{\|\cdot\|}\left(\left\{ \frac{\vartheta \boldsymbol{d}_k}{\|\boldsymbol{d}_k\|} \in A : \quad \frac{\vartheta f(\boldsymbol{d}_k)}{\|\boldsymbol{d}_k\|} \in V \right\} \right)}{\Gamma^{\|\cdot\|}\left(\left\{ \frac{\vartheta \boldsymbol{d}_k}{\|\boldsymbol{d}_k\|} \in C_{m+h+1}^{\|\cdot\|} : \quad \frac{\vartheta f(\boldsymbol{d}_k)}{\|\boldsymbol{d}_k\|} \in V \right\} \right)}, \tag{24}$$

for any Borel sets $A \subset C_{m+h+1}^{\|\cdot\|}$, $V \subset S_{m+1}^{\|\cdot\|}$ such that $\left\{ \frac{\vartheta \boldsymbol{d}_k}{\|\boldsymbol{d}_k\|} \in C_{m+h+1}^{\|\cdot\|} : \quad \frac{\vartheta f(\boldsymbol{d}_k)}{\|\boldsymbol{d}_k\|} \in V \right\} \neq \emptyset$, $\Gamma^{\|\cdot\|}\Big(\partial(A \cap B(V))\Big) = \Gamma^{\|\cdot\|}(\partial B(V)) = 0$, where $B(V) = V \times \mathbb{R}^h$ and f is as in (22).

From Proposition 4 and by setting $V = S_{m+1}^{\|\cdot\|}$, and A an arbitrarily small closed neighborhood of all the points $(\vartheta \boldsymbol{d}_k/\|\boldsymbol{d}_k\|)_{\vartheta,k}$, we can see that $\lim_{x \to +\infty} \mathbb{P}\Big(\boldsymbol{X_t}/\|\boldsymbol{X_t}\| \in A \Big| \|\boldsymbol{X_t}\| > x\Big) = 1$. In

other words, when far from central values, the trajectory of the process (X_t) necessarily features patterns of the same shape as some $\vartheta \boldsymbol{d}_k/\|\boldsymbol{d}_k\|$, which is a finite piece of a moving average's coefficient sequence. The index k indicates to which piece $(d_{k+m}, \ldots, d_k, d_{k-1}, \ldots, d_{k-h})$ of this moving average it corresponds to, and, $\vartheta \in \{-1, +1\}$ indicates whether the pattern is flipped upside down (in case the extreme event is driven by a negative value of an error ε_τ). The likelihood of a pattern $\vartheta \boldsymbol{d}_k/\|\boldsymbol{d}_k\|$ can be evaluated by setting A to be a small neighborhood of that point. From this viewpoint, the observed path, $(X_{t-m}, \ldots, X_{t-1}, X_t)/\|\boldsymbol{X_t}\|$ will *a fortiori* be of the same shape as some $\vartheta(d_{k+m}, \ldots, d_{k+1}, d_k)/\|\boldsymbol{d}_k\|$ when an extreme event approaches in time. Observing the initial part of the pattern can provide information about the remaining unobserved part: the conditional likelihood of the latter can be assessed by setting V to be a small neighborhood of the observed pattern.

Remark 2 The tail conditional distribution given in (24) highlights three types of uncertainty/approximation for prediction. We deal with two of them in the rest of the paper and leave the third for future research.[12]

(ι) In practice, events of the type

$$\{(X_{t-m}, \ldots, X_{t-1}, X_t)/\|\boldsymbol{X_t}\| = \vartheta(d_{k+m}, \ldots, d_{k+1}, d_k)/\|\boldsymbol{d}_k\|\}$$

have probability zero of occurring, and only noisy observations such as $(X_{t-m}, \ldots, X_{t-1}, X_t)/\|\boldsymbol{X_t}\| \approx \vartheta(d_{k+m}, \ldots, d_{k+1}, d_k)/\|\boldsymbol{d}_k\|$ are available on a realized trajectory. The choice of an adequate conditioning neighborhood V in (24) given a piece of the trajectory, will thus need to rely on a statistical approach. One could envision hypothesis tests to determine whether a piece of the realized (noisy) trajectory "is more similar" to a certain pattern 1 or to another pattern 2.

($\iota\iota$) Even if the observed path can be confidently identified with a particular pattern,

[12]The considerations developed in this remark focus solely on the probabilistic uncertainty of the prediction, assuming that the process (X_t) is entirely known; that is, no parameter nor any sequence $(d_{j,k})$ needs to be inferred from the data. We leave for future research issues related to statistical uncertainty.

uncertainty regarding the future trajectory may remain. It could indeed be that several patterns $\vartheta \boldsymbol{d}_k/\|\boldsymbol{d}_k\|$ coincide on their first $m+1$ components, but differ in the last h components. The stable anticipative AR(1) process is a typical example of this phenomenon, which will be studied in the next section. Interestingly, the stable anticipative AR(2) process, and more generally persistent enough processes, avoid this issue, as discussed hereafter.

($\iota\iota\iota$) The tail conditional distribution (24) represents an asymptotic behavior as the (semi-) norm of $\boldsymbol{X_t}$ grows infinitely large. It is, therefore, only an approximation of the true dynamics during extreme events. It would be interesting to obtain a more refined asymptotic development in x of the above convergence in order to gauge the approximation error of the true conditional distribution. We partially investigate this issue by means of Monte Carlo simulation. In particular, we numerically quantify how far from„ the predicted patterns the future path can be, depending on the quantile of the observed realizations.

4.2 The anticipative AR(1)

We now consider (X_t) the stable anticipative AR(1) process defined by

$$X_t = \rho X_{t+1} + \varepsilon_t, \quad 0 < |\rho| < 1, \tag{25}$$

where $(\varepsilon_t)_{t\in\mathbb{Z}} \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, \beta, 1, 0)$. The moving average coefficient is of the form $(\rho^k \mathbb{1}_{\{k\geq 0\}})_k$, and thus, $m_0 = 0$ as stated in (18). By Corollary 2, we know that for any $m \geq 0$, $h \geq 1$, (X_t) is (m, h)-past-representable. The spectral measures of paths $\boldsymbol{X}_t$ simplify and charge finitely many points. Their forms are given in the following lemma.

Lemma 1 *Let (X_t) be an α-stable anticipative AR(1) processes as in (25). Letting $\boldsymbol{X_t}$ as in (14) for $m \geq 0$, $h \geq 1$, its spectral measure on $C^{\|\cdot\|}_{m+h+1}$ for a seminorm satisfying (10) is*

given by

$$\Gamma^{\|\cdot\|} = \sum_{\vartheta\in S_1}\left[w_\vartheta\delta_{\{(\vartheta,0,\dots,0)\}} + w_\vartheta\sum_{k=-m+1}^{h-1}\|\boldsymbol{d}_k\|^\alpha\delta_{\left\{\frac{\vartheta\boldsymbol{d}_k}{\|\boldsymbol{d}_k\|}\right\}} + \frac{\bar{w}_\vartheta}{1-|\rho|^\alpha}\|\boldsymbol{d}_h\|^\alpha\delta_{\left\{\frac{\vartheta\,\boldsymbol{d}_h}{\|\boldsymbol{d}_h\|}\right\}}\right], \quad (26)$$

where for all $\vartheta\in S_1$ *and* $-m+1\le k\le h$,

$$\begin{aligned}\boldsymbol{d}_k &= (\rho^{k+m}\mathbb{1}_{\{k\ge -m\}},\dots,\rho^k\mathbb{1}_{\{k\ge 0\}},\rho^{k-1}\mathbb{1}_{\{k\ge 1\}},\dots,\rho^{k-h}\mathbb{1}_{\{k\ge h\}}),\\ w_\vartheta &= (1+\vartheta\beta)/2,\\ \bar{w}_\vartheta &= (1+\vartheta\bar{\beta})/2,\\ \bar{\beta} &= \beta\frac{1-\rho^{<\alpha>}}{1-|\rho|^\alpha},\end{aligned}$$

and if $h=1$ *and* $m=0$, *the sum* $\sum_{k=-m+1}^{h-1}$ *vanishes by convention.*

The next proposition provides the tail conditional distribution of future paths in the case where ρ is positive. Let us first introduce useful neighborhoods of the distinct charged points of $\Gamma^{\|\cdot\|}$. Denote $\boldsymbol{d}_{0,-m} = (\overbrace{1,0,\dots,0}^{m+h+1})$ so that the charged points of $\Gamma^{\|\cdot\|}$ are all of the form $\vartheta\boldsymbol{d}_k/\|\boldsymbol{d}_k\|$ with indexes (ϑ,k) in the set $\mathcal{I} := S_1\times\Big(\{-m,h\}\cup\{(0,-m)\}\Big)$. With f as in (22), define for any $(\vartheta_0,k_0)\in\mathcal{I}$, the set V_0 as any closed neighborhood of $\vartheta_0 f(\boldsymbol{d}_{k_0})/\|\boldsymbol{d}_{k_0}\|$ such that

$$\forall(\vartheta',k')\in\mathcal{I},\qquad \frac{\vartheta' f(\boldsymbol{d}_{k'})}{\|\boldsymbol{d}_{k'}\|}\in V_0 \implies \frac{\vartheta' f(\boldsymbol{d}_{k'})}{\|\boldsymbol{d}_{k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{k_0})}{\|\boldsymbol{d}_{k_0}\|}, \quad (27)$$

In other terms, $V_0\times\mathbb{R}^d$ is a subset of $C_{m+h+1}^{\|\cdot\|}$ where the only points charged by $\Gamma^{\|\cdot\|}$ have their first $(m+1)^{\text{th}}$ components coinciding with $\vartheta_0 f(\boldsymbol{d}_{k_0})/\|\boldsymbol{d}_{k_0}\|$. Define $A_{\vartheta,k}$ for any (ϑ,k) as any closed neighborhood of $\vartheta\boldsymbol{d}_k/\|\boldsymbol{d}_k\|$ that does not contain any other charged point of $\Gamma^{\|\cdot\|}$, meaning:

$$\forall(\vartheta',k')\in\mathcal{I},\qquad \frac{\vartheta'\boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|}\in A_{\vartheta,k} \implies (\vartheta',k') = (\vartheta,k). \quad (28)$$

Proposition 5 *Let* (X_t) *be an* α*-stable anticipative AR(1) processes as in* (25) *with* $\rho\in(0,1)$. *Let* $\boldsymbol{X_t}$, *the* $\boldsymbol{d}_k$*'s and the spectral measure of* $\boldsymbol{X_t}$ *be as given in Lemma 1, for any* $m\ge 0$, $h\ge 1$. *Let* V_0 *be any small closed neighborhood of* $\vartheta_0 f(\boldsymbol{d}_{k_0})/\|\boldsymbol{d}_{k_0}\|$ *in the sense of*

(27) *for some* $(\vartheta_0, k_0) \in \mathcal{I}$ *and let* $B(V_0) = V_0 \times \mathbb{R}^h$. *Then, with* $A_{\vartheta,k}$ *an arbitrarily small neighborhood of some* $\vartheta \boldsymbol{d}_k/\|\boldsymbol{d}_k\|$ *as in* (28), *the following hold.*

(ι) *Case* $m \geq 1$.

