Forecasting extreme trajectories using seminorm representations

Gilles de Truchis[*], Sébastien Fries[†], Arthur Thomas[‡]

November 2, 2026

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 $\boldsymbol{X}_t = (X_{t-m}, \ldots, X_{t+h})$ are representable on $\{\boldsymbol{s} \in \mathbb{R}^{m+h+1} : \|\boldsymbol{s}\| = 1\}$ for $\|\cdot\|$ an appropriate seminorm. The key finding is that purely causal processes are inherently non-representable and that anticipativeness is required. For such noncausal 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, so that the conditional law of the future path degenerates to a single point in the large-deviation limit. 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: Anticipative processes, Extreme value theory, Pattern identification, Stable random vectors

[*]Université d'Orléans, Laboratoire d'Economie d'Orléans (LEO), 45000 PARIS, FRANCE, gilles.detruchis@univ-orleans.fr.

[†]Department of Econometrics and Data Science, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[‡]Université Paris Dauphine, Université PSL, LEDA, CNRS, IRD, 75016 PARIS, FRANCE, 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 (Fries and Zakoian, 2019; Gouriéroux and Zakoian, 2017; Hecq et al., 2017a; Cavaliere et al., 2020; Gouriéroux and Jasiak, 2017; Bec et al., 2020; Fries, 2022; Hecq and Voisin, 2021; Blasques et al., 2025; Hecq and 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 and Saikkonen, 2013; Chen et al., 2017; Gouriéroux et al., 2019; Chahrour and 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 and Saikkonen, 2011, 2013; Hecq et al., 2016; Fries and Zakoian, 2019; Hecq et al., 2017b; Cavaliere et al., 2020; Velasco and Lobato, 2018; Hecq et al., 2020; Gouriéroux and Jasiak, 2016, 2017; Gourieroux and Jasiak, 2023; Hecq and Velasquez-Gaviria, 2025; Jasiak and Manafi Neyazi, 2026), 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; Gouriéroux and Jasiak, 2016; Gourieroux et al., 2021; Dumitrescu et al., 2026). Recently, Gourieroux and Jasiak (2026) 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 guarantees and are computationally expensive for long horizons and they fail to capture dynamics during explosive episodes accurately (Hecq and Voisin, 2021).

An exception is the anticipative α-stable AR(1) process, for which partial results were obtained in Gouriéroux and 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

[1] Fries and 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 and Jasiak (2026) extends Gouriéroux and Jasiak (2016)'s forecasting approach to multivariate mixed-causal (S)VAR, but their kernel-based conditioning methods remain computationally demanding.

anticipative stable AR(1) process simplify during extreme events, appearing to follow an explosive exponential path with a determined killing probability (see Gouriéroux and Zakoian, 2017; 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, $(X_{t+1},\dots,X_{t+h})$ given the observed trajectory $(X_{t-m},\dots,X_t)$ 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 $\boldsymbol{X_t} = (X_{t-m},\dots,X_{t+h})$ 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 and 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 and 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 $X_{t-m},\dots,X_t$ are assumed to be observed. This is because, through the Euclidean norm of $\boldsymbol{X}_t$, the conditioning also depends on the future realizations $X_{t+1},\dots,X_{t+h}$. 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

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

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

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

[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.

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, we explore a new representation of stable random vectors on the "unit cylinder" $C_{m+h+1}^{\|\cdot\|} := \{\boldsymbol{s} \in \mathbb{R}^{m+h+1} : \ \|\boldsymbol{s}\| = 1\}$, where $\|\cdot\|$ is such a seminorm. Contrary to representations involving norms (see Theorem 2.3.8 in Samorodnitsky and Taqqu (1994)), not every stable random vector can be characterized on such a unit cylinder: we show 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 Borel sets 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 and Hsing, 1995; Konstantinides and Mikosch, 2005; Meinguet and Segers, 2010; Buraczewski et al., 2013; Janßen and Segers, 2014; Basrak et al., 2016; Dombry et al., 2017; Janßen, 2019; Planinić and 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 