(a) *If* $0 \leq k_0 \leq h$*:*

$$\mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t}, A_{\vartheta,k}\Big|B(V_0)\Big) \underset{x\to\infty}{\longrightarrow} \begin{cases} |\rho|^{\alpha k}(1-|\rho|^\alpha)\delta_{\vartheta_0}(\vartheta), & 0 \leq k \leq h-1, \\ \\ |\rho|^{\alpha h}\delta_{\vartheta_0}(\vartheta), & k = h. \end{cases}$$

(b) *If* $-m \leq k_0 \leq -1$*:*

$$\mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t}, A_{\vartheta,k}\Big|B(V_0)\Big) \underset{x\to\infty}{\longrightarrow} \delta_{\vartheta_0}(\vartheta)\delta_{k_0}(k).$$

($\iota\iota$) *Case* $m = 0$.

$$\mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t}, A_{\vartheta,k}\Big|B(V_0)\Big) \underset{x\to\infty}{\longrightarrow} \begin{cases} \dfrac{w_{\vartheta_0}}{p_{\vartheta_0}}\delta_{\{\vartheta_0\}}(\vartheta), & k = 0 \\ \\ |\rho|^{\alpha k}(1-|\rho|^\alpha)\delta_{\{\vartheta_0\}}(\vartheta), & 1 \leq k \leq h-1, \\ \\ |\rho|^{\alpha h}\delta_{\{\vartheta_0\}}(\vartheta), & k = h, \end{cases}$$

with $p_{\vartheta_0} = w_{\vartheta_0}/(1-|\rho|^\alpha)$.

For $m \geq 1$, meaning the observed path is assumed to be of length at least 2, there is a significant distinction depending on whether $k_0 \in \{0, \ldots, h\}$ or $k_0 \in \{-m, \ldots, -1\}$. In the latter case, this means that the shock is already observed. This implies that, given the observed path, the shape of the future trajectory is completely determined, as the asymptotic probability of the entire path $\boldsymbol{X_t}/\|\boldsymbol{X_t}\|$ being in an arbitrarily small neighborhood of $\vartheta \boldsymbol{d}_k/\|\boldsymbol{d}_k\|$ is 1 when $\vartheta = \vartheta_0$, $k = k_0$. For the former case, this probability is strictly positive if and only if $\vartheta = \vartheta_0$, but the observed pattern is compatible with several distinct future trajectories. This can be understood by examining the form of the sequences $\boldsymbol{d}_k/\|\boldsymbol{d}_k\|$ and

their restrictions to the first $m+1$ components $f(\boldsymbol{d}_k)/\|\boldsymbol{d}_k\|$. On the one hand (omitting ϑ),

$$\frac{\boldsymbol{d}_k}{\|\boldsymbol{d}_k\|} = \begin{cases} \dfrac{(\overbrace{\rho^{k+m},\dots,\rho^k}^{m+1},\overbrace{\rho^{k-1},\dots,\rho,1,0,\dots,0}^{h})}{\|(\rho^{k+m},\dots,\rho^k,\rho^{k-1},\dots,\rho,1,0,\dots,0)\|}, & \text{for} \qquad k\in\{0,\dots,h\}, \\ \dfrac{(\rho^{k+m},\dots,\rho,1,0,\dots,0,0,\dots,0)}{\|(\underbrace{\rho^{k+m},\dots,\rho,1,0,\dots,0}_{m+1},\underbrace{0,\dots,0}_{h})\|}, & \text{for} \qquad k\in\{-m,\dots,-1\}. \end{cases}$$

We can observe that all the above sequences are segments of explosive exponential functions, each truncated at a certain coordinate. For $k \in \{0,\dots,h\}$, the first zero component, representing the "crash of the bubble", is located at or after the $(m+2)^{\text{th}}$ component, whereas for $k \in \{-m,\dots,-1\}$, it is located at or before the $(m+1)^{\text{th}}$ component. Using the homogeneity of the seminorm and (3), we have on the other hand that

$$\frac{f(\boldsymbol{d}_k)}{\|\boldsymbol{d}_k\|} = \begin{cases} \dfrac{(\overbrace{\rho^m,\dots,\rho,1}^{m+1})}{\|(\underbrace{\rho^m,\dots,\rho,1}_{m+1},\underbrace{0,\dots,0,0,\dots,0}_{h})\|}, & \text{for} \qquad k\in\{0,\dots,h\}, \\ \dfrac{(\overbrace{\rho^{k+m},\dots,\rho,1,0,\dots,0}^{m+1})}{\|(\underbrace{\rho^{k+m},\dots,\rho,1,0,\dots,0}_{m+1},\underbrace{0,\dots,0}_{h})\|}, & \text{for} \qquad k\in\{-m,\dots,-1\}. \end{cases}$$

Thus, conditioning the trajectory on the event $\{f(\boldsymbol{X_t})/\|\boldsymbol{X_t}\| \approx f(\boldsymbol{d}_{k_0})/\|\boldsymbol{d}_{k_0}\|\}$ for some $k_0 \in \{-m,\dots,-1\}$ amounts to conditioning on the burst of a bubble being observed in the past trajectory, with no new bubble forming yet. This allows for the exact identification of the position of the pattern within the moving average's coefficient sequence.

When conditioning with $k_0 \in \{0,\dots,h\}$ however, the crash date is not observed and can occur either within the next $h-1$ periods or after h. The shape of the observed path, though, corresponds to a segment of exponential growth with a rate of ρ^{-1} regardless of how much time remains before the burst. This leaves several potential future paths. The likelihood of each possible scenario can be quantified as follows: the quantity $|\rho|^{\alpha k}(1-|\rho|^{\alpha})$ represents the probability that the bubble will peak in exactly k periods ($0 \le k < h$), while $|\rho|^{\alpha h}$ represents the probability that the bubble will last at least h more periods. This confirms

the interpretation of the conditional moments proposed in Fries (2022). Additionally, it extends this interpretation by considering entire paths rather than just point predictions. For $m = 0$, meaning only the present value is observed, no pattern can be identified, only the sign of the shock. Hence, the growth rate ρ^{-1} of the ongoing event becomes unidentifiable. This is reflected in the fact that the asymptotic probabilities of paths with growth rates ρ^{-1}, are positive (as shown in case $(\iota\iota)$ of Proposition 5).

4.3 The anticipative AR(2) and fractionally integrated white noise

We focus here on two processes that both share the interesting property of having a 0-1 tail conditional distribution whenever the observed path has length at least 2 (i.e., $m \geq 1$): the anticipative AR(2) and the anticipative fractionally integrated white noise (FWN). With an appropriate choice of parameters, the former can generate bubble-like trajectories with accelerating or decelerating growth rates, while the latter can accommodate hyperbolic bubbles. In contrast with the anticipative AR(1), these bubbles do not display an exponential profile, but they still exhibit an inflation-peak-collapse behavior. Any extension of these two minimal specifications should preserve the following properties.

Anticipative AR(2)

The anticipative AR(2) is the strictly stationary solution of

$$(1 - \lambda_1 F)(1 - \lambda_2 F)X_t = \varepsilon_t, \qquad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, \beta, \sigma, 0), \tag{29}$$

where $\lambda_i \in \mathbb{C}$ and $0 < |\lambda_i| < 1$ for i=1,2. In case $\lambda_i \in \mathbb{C} \setminus \mathbb{R}$, $i = 1, 2$, we impose that $\lambda_1 = \bar{\lambda}_2$ to ensure (X_t) is real-valued. We further assume that $\lambda_1 + \lambda_2 \neq 0$, to exclude the cases where (X_{2t}) and (X_{2t+1}) are independent anticipative AR(1) processes. The solution of (29) admits the moving average representation $X_t = \sum_{k\in\mathbb{Z}} d_k \varepsilon_{t+k}$ with

$$d_k = \begin{cases} \dfrac{\lambda_1^{k+1} - \lambda_2^{k+1}}{\lambda_1 - \lambda_2} \mathbb{1}_{\{k\geq 0\}}, & \text{if} \quad \lambda_1 \neq \lambda_2, \\ (k+1)\lambda^k \mathbb{1}_{\{k\geq 0\}}, & \text{if} \quad \lambda_1 = \lambda_2 = \lambda. \end{cases} \tag{30}$$

Anticipative fractionally integrated white noise

The anticipative FWN process can be defined as the stationary solution of

$$(1-F)^d X_t = \varepsilon_t, \qquad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, \beta, \sigma, 0), \tag{31}$$

with $\alpha(d-1) < -1$. The solution of (31) admits the moving average representation $X_t = \sum_{k=0}^{+\infty} d_k \varepsilon_{t+k}$ with $d_0 = 1$ and

$$d_k = \frac{\Gamma_f(k+d)}{\Gamma_f(d)\Gamma_f(k+1)} \mathbb{1}_{\{k \geq 0\}}, \quad \text{for} \quad k \neq 0, \tag{32}$$

where $\Gamma_f(\,\cdot\,)$ denotes the Gamma function.

It can be shown that both process are necessarily (m,h)-past-representable for $m \geq 1$ and $h \geq 1$. The 0-1 tail conditional distribution property, when the observed path has length at least 2, is demonstrated in the next proposition:

Proposition 6 *Let (X_t) be the α-stable anticipative AR(2) or anticipative fractionally integrated white noise. For any $m \geq 1$ and $h \geq 1$, let $\boldsymbol{X_t}$ as in (14) and $\boldsymbol{d}_k = (d_{k+m}, \ldots, d_k, d_{k-1}, \ldots, d_{k-h})$ where (d_k) is as in (30) or (32). Also V_0 is a small neighborhood of $\vartheta_0 \boldsymbol{d}_{k_0}/\|\boldsymbol{d}_{k_0}\|$ as in (27) for some $\vartheta_0 \in S_1$, $k_0 \in \{-m, \ldots, h\}$, and $B(V_0) = V_0 \times \mathbb{R}^h$. Then,*

$$\mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t}, A \Big| B(V_0)\Big) \underset{x \to \infty}{\longrightarrow} \begin{cases} 1, & \text{if} \quad \dfrac{\vartheta_0 \boldsymbol{d}_{k_0}}{\|\boldsymbol{d}_{k_0}\|} \in A, \\ 0, & \text{otherwise,} \end{cases}$$

for any closed neighborhood $A \subset C_{m+h+1}^{\|\cdot\|}$ such that

$$\partial A \cap \{\vartheta \boldsymbol{d}_k/\|\boldsymbol{d}_k\| : \ \vartheta \in S_1, \ k \in \{-m, \ldots, h\}\} = \emptyset.$$

Remark 3 Contrary to the anticipative AR(1), the trajectories of the anticipative AR(2) and fractionally integrated processes do not leave room for indeterminacy of the future path. Asymptotically, given any observed path of length at least 2, the shape of the future trajectory can be deduced deterministically. This holds even if the peak/collapse of a bubble is not yet present in the observed piece of trajectory. Therefore, provided the current pattern is properly identified,[13] it appears possible in the framework of these models to infer in

[13] See point (ι) of Remark 2.

advance the peak and crash dates of bubbles with very high confidence, in principle, with certainty.

5 Forecasting climate anomalies

Providing a forecast of El Niño (and La Niña) weather shocks is of primary importance, as it offers numerous societal benefits, ranging from extreme weather warnings to agricultural planning (Alley et al. 2019). Furthermore, there is an extensive literature documenting the significant economic effects of these climate anomalies, including their impact on primary commodity prices (Brenner 2002), macroeconomic fluctuations (Cashin et al. 2017), and broader economic development outcomes (Dell et al. 2014). El Niño (or La Niña) intensity is defined as a value constructed from the Southern Oscillation Index (SOI).[14] This section discusses the performance of our forecasting procedures in detecting the peak and the end of an El Niño (and La Niña) shock. We split the data into an in-sample period (from 01/1951 to 12/1991) and an out-of-sample period (from 01/1992 to 01/2024) to test the robustness of our forecasting procedure. Figure 5 in Appendix B.1 of the Supplementary Material displays the data sample, where the shaded area corresponds to the out-of-sample data. The alternation of boom and bust, which appears to be an identifiable pattern, is clearly distinguishable. Following the analysis of the ACF and PACF (see Figure 7 in Appendix B.1 of the Supplementary Material), we estimate an anticipative AR(2) model for the in-sample SOI data using the Generalized Covariance (GCoV) estimator of Gourieroux & Jasiak (2023) and using the alternative AML methods of Lanne & Saikkonen (2011), adapted for α-stable laws. Both approach confirm the retained specification, a purely anticipative AR(2), and the coefficients estimated are similar.[15]

[14] Data and methodology for constructing the SOI are available at https://www.ncei.noaa.gov/access/monitoring/enso/soi. The SOI is a monthly variable derived from air-pressure differentials in the South Pacific, measured between Tahiti and Darwin.

[15] The detailed estimation methodology, including the model selection procedure and alternative estimation methods, is presented in Appendix B of the Supplementary Material.

5.1 Forecasting reversal dates

A La Niña (resp. El Niño) episode is defined as the SOI exceeding 1 (resp. -1) for at least three consecutive periods. We hence take advantage of Proposition 6 to predict the reversal date of the El Niño occurrence that presumably starts at the end of the in-sample period. As this last observation is below -1, we admit that x is far from central values. We focus on the case where k belongs to $\{-m, \ldots, -1\}$ and recall that for large values of x and sufficiently persistence anticipative process, the approximation

$$(X_{t-m}, \ldots, X_{t-1}, X_t)/\|\boldsymbol{X_t}\| \approx \vartheta(d_{k+m}, \ldots, d_{k+1}, d_k)/\|\boldsymbol{d}_k\|, \quad \boldsymbol{X_t} = (X_{t-m}, \ldots, X_{t-1}, X_t),$$

can be used to derive the next crash date and then estimate the future path up to $t+h$. As for a range of realizations, we ignore to which piece of the moving average trajectory it corresponds, we pay particular attention to the selection of k_0 and the impact of m. ϑ_0 is assumed to be known here as in general it can be deduced from the data (positive or negative extreme event). Our forecasting procedure proceeds in 4 steps. First, we compute $(X_{t-m}, \ldots, X_{t-1}, X_t)/\|\boldsymbol{X_t}\|$ for a given $m \geq 1$. Second, we evaluate $\vartheta \boldsymbol{d}_k/\|\boldsymbol{d}_k\|$ for $k \in (\underline{k}, \bar{k})$ where $\underline{k} = 30 > m$ and $\bar{k} = 0$. Third, we check whether some $\vartheta \boldsymbol{d}_k/\|\boldsymbol{d}_k\|$ belong to a small neighborhood A of $(X_{t-m}, \ldots, X_{t-1}, X_t)/\|\boldsymbol{X_t}\|$. If k_0 cannot be identified because several values k satisfy this condition, we reduce the neighborhood until a unique $k = k_0$ remains. At this stage, we structure $\boldsymbol{X_t}$ as in (14) and $\boldsymbol{d}_{k_0}$ as $(d_{k+m}, \ldots, d_k, d_{k-1}, \ldots, d_{k-h})$. The last step simply consists of using the deterministic trajectory of $\boldsymbol{d}_{k_0}$ to iterate up to $d_{k-h} = 0$ and hence obtain the bubble burst date. From Proposition 6 we know that if X_t is anticipative enough, its future path will follow the one of $\boldsymbol{d}_{k_0}$ with a very high level of confidence such that

$$\boldsymbol{X_t}/\|\boldsymbol{X_t}\| \approx \vartheta_0 \boldsymbol{d}_{k_0}/\|\boldsymbol{d}_{k_0}\|,$$

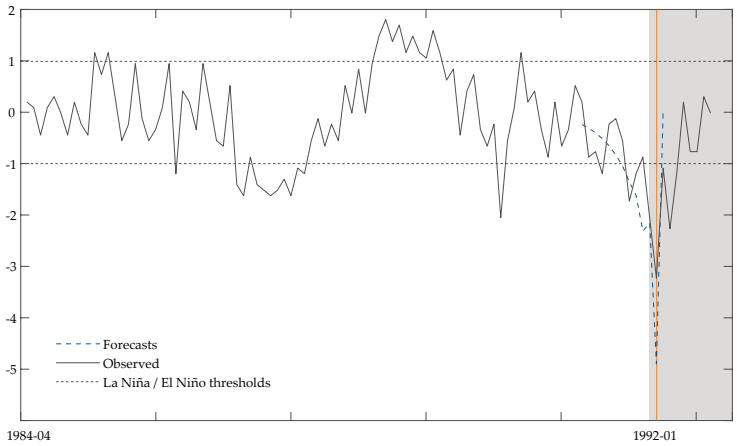
hence offering the possibility to predict $X_{t+1}, \ldots, X_{t+h}$. Finite sample performance of this 4-step forecasting procedure for crash date prediction, including a comprehensive bias analysis across multiple MAR specifications (MAR(0,2), MAR(1,2), and MAR(0,3)), and

systematic evaluation of parameter sensitivity (particularly for m and k_0 selection), are presented in Appendix A.1 of the supplementary material.[16] Following the simulation study presented in this appendix, we determine k_0 for various values of $m \in [1, 10]$. As $\hat{\alpha} \approx 1.9$ exhibits light tails, we are likely to encounter some difficulties in applying our pattern recognition procedure: far from the peak (m large) we might observe values from the center of the distribution. On the other hand, small m might lead to imprecise results as few past information is used to determine the piece of trajectory and the process is not strongly anticipative given the estimated coefficients. Our findings offer some robustness in this particular case as for $m = \{1, 2\}$ and $m \in [5, 10]$ our procedure always points toward $k_0 = 1$. For $m = 3$ and $m = 4$ we find $k_0 = 5$ and $k_0 = 3$ respectively. We hence retain $k_0 = 1$ and $m = 10$, therefore implying an imminent reversal date as we are close to the last piece of the trajectory described by $\vartheta_0 \boldsymbol{d}_{k_0}$, with $\vartheta_0 = -1$. The selected piece of the trajectory is represented in Figure 1. We then deduce the reversal date and we compute the future values of X_t up to $X_{t+h} = 0$, with $h = k_0 + 1$, that is when the SOI goes back to its central value. We find that El Niño should reverse just after February 1992, reaching a peak at $\hat{x}_{t+1} = -4.60$. When compared with the out-of-sample period, the reversal date appeared to be very accurately predicted. However, the magnitude of the peak reached during this El Niño occurrence is overestimated as $x_{t+1} = -3.04$.

To ensure the robustness of our approach, following the same procedure used to predict the reversal date in Figure 1, we predict all El Niño and La Niña anomalies in the out-of-sample dataset (from 01/1992 to 01/2024). The results are summarized in Table 1. For an El Niño (or La Niña) event, we start our forecasting procedure at the first date when the SOI is below -1 (or above 1) before the end date of an identified El Niño phenomenon, called the start date in Table 1; the end date is when the SOI returns between -1 and 0 (or 1 and 0). We

[16] This appendix quantifies how sources of uncertainty from Remark 2 affect finite-sample performance, demonstrating that bias decreases with sample size and varies systematically with the tail parameter α and lag selection m.

Figure 1: El Niño reversal forecast



Notes: The shaded area corresponds to the out-of-sample data.

also forecast the peak date, defined as the minimum (or maximum) value of the SOI before the start and end dates. Table 1 shows that for all 14 El Niño and La Niña occurrences in the out-of-sample dataset, our procedure leads to an average error of 0.42 months in finding the peak date and 0.57 months in finding the end date compared to the true peak and end dates. We also report in Table 1 the selected k_0 and m from our procedure.

Table 1: Forecasting out-of-sample El Niño and La Niña anomalies

Type of anomaly	El Niño	El Niño	La Niña	El Niño	La Niña	La Niña	El Niño	La Niña	La Niña	La Niña	La Niña
Start date	12/1991	07/1994	11/2007	12/2009	07/2010	11/2010	07/2015	11/2021	02/2022	08/2022	11/2022
Peak date	01/1992	09/1994	02/2008	02/2010	09/2010	12/2010	10/2015	01/2021	03/2022	10/2022	12/2022
End date	04/1992	10/1994	03/2008	03/2010	11/2010	04/2011	11/2015	03/2021	05/2022	11/2022	02/2023
Forecasted Peak	01/1992	09/1994	02/2008	03/2010	08/2010	01/2011	09/2015	01/2021	04/2022	10/2022	01/2023
Forecasted End	02/1992	10/1994	03/2008	04/2010	09/2010	02/2011	10/2015	02/2021	05/2022	11/2022	02/2023
Peak forecast error	0	0	0	1	-1	1	-1	0	1	0	-1
End forecast error	-2	0	0	1	-1	-2	-1	-1	0	0	0
k_0	1	2	3	3	1	2	2	2	2	2	2
m	10	10	10	9	10	10	10	10	10	10	10

6 Conclusion

This paper develops a spectral representation theory for α-stable random vectors on unit cylinders defined by seminorms, extending the classical representation on unit spheres. We establish necessary and sufficient conditions for the existence of such representations. For two-sided α-stable moving averages $(X_t) = \sum_{k\in\mathbb{Z}} d_k \varepsilon_{t+k}$, we characterize when trajectory vectors $\boldsymbol{X}_t = (X_{t-m}, \ldots, X_{t+h})$ are representable. In particular, we show that non-anticipative or pure causal processes are non-representable. For non-causal and mixed-causal processes, we derive the asymptotic conditional distribution of normalized – by an appropriate seminorm – future paths given large observed trajectories. For anticipative AR($p \geq 2$) and fractionally integrated processes, this conditional distribution degenerates to a single Dirac mass, yielding deterministic prediction of extreme trajectories. We use Monte Carlo simulations to demonstrate the ability of our forecasting theory to predict reversal dates. The numerical analysis confirms that our procedures perform well in finite sample across a wide range of scenarios. To provide further insight into the empirical relevance of the seminorm representation of α-stable moving averages, we demonstrate its ability to accurately predict climate anomalies. Specifically, we predict, out-of-sample, the reversal date of La Niña and El Niño episodes episodes.

7 Disclosure statement

The authors declare that they have no conflicts of interest.

8 Data Availability Statement

To replicate our results and demonstrate the generality of our approach, we developed a web application enabling users to reproduce our findings and apply our methods to other time series in macroeconomics, finance, and climate science. The application is available at https://marforecast.streamlit.app/?utm_medium=oembed. SOI data and methodology are

available at https://www.ncei.noaa.gov/access/monitoring/enso/soi.

SUPPLEMENTARY MATERIALS

Supplementary Material This supplement contains Monte Carlo simulations and numerical analysis in Appendix A and complementary empirical results in Appendix B. The proofs are collected in Appendix C.

Reproduction code This supplement contains code (Python and R) and data to reproduce all the analyses and figures in this paper.

References

Alley, R. B., Emanuel, K. A. & Zhang, F. (2019), 'Advances in weather prediction', *Science* **363**(6425), 342–344.

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

Basrak, B., Planinić, H. & Soulier, P. (2016), 'An invariance principle for sums and record times of regularly varying stationary sequences', *Probability Theory and Related Fields* pp. 1–46.

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

Bec, F., Nielsen, H. B. & Saïdi, S. (2020), 'Mixed causal–noncausal autoregressions: Bimodality issues in estimation and unit root testing', *Oxford Bulletin of Economics and Statistics* **82**(6), 1413–1428.

Blasques, F., Koopman, S. J., Mingoli, G. & Telg, S. (2025), 'A novel test for the presence of local explosive dynamics', *Journal of Time Series Analysis* **46**(5), 966–980.

Brenner, A. D. (2002), 'El niño and world primary commodity prices: Warm water or hot air?', *The Review of Economics and Statistics* **84**(1), 176–183.

Brockwell, P. & Davis, R. (1991), *Time Series: Theory and Methods*, Springer series in statistics, Springer.

Buraczewski, D., Damek, E., Mikosch, T. & Zienkiewicz, J. (2013), 'Large deviations for solutions to stochastic recurrence equations under Kesten's condition', *The Annals of Probability* **41**(4), 2755 – 2790.

Cashin, P., Mohaddes, K. & Raissi, M. (2017), 'Fair weather or foul? the macroeconomic effects of el niño', *Journal of International Economics* **106**, 37–54.

Cavaliere, G., Nielsen, H. B. & Rahbek, A. (2020), 'Bootstrapping noncausal autoregressions: With applications to explosive bubble modeling', *Journal of Business & Economic Statistics* **38**(1), 55–67.

Chahrour, R. & Jurado, K. (2021), 'Recoverability and expectations-driven fluctuations', *The Review of Economic Studies* **89**(1), 214–239.

Chen, B., Choi, J. & Escanciano, J. C. (2017), 'Testing for fundamental vector moving average representations', *Quantitative Economics* **8**, 149–180.

Cioczek-Georges, R. & Taqqu, M. S. (1994), 'How do conditional moments of stable vectors depend on the spectral measure?', *Stochastic processes and their applications* **54**, 95–111.

Cioczek-Georges, R. & Taqqu, M. S. (1998), 'Sufficient conditions for the existence of conditional moments of stable random variables', *Stochastic Processes and Related Topics* pp. 35–67.

Davis, R. A. & Hsing, T. (1995), 'Point Process and Partial Sum Convergence for Weakly Dependent Random Variables with Infinite Variance', *The Annals of Probability* **23**(2), 879 – 917.

Davis, R. A. & Mikosch, T. (2009), 'The extremogram: A correlogram for extreme events', *Bernoulli* **15**(4), 977 – 1009.

Dell, M., Jones, B. F. & Olken, B. A. (2014), 'What Do We Learn from the Weather? The New Climate-Economy Literature', *Journal of Economic Literature* **52**(3), 740–798.

Dombry, C., Hashorva, E. & Soulier, P. (2017), 'Tail measure and tail spectral process of regularly varying time series', *arXiv preprint arXiv:1710.08358* .

Dombry, C., Hashorva, E. & Soulier, P. (2018), 'Tail measure and spectral tail process of regularly varying time series', *The Annals of Applied Probability* .

Dombry, C., Tillier, C. & Wintenberger, O. (2022), 'Hidden regular variation for point processes and the single/multiple large point heuristic', *The Annals of Applied Probability* **32**(1), 191 – 234.

Fries, S. (2022), 'Conditional moments of noncausal alpha-stable processes and the prediction of bubble crash odds', *Journal of Business & Economic Statistics* **40**(4), 1596–1616.

Fries, S. & Zakoian, J.-M. (2019), 'Mixed causal-noncausal ar processes and the modelling of explosive bubbles', *Econometric Theory* **35**(6), 1234–1270.

Giancaterini, F. (2023), Essays on univariate and multivariate noncausal processes, PhD thesis, Maastricht University, Netherlands.

Giancaterini, F., Hecq, A., Jasiak, J. & Neyazi, A. M. (2025), 'Bubble detection with application to green bubbles: A noncausal approach'.

Giancaterini, F., Hecq, A. & Morana, C. (2022), 'Is climate change time-reversible?', *Econometrics* **10**(4).

Gourieroux, C., Hencic, A. & Jasiak, J. (2021), 'Forecast performance and bubble analysis in noncausal mar (1, 1) processes', *Journal of Forecasting* **40**(2), 301–326.

Gourieroux, C. & Jasiak, J. (2023), 'Generalized covariance estimator', *Journal of Business & Economic Statistics* **41**(4), 1315–1327.

Gourieroux, C. & Jasiak, J. (2025), 'Nonlinear fore(back)casting and innovation filtering for causal-noncausal var models'.

Gourieroux, C., Jasiak, J. & Monfort, A. (2020), 'Stationary bubble equilibria in rational expectation models', *Journal of Econometrics* **218**(2), 714–735.

Gourieroux, C., Jasiak, J. & Tong, M. (2021), 'Convolution-based filtering and forecasting: An application to wti crude oil prices', *Journal of Forecasting* **40**(7), 1230–1244.

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

Gouriéroux, C. & Jasiak, J. (2017), 'Noncausal vector autoregressive process: Representation, identification and semi-parametric estimation', *Journal of Econometrics* **200**, 118–134.

Gouriéroux, C., Lu, Y. & Robert, C.-Y. (2025), 'The causal-noncausal tail processes: An introduction'.

Gouriéroux, C., Monfort, A. & Renne, J.-P. (2019), 'Identification and estimation in non-fundamental structural varma models', *The Review of Economic Studies* **87**(4), 1915–1953.

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.

Gouriéroux, C. & Zakoian, J.-M. (2017), 'Local explosion modelling by non-causal process', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79**, 737–756.

Hecq, A., Issler, J. V. & Telg, S. (2020), 'Mixed causal–noncausal autoregressions with exogenous regressors', *Journal of Applied Econometrics* **35**(3), 328–343.

Hecq, A., Lieb, L. & Telg, S. M. (2016), 'Identification of mixed causal-noncausal models in finite samples', *Annals of Economics and Statistics* **123/124**, 307–331.

Hecq, A., Telg, S. & Lieb, L. (2017*a*), 'Do seasonal adjustments induce noncausal dynamics in inflation rates?', *Econometrics* **5**, 48.

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

Hecq, A. & Velasquez-Gaviria, D. (2025), 'Non-causal and non-invertible arma models: Identification, estimation and application in equity portfolios', *Journal of Time Series Analysis* **46**(2), 325–352.

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

Hencic, A. & Gouriéroux, C. (2015), 'Noncausal autoregressive model in application to bitcoin/usd exchange rates', *Econometrics of Risk* pp. 17–40.

Janßen, A. (2019), 'Spectral tail processes and max-stable approximations of multivariate regularly varying time series', *Stochastic Processes and their Applications* **129**(6), 1993–2009.

Janßen, A. & Segers, J. (2014), 'Markov tail chains', *Journal of Applied Probability* **51**, 1133–1153.

Jasiak, J. & Neyazi, A. M. (2025), 'Gcov-based portmanteau test'.

Konstantinides, D. G. & Mikosch, T. (2005), 'Large deviations and ruin probabilities for solutions to stochastic recurrence equations with heavy-tailed innovations', *The Annals of Probability* **33**(5), 1992 – 2035.

Kulik, R. & Soulier, P. (2020), *Heavy-tailed time series*, Springer.

Lanne, M. & Luoto, J. (2016), 'Noncausal bayesian vector autoregression', *Journal of Applied Econometrics* **31**(7), 1392–1406.

Lanne, M., Luoto, J. & Saikkonen, P. (2012), 'Optimal forecasting of noncausal autoregressive time series', *International Journal of Forecasting* **28**(3), 623–631.

Lanne, M. & Saikkonen, P. (2011), 'Noncausal autogressions for economic time series', *Journal of Time Series Econometrics* **3**.

Lanne, M. & Saikkonen, P. (2013), 'Noncausal vector autoregression', *Econometric Theory* **29**, 447–481.

Meinguet, T. & Segers, J. (2010), 'Regularly varying time series in banach spaces', *arXiv preprint arXiv:1001.3262* .

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.

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

Planinić, H. & Soulier, P. (2017), 'The tail process revisited', *Extremes* pp. 1–29.

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

Velasco, C. & Lobato, I. N. (2018), 'Frequency domain minimum distance inference for possibly noninvertible and noncausal ARMA models', *The Annals of Statistics* **46**(2), 555 – 579.

SUPPLEMENTARY MATERIAL TO "FORECASTING EXTREME TRAJECTORIES USING SEMINORM REPRESENTATIONS"

A 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 A.1 and reversal probability prediction in Appendix A.2. Furthermore, Appendix A.3 illustrates the unit cylinder geometry through numerical simulations, while Appendix A.4 presents the associated directional histograms that visualize the concentration patterns in extreme observations.

A.1 Forecasting crash dates

One can apply Proposition 6 to infer information on future paths from the observed "persistant" trajectory, as long as it deviates far enough from central values. We focus on the case where k belongs to $\{-m, \dots, -1\}$ and document that in practice, for finitely large values of x and sufficiently persistence anticipative process, the approximation

$$(X_{t-m}, \dots, X_{t-1}, X_t)/\|\boldsymbol{X_t}\| \approx \vartheta(d_{k+m}, \dots, d_{k+1}, d_k)/\|\boldsymbol{d}_k\|,$$

can be used to derive the next crash date and then estimate the future path up to $t+h$. We also discuss to what extent the sources of uncertainty listed in Remark 2 affect the performance of our procedure in the presence of finitely large realizations. As for a range of realizations, we ignore to which piece of the moving average trajectory it corresponds, we pay particular attention to the selection of k_0 and the impact of m. ϑ_0 is assumed to be known here as in general it can be deduced from the data.

Our forecasting procedure proceeds in 4 steps. First, we compute $(X_{t-m}, \dots, X_{t-1}, X_t)/\|\boldsymbol{X_t}\|$ for a given $m \geq 1$. Second, we evaluate $\vartheta \boldsymbol{d}_k/\|\boldsymbol{d}_k\|$ for $k \in (\underline{k}, \bar{k})$ where $\underline{k} = 30$ and $\bar{k} = 0$. Third, we check whether some $\vartheta \boldsymbol{d}_k/\|\boldsymbol{d}_k\|$ belong to a small neighborhood A of

$(X_{t-m},\dots,X_{t-1},X_t)/\|\boldsymbol{X_t}\|$. If k_0 cannot be identified because several values k satisfy this condition, we reduce the neighborhood until a unique $k=k_0$ remains. The last step simply consists of using the deterministic trajectory of $\boldsymbol{d}_{k_0}$ to iterate up to $d_{k-h}=0$ and hence obtain the bubble burst date. From Proposition 6 we know that if X_t is anticipative enough, its future path will follow the one of $\boldsymbol{d}_{k_0}$ with a very high level of confidence such that

$$\boldsymbol{X_t}/\|\boldsymbol{X_t}\| \approx \vartheta_0 \boldsymbol{d}_{k_0}/\|\boldsymbol{d}_{k_0}\|,$$

hence offering the possibility to predict $X_{t+1},\dots,X_{t+h}$. This procedure is likely to be sensible to the selection of m. We investigate this issue by considering $m=\{1,3,5,7,9,11\}$. We also anticipate that, how far we deviate from the Gaussian distribution, in terms of tail index, is likely to affect the results, and hence we consider $\alpha=\{0.9,1.2,1.5,1.8\}$ (see Remark 2 ($\iota\iota\iota$)). We simulate the three following processes: MAR(0,2), MAR(1,2), MAR(0,3). The MAR(1,1) process was ruled out because Proposition 6 does not apply directly, as we do not obtain certainty in the path but rather some probabilities. We first present the results corresponding to the purely anticipative AR(2) in Table 4, as it is the main specification in the empirical section 5. Let start by presenting the simulation procedure. Each simulated path is governed by a SαS anticipative AR(2) of the following form: $X_t = 0.7X_{t+1} + 0.1X_{t+2} + \varepsilon_t$ where $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha,0,0.5,0)$. For a given artificial time series x_t, we identify a positive bubble peak as $max(x_t)$ and treat as unobserved the remaining values of the series and the $\lceil N\times 0.01\rceil$ periods preceding the bubble burst. We then explore all these scenarios for $N=\{250,500,1000\}$ (i.e. $k_0=\{3,5,10\}$) and 1000 trajectories. In theory, the sample size N should not impact the prediction performance but we use it here to control the quantile of the last in-sample observation. More precisely, our simulation framework results in the quantiles reported in Table 3 and allows us to investigate the impact of departing from the asymptotic theory ($x\to\infty$). For instance, we can see that for $N=1000$, the last in-sample observation used to predict an extreme event that surge 10 periods ahead, actually corresponds to the quantile 0.91 when $\alpha=1.5$. In such a configuration, the realizations of X_t are likely to be only moderately large compared

to the asymptotic requirements ($x \to \infty$).

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

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, 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)

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.

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.

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.

A.2 Forecasting reversal probabilities

Our main theoretical result (Proposition 6) demonstrates that for sufficiently anticipative processes such as the AR(2), trajectory forecasting becomes deterministic, conditional on the precise identification of the past pattern, indexed by k_0. 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 probability of reversal $\hat{p}_{q,\vartheta_0}(h)$ using the relative frequency over the simulation sample:

$$\hat{p}_{q,\vartheta_0}(h) = \frac{\sum_{t=1}^{N-h} \mathbb{1}_{\{t \in R_h(\delta) \cap C_{\vartheta_0}(q)\}}}{\sum_{t=1}^{N-h} \mathbb{1}_{\{t \in C_{\vartheta_0}(q)\}}} \tag{34}$$

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

Table 7: 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 (34) 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 8, 9, and 10, respectively.

Table 8: 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 (34) 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 $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(1.5, 0, 0.5, 0)$

The results for the MAR(1,1) process, presented in Table 8, 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 7). This empirical finding aligns directly with the theoretical predictions outlined in Proposition 5 (Case 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

Table 9: 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 (34) 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 $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(1.5, 0, 0.5, 0)$

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 8 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 more precise, deterministic forecasts achievable through pattern identification as demonstrated in Appendix A.1.

Table 10: 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 (34) 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 $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(1.5, 0, 0.5, 0)$

A.3 Visualisation of the unit cylinder

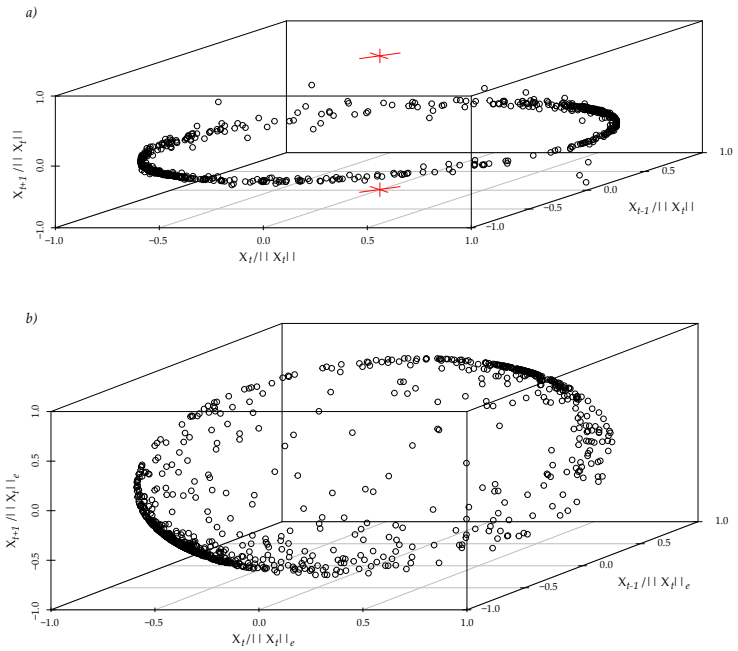
In the spirit of the Remark 1, we consider an α-stable vector $\boldsymbol{X_t} = (X_{t-1}, X_t, X_{t+1})$ where X_t is an anticipative AR(2) specified as in A.2. X_t being past-representable, it admits a representation on the unit-cylinder, as the Theorem 1 applies. Furthermore, as discussed in 6, its 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

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



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.

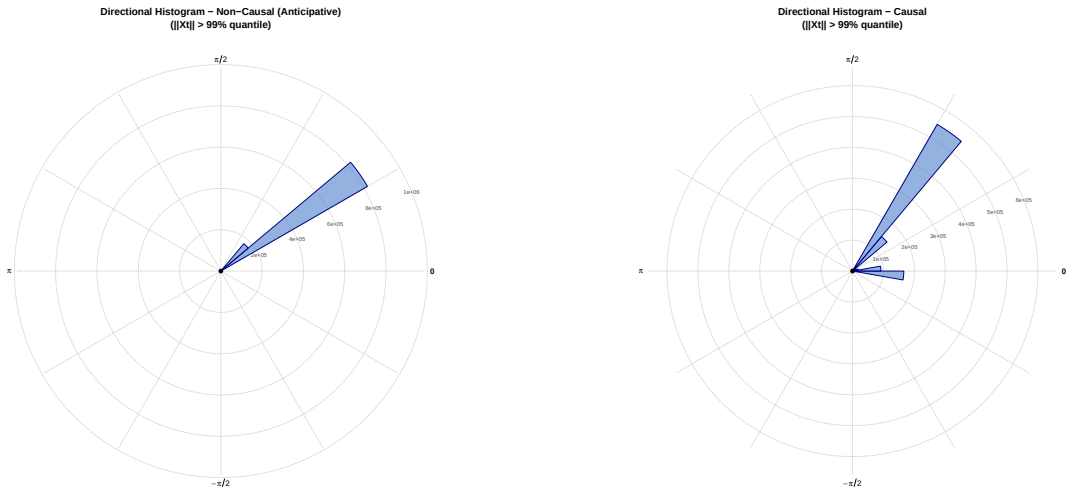
A.4 Directional histograms

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:

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

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

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.



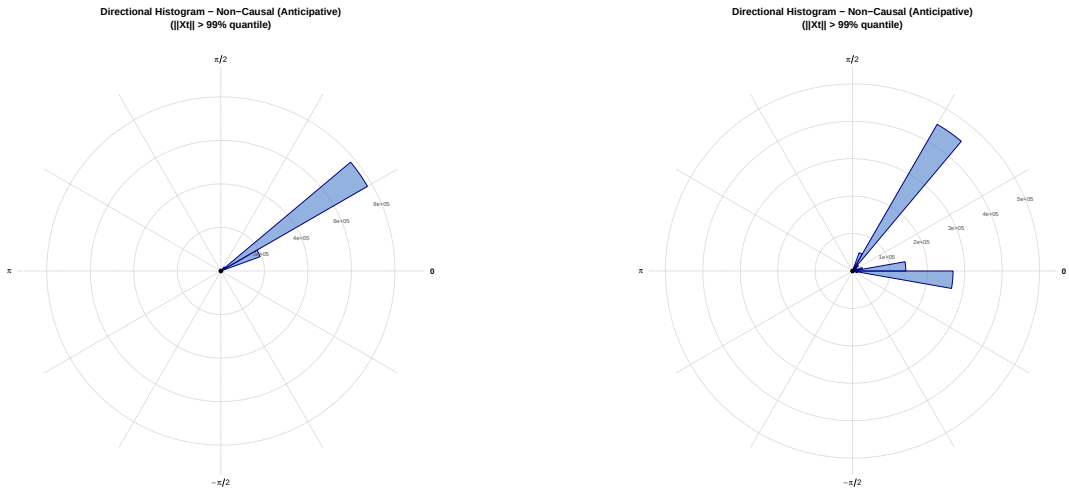
(a) MAR(0,2) (b) MAR(2,0)

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* $\arctan(X_{t-1}/X_t)$.

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 (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.

B 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 & Jasiak (2023), presented in Appendix B.2, and the parameter estimates obtained using the alternative approximate maximum likelihood (AML) method of Lanne & Saikkonen (2011), adapted to α-stable distributions, presented in Appendix B.3.

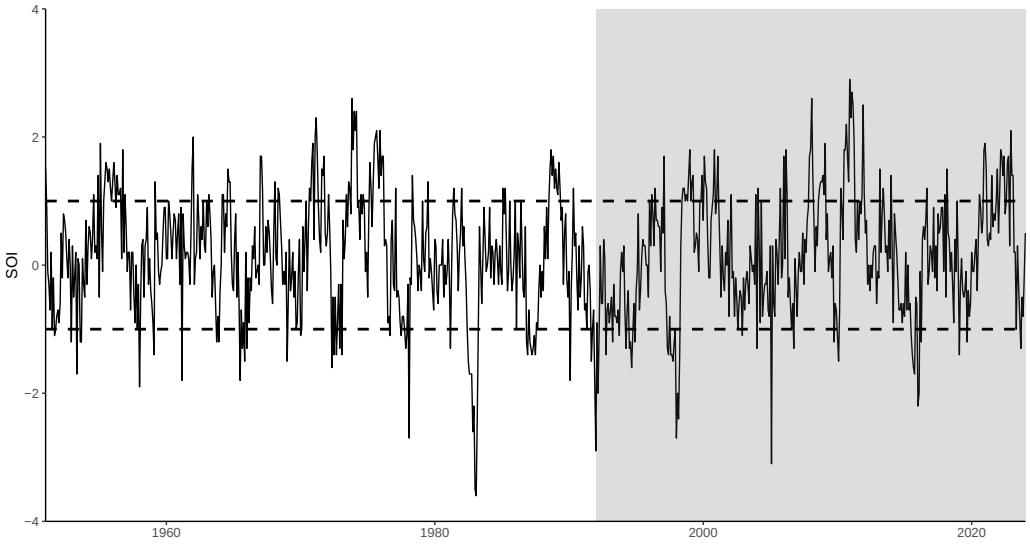
B.1 Distributional Properties and Autocorrelation Structure of the SOI

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.

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 11, 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 7 displays the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) for the in-sample SOI data. The PACF exhibits significant spikes at lags 1 and 2, suggesting an autoregressive structure of order 2. Based on this analysis, we set $p_{\text{ap}} =$

Figure 5: Southern Oscillation Index (SOI)



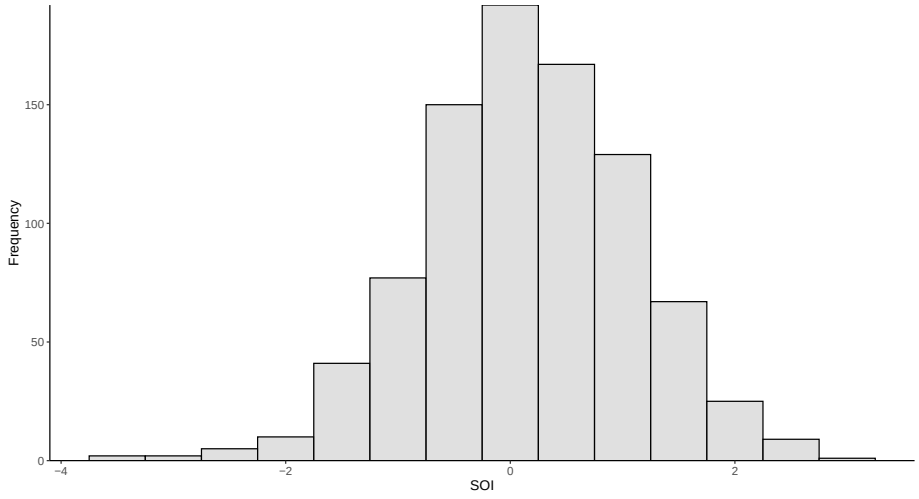
Notes: The shaded area represents the out-of-sample data. The horizontal black dashed lines indicate values of 1 *and* -1*. An El Niño event is defined as an SOI below* -1*, while a La Niña event corresponds to an SOI above* 1*, each persisting for at least three periods.*

$p + q = 2$. This specification is further validated using the Bayesian Information Criterion (BIC), as reported in Table 13.

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 & Jasiak 2023), referred to as GCoV, and the approximate maximum likelihood method (AML), introduced by Lanne & 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.[17] Gourieroux & Jasiak (2023) shows that the consistency and asymptotic properties of the semi-parametric estimator require the existence

[17] 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.

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.

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 moments exist. However, as pointed out in Remark 2 on p. 1318 of Gourieroux & 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 B.2 and include the AML results in the subsection B.3. Nonetheless, the retained specification—a purely anticipative AR(2)—and the coefficients estimated using AML are similar.

B.2 GCoV Estimation of Gourieroux & Jasiak (2023)

Let us begin by recalling the $MAR(p, q)$ model from Lanne & Saikkonen (2011), which imposes a multiplicative representation of the two-sided alpha-stable MA(∞) form in

Table 11: 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

equation (11).[18] 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:

$$\Psi(F)\Phi(B)X_t = \varepsilon_t, \qquad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, \beta, \sigma, 0), \tag{35}$$

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

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

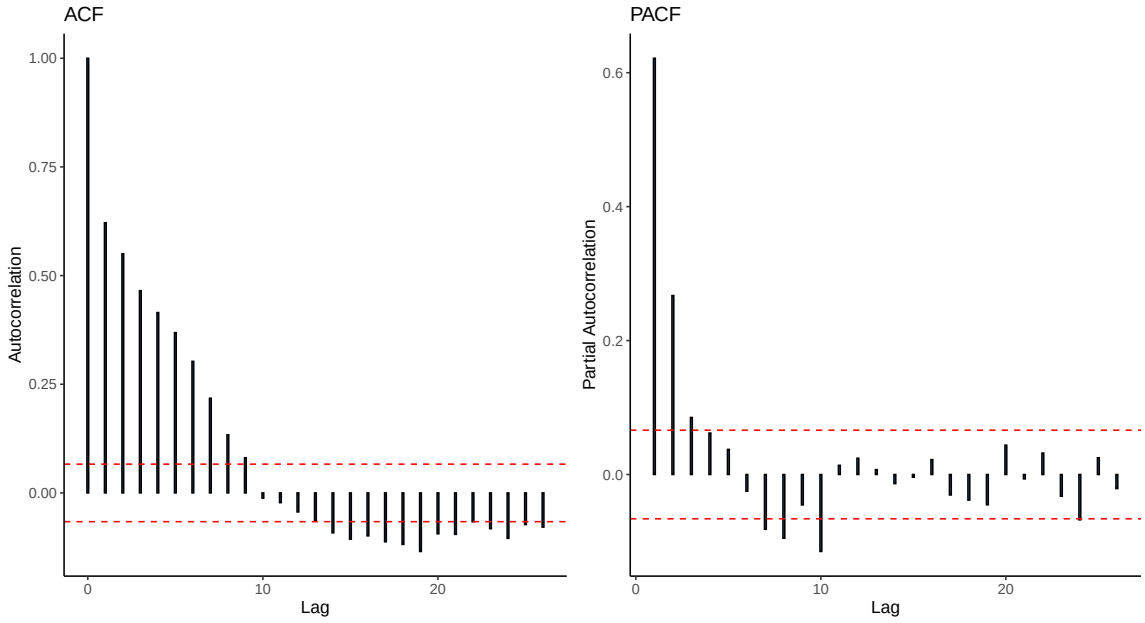
This assumption ensures stationarity in both the causal and noncausal components. A key challenge in this context is the identification of p and q. One solution, proposed by Lanne & Saikkonen (2011) and Hecq et al. (2017*b*, 2020), is to first select $p_{ap} = p + q$ by relying on the ACF and PACF of the SOI in the training sample, and then confirm the choice using the BIC criterion.[19]

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

[18] In the approach of Gourieroux & 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).

[19] 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 & Zakoian 2019).

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



to 2 in this case. We also apply standard information criteria, such as BIC, to validate this choice.[20] Once p_{ap} is selected, we follow the procedure outlined by Lanne & Saikkonen (2011) and (see 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.[21] 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

[20] Table 13 confirms the selection of 2.

[21] See again Table 13 , which confirms the choice of purely anticipative AR(2)

(35), where Θ is the entire parameter space, the GCoV estimator minimizes the following portmanteau statistic:

$$\hat{\theta} = \underset{\Theta}{\operatorname{argmin}} \sum_{h=1}^{\tilde{H}} Tr\left[\hat{\gamma}_g(h;\theta)\hat{\gamma}_g(0;\theta)^{-1}\hat{\gamma}_g(h;\theta)'\hat{\gamma}_g(0;\theta)^{-1}\right] \tag{36}$$

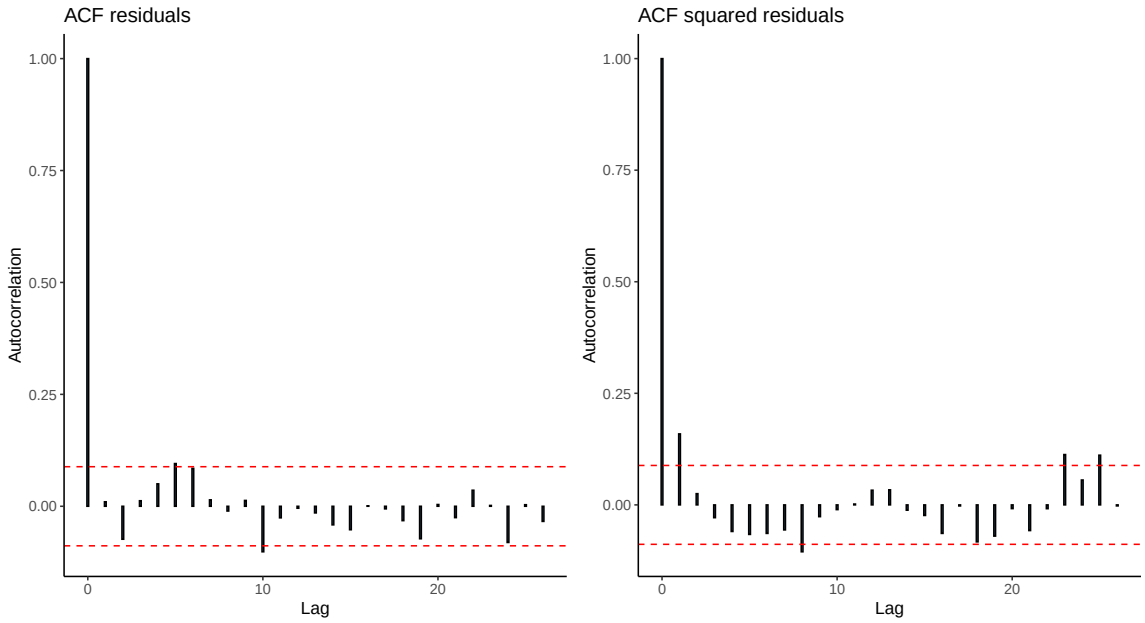
where $\tilde{H}$ is the highest selected lag, $\hat{\gamma}_g(h;\theta)$ is the sample autocovariance between $g(\varepsilon_t)$ and $g(\varepsilon_{t-h})$, with $g(\varepsilon_t) = [g_1(\varepsilon_t), \cdots, g_K(\varepsilon_t)]$, where ε_t, is the residuals obtained using equation (35).

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 & 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 & Jasiak (2023). If θ_0 represents the parameters estimated by GCoV, the standard errors can be computed as described in Gourieroux & 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 future research. Recall that the aim of the GCoV approach is to target i.i.d. residuals.[22] We test all $2^{p_{ap}} - 1$ possible combinations for MAR(r, s) with different combinations of

[22]The residuals are extracted from the SOI data using the estimated parameters and the equation (35)

$H = \{1, 2, 3\}$ 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 8 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 12. Figure 8 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 12).

Figure 8: ACF of Residuals and Squared Residual using GCoV



However, we also implement the Portmanteau Test from Jasiak & Neyazi (2025), 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 12). Table 12 presents the estimated parameters for ψ_1 and ψ_2, which closely resemble those obtained using the AML approach, as detailed in Table 14.

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 12: 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 & Neyazi (2025)'s test	5.5933	12.5915	0.4702
JB-Test	21.261	5.9900	0.0000

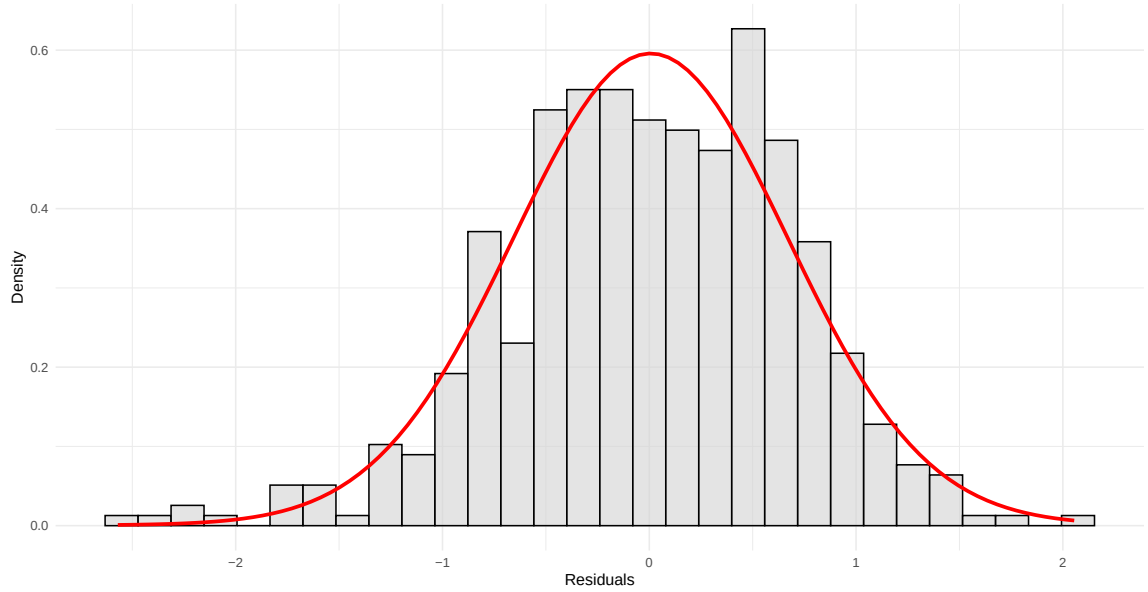
Notes: Estimated parameters of the α-stable anticipative AR(2) process associated with the SOI series for the period 01/1951–12/1991. Standard deviations are provided in parentheses. The specification tests are as follows: the Ljung-Box test on residuals, the Ljung-Box test on the squared residuals, the Portmanteau test of Jasiak & Neyazi (2025), and a non-Gaussianity test.

Table 12 confirms that the residuals are indeed non-Gaussian. Next, we fit an α-stable distribution $\mathcal{S}(\alpha, \beta, \sigma, 0)$ to the residuals using the characteristic function-based estimation for α-stable distributions described in Nolan (2021). The estimated parameters are reported in Table 12. These results are consistent with those obtained using the AML approach (see Table 14). Figure 9, 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.

B.3 AML estimation of Lanne & Saikkonen (2011)

This section, presents the results on in-sample SOI data using the AML estimation procedure from, Lanne & Saikkonen (2011) and Hecq et al. (2020, 2017*b*). As is indicated by the Figure 7, we set $p_{ap} = 2$, and we estimate all the possible $2^{p_{ap}}) - 1$ combinaison of $MAR(p, q)$.

Figure 9: 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: $\alpha = 1.9754$, $\beta = -0.0216$, $\sigma = 0.4735$.

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*b*). For the $MAR(p,q)$ as in eq (35), the ensemble of parameters to be estimated is $\theta = (\Psi, \Phi, \alpha, \beta, \sigma) \in \Theta$ then the Lanne & Saikkonen (2011) AML estimator, is defined as:

$$\hat{\theta} = \underset{\Theta}{\operatorname{argmax}} \sum_{t=p+1}^{T-q} \ln f^{-1} \left[\Psi(F)\Phi(B)X_t; (\alpha, \beta, \sigma, 0) \right] \tag{37}$$

where $f(.;(\alpha,\beta,\sigma,0))$ denotes the probability function of the ε_t.[23] Standard deviations are estimated using finite differences gradient and Hessian for the parameters in the right space. Table 13, according to the likelihood criteria, shows that the best specification for causal-non-causal models is the anticipative AR(2)

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

[23] Lanne & 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 13: 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

Table 14 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 14, 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 residuals, shown in Figure 10. However, Figure 10 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 14.

The Jarque-Bera (JB) test in Table 14 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 11.

In Figure 10, 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

Table 14: 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.

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 11, 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 the underlying distribution of the residuals.

Figure 10: ACF of Residuals and Squared Residual using AML

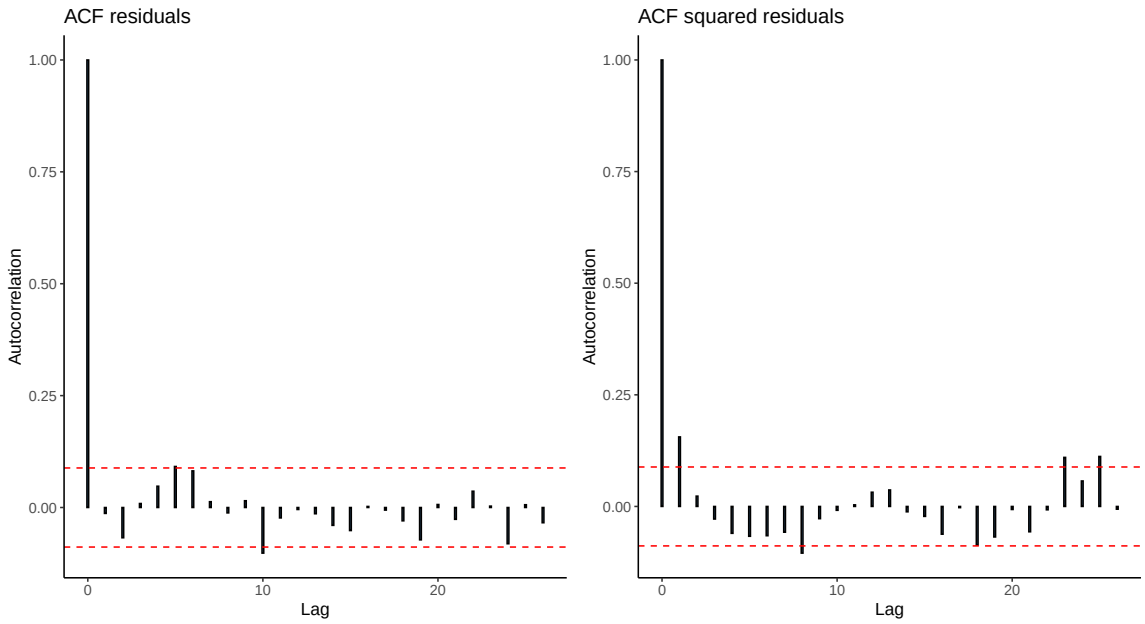
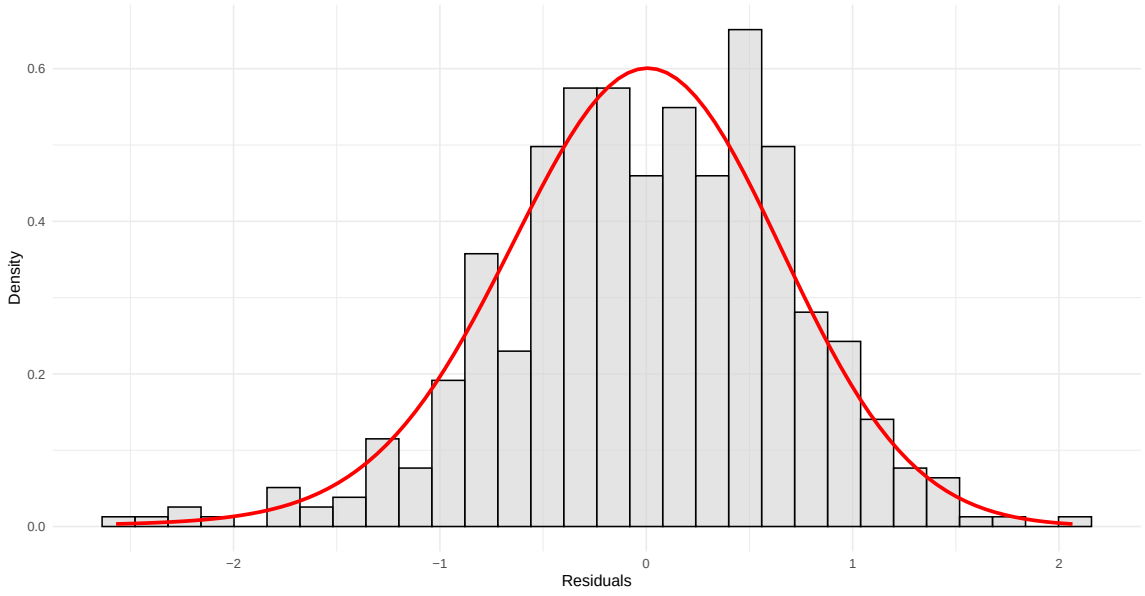


Figure 11: Histogram of residuals using AML



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: $\alpha = 1.9390$, $\beta = -0.9970$, $\sigma = 0.4696$.

C Proofs

C.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 $\boldsymbol{m}_{\|\cdot\|}^0 \in \mathbb{R}^d$ such that

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

With the change of variable $\boldsymbol{s}' = T_{\|\cdot\|}^{-1}(\boldsymbol{s}) = \boldsymbol{s}/\|\boldsymbol{s}\|_e$,

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

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

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

By the unicity of the spectral representation of $\boldsymbol{X}$ on S_d, we necessarily have $(\Gamma, \boldsymbol{\mu}^0) = (\overline{\gamma}, \boldsymbol{m}_{\|\cdot\|}^0)$. 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 $a = 2/\pi$,

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

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,

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

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

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

Reciprocally, assume there exists a measure $\gamma^{\|\cdot\|}$ on $C_d^{\|\cdot\|}$ satisfying (7) and a non-random vector $\boldsymbol{m}^0_{\|\cdot\|} \in \mathbb{R}^d$ such that

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

First, we can see that

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

We will later show the following result:

Lemma 2 *Let $\gamma^{\|\cdot\|}$ a Borel measure on $C_d^{\|\cdot\|}$ satisfying (7). Then,*

$$\int_{C_d^{\|\cdot\|}} \|\boldsymbol{s}\|_e\Big|\ln\|\boldsymbol{s}\|_e\Big|\gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty. \tag{38}$$

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

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

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 $\tilde{\boldsymbol{m}} := (\tilde{m}_i)$ 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

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

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

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

$$\tilde{m}_i = \int_{S_d \setminus K^{\|\cdot\|}} s_i \ln \|\boldsymbol{s}\| \Gamma(d\boldsymbol{s}), \quad i = 1, \ldots, d.$$

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 (7), we know that for all $\boldsymbol{u} \in \mathbb{R}^d$

$$\int_{C_d^{\|\cdot\|}} |\langle \boldsymbol{u}, \boldsymbol{s}\rangle| \Big| \ln |\langle \boldsymbol{u}, \boldsymbol{s}\rangle| \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) = \int_{C_d^{\|\cdot\|} \cap \{b \leq \|\boldsymbol{s}\|_e \leq M\}} + \int_{C_d^{\|\cdot\|} \cap \{\|\boldsymbol{s}\|_e > M\}} < +\infty. \qquad (39)$$

and thus, in particular

$$\int_{\{\boldsymbol{s}'\in C_d^{\|\cdot\|}:\ \|\boldsymbol{s}'\|_e>M\}} |\langle \boldsymbol{u},\boldsymbol{s}\rangle| \Big|\ln|\langle \boldsymbol{u},\boldsymbol{s}\rangle|\Big| \gamma^{\|\cdot\|}(d\boldsymbol{s})$$
$$= \int_{\{\boldsymbol{s}'\in C_d^{\|\cdot\|}:\ \|\boldsymbol{s}'\|_e>M\}} |\langle \boldsymbol{u},\boldsymbol{s}\rangle| \Big|\ln\|\boldsymbol{s}\|_e + \ln|\langle \boldsymbol{u}, \frac{\boldsymbol{s}}{\|\boldsymbol{s}\|_e}\rangle|\Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty. \tag{40}$$

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

$$\int_{\{\boldsymbol{s}'\in C_d^{\|\cdot\|}:\ \|\boldsymbol{s}'\|_e>M\}} |\langle \boldsymbol{u},\boldsymbol{s}\rangle| \Big|\ln\|\boldsymbol{s}\|_e + \ln|\langle \boldsymbol{u}, \frac{\boldsymbol{s}}{\|\boldsymbol{s}\|_e}\rangle|\Big| \gamma^{\|\cdot\|}(d\boldsymbol{s})$$
$$= \int_{\{\boldsymbol{s}'\in C_d^{\|\cdot\|}:\ \|\boldsymbol{s}'\|_e>M\}} |\langle \boldsymbol{u},\boldsymbol{s}\rangle| \Big|\ln\|\boldsymbol{s}\|_e\Big|\Big|1+\frac{\ln|\langle \boldsymbol{u},\boldsymbol{s}/\|\boldsymbol{s}\|_e\rangle|}{\ln\|\boldsymbol{s}\|_e}\Big| \gamma^{\|\cdot\|}(d\boldsymbol{s})$$
$$\geq \int_{\{\boldsymbol{s}'\in C_d^{\|\cdot\|}:\ \|\boldsymbol{s}'\|_e>M\}} |\langle \boldsymbol{u},\boldsymbol{s}\rangle| \Big|\ln\|\boldsymbol{s}\|_e\Big|\Big|1-\Big|\frac{\ln|\langle \boldsymbol{u},\boldsymbol{s}/\|\boldsymbol{s}\|_e\rangle|}{\ln\|\boldsymbol{s}\|_e}\Big|\Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) \tag{41}$$

Let us now partition the space $\mathbb{R}^d$ into subsets $R_1,\dots,R_d$ such that, for any $i=1,\dots,d$ and any $\boldsymbol{s}=(s_1,\dots,s_d)\in R_i$, $\sup_j |s_j| = |s_i|$.[24] We have by (40)-(41) that for any $i=1,\dots,d$, any $\boldsymbol{u}\in\mathbb{R}^d$,

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

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

$$\int_{\{\boldsymbol{s}'\in C_d^{\|\cdot\|}:\ \|\boldsymbol{s}'\|_e>M\}\cap R_i} |\langle \boldsymbol{e}_i,\boldsymbol{s}\rangle| \Big|\ln\|\boldsymbol{s}\|_e\Big|\Big|1-\Big|\frac{\ln|\langle \boldsymbol{e}_i,\boldsymbol{s}/\|\boldsymbol{s}\|_e\rangle|}{\ln\|\boldsymbol{s}\|_e}\Big|\Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty. \tag{42}$$

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

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

[24] Strictly speaking, $(R_1,\dots,R_d)$ 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.

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

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

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 $\left\| \boldsymbol{s}_n / \|\boldsymbol{s}_n\|_e \right\|_e = 1$ 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 $\left| \ln |\langle \boldsymbol{e}_i, \boldsymbol{s} / \|\boldsymbol{s}\|_e \rangle| \right| \leq A$ for some $A > 0$. Provided M is taken large enough (e.g., $M > 2A$), we will have in (42)

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

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

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

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

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

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 (39), 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 $i = 1, \ldots, d$. 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. □

C.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 & Taqqu (1994) to seminorms.

Proposition 7 *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 $\Gamma^{\|\cdot\|}(\partial A) = 0$,*

$$\lim_{x \to +\infty} x^{\alpha} \mathbb{P}\Big(\|\boldsymbol{X}\| > x, \frac{\boldsymbol{X}}{\|\boldsymbol{X}\|} \in A\Big) = C_\alpha \Gamma^{\|\cdot\|}(A), \tag{43}$$

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 & 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

$$Z_k = (C_\alpha \Gamma^{\|\cdot\|}(C_d^{\|\cdot\|}))^{1/\alpha} \sum_{i=1}^{\infty} [\Gamma_i^{-1/\alpha} S_i^{(k)} - b_{i,k}(\alpha)], \quad k = 1, \ldots, d, \tag{44}$$

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 (4). This allows to restate the integral representation Theorem 3.5.6 in Samorodnitsky & 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 & 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

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

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 & 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 (44). The rest of the proof is similar to that of Theorem 4.4.8 in Samorodnitsky & 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.[25] □

C.3 Proof of Theorem 1

From Proposition 1, we know that a necessary condition for the representability of $\boldsymbol{X_t}$ on $C_{m+h+1}^{\|\cdot\|}$ 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

[25] 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 $x \in \mathbb{R}^d$.

atoms on $C^{\|\cdot\|}_{m+h+1}$,

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

by (9). Now assume that the following holds:

$$\forall k \in \mathbb{Z}, \quad \Big[(d_{k+m}, \dots, d_k) = \mathbf{0} \Longrightarrow (d_{k+m}, \dots, d_{k-h}) = \mathbf{0}\Big]. \tag{45}$$

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

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

Therefore, (45) 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 (15) 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 $\int_{S_d} \Big| \ln \|\boldsymbol{s}\| \Big| \Gamma(d\boldsymbol{s}) < +\infty$. That is

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

Substituting Γ by its expression in (14), the above condition holds if and only if (15) 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

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

C.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,

$$\begin{aligned}
(X_t) \text{ is past-representable} &\iff \text{There exist } m \geq 0,\ h \geq 1, \text{ such that (15) holds} \\
&\iff \exists m \geq 0, \forall k \in \mathbb{Z}, \Big[d_{k+m} = \ldots = d_k = 0 \implies \forall \ell \leq k-1,\ d_\ell = 0 \Big].
\end{aligned}$$

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 (15) 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 (15) 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

$$d_{k+m_0} = \ldots = d_{k+1} = 0, \quad \text{and} \quad d_k \neq 0. \tag{46}$$

Because (X_t) is (m, h)-past-representable, we have by Theorem 1 that (15) 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,

$$(X_t) \text{ is past-representable} \iff \text{There exist } m \geq 0,\ h \geq 1, \text{ such that (15) and (16) hold}$$

From the previous proof, we moreover have that

$$\exists\, m \geq 0, \text{ such that (15) holds} \iff m_0 < +\infty \iff \begin{cases} m_0 < +\infty \\ \forall\, m' \geq m_0, \text{ (15) holds} \\ \forall\, m' < m_0, \text{ (15) does not hold} \end{cases}$$

Hence

$$\exists\, m \geq 0,\ h \geq 1, \text{ such that (15) and (16) hold}$$

$$\iff \begin{cases} m_0 < +\infty \\ \forall\, m' \geq m_0,\ (15) \text{ holds} \\ \forall\, m' < m_0,\ (15) \text{ does not hold} \\ \exists\, m \geq 0,\ h \geq 1, \text{ such that (15) and (16) hold.} \end{cases}$$

The latter in particular implies $m_0 < +\infty$ and the existence of $m \geq m_0$, $h \geq 1$ such that (16) holds. Reciprocally,

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

$$\implies \begin{cases} m_0 < +\infty \\ \forall\, m' \geq m_0,\ (15) \text{ holds} \\ \exists\, m \geq m_0,\ h \geq 1, \text{ such that (16) holds,} \end{cases}$$

which in particular implies that there exists $m \geq m_0$, $h \geq 1$ such that both (15) and (16) 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.

C.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 $m_0 = +\infty$.

C.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 (18) 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

$$\begin{aligned} & \sum_{i=-q}^{p} \varphi_i X_{t+i} = \sum_{k=-r}^{s} \theta_k \varepsilon_{t+k} \\ \iff \quad & \sum_{i=-q}^{p} \varphi_i \sum_{k\in\mathbb{Z}} d_k \varepsilon_{t+k+i} = \sum_{k=-r}^{s} \theta_k \varepsilon_{t+k} \\ \iff \quad & \sum_{k\in\mathbb{Z}} \Big(\sum_{i=-q}^{p} \varphi_i d_{k-i} \Big) \varepsilon_{t+k} = \sum_{k=-r}^{s} \theta_k \varepsilon_{t+k}. \end{aligned} \tag{47}$$

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

$$\sum_{i=-q}^{p} \varphi_i d_{k-i} = 0. \tag{48}$$

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 (48) 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

$$X_t = \sum_{k=-\infty}^{\bar{k}} d_k \varepsilon_{t+k}, \qquad t \in \mathbb{Z}. \tag{49}$$

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

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

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 (49), 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

$$\Theta(z^{-1})H(z)b_1(z) = \Psi(z^{-1})(P(z^{-1}) + Q(z)), \qquad z \in \mathbb{C}. \tag{50}$$

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 (50) 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 $\boldsymbol{d}_k \neq \boldsymbol{0}$, 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

$$\|\boldsymbol{d}_k\|_e \left| \ln\left(\|\boldsymbol{d}_k\|/\|\boldsymbol{d}_k\|_e\right)\right| \underset{k\to\pm\infty}{\sim} \text{const } k^b\lambda^k.$$

Therefore, for any $m \geq m_0$, $h \geq 1$,

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

The equivalence between (ι) and $(\iota\iota\iota)$ 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 (15) never holds with $m < m_0$ (*a fortiori*, with $m < m_0 = +\infty$), concluding the proof.

C.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 (22).

C.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,

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

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

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

C.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

$$I_0=\begin{cases}\left\{\dfrac{\vartheta_0\boldsymbol{d}_{k_0}}{\|\boldsymbol{d}_{k_0}\|}\right\}, & \textit{if}\quad -m+1\leq k_0\leq -1\\[2ex] \left\{\dfrac{\vartheta_0\boldsymbol{d}_{0,k_0}}{\|\boldsymbol{d}_{0,k_0}\|}\right\}=\{(\vartheta_0,0,\ldots,0)\}, & \textit{if}\quad k_0=-m.\end{cases}$$

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

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

and by (9) we also have that

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

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

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

$$f(\boldsymbol{d}_{k'}) = (\overbrace{\rho^{k'+m}, \ldots, \rho, 1, 0, \ldots, 0}^{m+1}, \overbrace{0, \ldots, 0}^{h}),$$

$$f(\boldsymbol{d}_{k_0}) = (\underbrace{\rho^{k_0+m}, \ldots, \rho, 1, 0, \ldots, 0}_{m+1}, \underbrace{0, \ldots, 0}_{h}),$$

and we also have that

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

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

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

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

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

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

If $k_0 = -m$, given that $(\vartheta_0, k_0) \in \mathcal{I} = S_1 \times \Big(\{-m, \ldots, -1, 0, 1, \ldots, h\} \cup \{(0, -m)\}\Big)$, and as $k' = k_0 = -m$, we have that $\boldsymbol{d}_{k_0} = \boldsymbol{d}_{0,-m} = (1, 0, \ldots, 0)$. 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 $k_0 \in \{1, \ldots, h\}$ then $f(\boldsymbol{d}_{k_0}) = \rho^{k_0}$ and by (9), $\|\boldsymbol{d}_{k_0}\| = |\rho|^{k_0}$. 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,

$$\mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t}, A_{\vartheta,k}\Big|B(V_0)\Big) \underset{x\to\infty}{\longrightarrow} \frac{\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta'\boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|}\in A_{\vartheta,k}:\ \ \frac{\vartheta' f(\boldsymbol{d}_{,k'})}{\|\boldsymbol{d}_{k'}\|}\in V_0\right\}\right)}{\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta'\boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|}\in C_{m+h+1}^{\|\cdot\|}:\ \ \frac{\vartheta' f(\boldsymbol{d}_{k'})}{\|\boldsymbol{d}_{k'}\|}\in V_0\right\}\right)}. \tag{51}$$

Focusing on the denominator, we have by (26)

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

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 $\bar{\beta}=\beta\dfrac{1-|\rho|^\alpha}{1-\rho^{<\alpha>}}=\beta$ and $\bar{w}_\vartheta = w_\vartheta$ in (25) for $\vartheta\in\{-1,+1\}$.

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

By Lemma 3,

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

By (9), for $k'\in\{0,1,\dots,h\}$

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

Thus,

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

Similarly for the numerator in (51), by (27),

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

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

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

The conclusion follows.

Case $m \ge 1$ and $-m \le k_0 \le -1$

We have by Lemma 3

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

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

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

and

$$\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta' \boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|} \in A_{\vartheta,k} : \ \frac{\vartheta' f(\boldsymbol{d}_{k'})}{\|\boldsymbol{d}_{k'}\|} \in V_0\right\}\right)$$
$$= \Gamma^{\|\cdot\|}\left(A_{\vartheta,k} \cap \left\{\frac{\vartheta_0 \boldsymbol{d}_{k_0}}{\|\boldsymbol{d}_{k_0}\|}\right\}\right)$$

$$= \begin{cases} \Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta_0 \boldsymbol{d}_{k_0}}{\|\boldsymbol{d}_{k_0}\|}\right\}\right), & \text{if } \vartheta = \vartheta_0, \text{ and } k = k_0, \\ \Gamma^{\|\cdot\|}(\emptyset), & \text{if } \vartheta \neq \vartheta_0 \text{ or } k \neq k_0, \end{cases}$$

$$= w_{\vartheta_0} \|\boldsymbol{d}_{k_0}\|^{\alpha} \delta_{\{\vartheta_0\}}(\vartheta) \delta_{\{k_0\}}(k).$$

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

$$\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta' \boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|} \in A_{\vartheta,k} : \ \frac{\vartheta' f(\boldsymbol{d}_{k'})}{\|\boldsymbol{d}_{k'}\|} \in V_0\right\}\right)$$
$$= \Gamma^{\|\cdot\|}\left(A_{\vartheta,k} \cap \left\{\frac{\vartheta_0 \boldsymbol{d}_{k_0}}{\|\boldsymbol{d}_{k_0}\|}\right\}\right)$$

$$= \begin{cases} \Gamma^{\|\cdot\|}\Big(A_{\vartheta,k} \cap \{\vartheta_0(1, 0, \ldots, 0)\}\Big), & \text{if } \vartheta = \vartheta_0, \text{ and } k = k_0 = -m, \\ \Gamma^{\|\cdot\|}(\emptyset), & \text{if } \vartheta \neq \vartheta_0 \text{ or } k \neq k_0, \end{cases}$$

$$= w_{\vartheta_0} \delta_{\{\vartheta_0\}}(\vartheta) \delta_{\{k_0\}}(k).$$

The conclusion follows as previously.

Case $m = 0$

By Lemma 3, as the ρ is positive

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

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

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

Similarly, by (27),

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

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

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

The conclusion follows.

C.10 Proof of Proposition 6

Lemma 4 *Let X_t be the α-stable anticipative AR(2) (resp. fractionally integrated AR) as in (28) (resp. (30)). With f as in (21), and for any $m \geq 1$, $h \geq 0$,*

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

Proof.

The result is clear for both processes for $-m \leq k, \ell \leq -1$. For $k, \ell \geq 0$,

$$\begin{aligned} \frac{f(\vartheta_1 \boldsymbol{d}_k)}{\|\boldsymbol{d}_k\|} = \frac{f(\vartheta_2 \boldsymbol{d}_\ell)}{\|\boldsymbol{d}_\ell\|} &\iff \left[\forall\, i = 0, \dots, m, \quad \frac{\vartheta_1 d_{k+i}}{\|\boldsymbol{d}_k\|} = \frac{\vartheta_2 d_{\ell+i}}{\|\boldsymbol{d}_\ell\|} \right] \\ &\iff \frac{d_k}{d_\ell} = \frac{d_{k+1}}{d_{\ell+1}} = \dots = \vartheta_1 \vartheta_2 \frac{\|\boldsymbol{d}_k\|}{\|\boldsymbol{d}_\ell\|}. \end{aligned} \tag{52}$$

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

$$\begin{aligned} \frac{d_k}{d_\ell} = \frac{d_{k+1}}{d_{\ell+1}} &\iff \frac{\lambda_1^{k+1} - \lambda_2^{k+1}}{\lambda_1^{\ell+1} - \lambda_2^{\ell+1}} = \frac{\lambda_1^{k+2} - \lambda_2^{k+2}}{\lambda_1^{\ell+2} - \lambda_2^{\ell+2}} \\ &\iff \lambda_1^{k-\ell} = \lambda_2^{k-\ell} \\ &\iff k = \ell. \end{aligned}$$

This 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,

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

□

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 (29) and (31) 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.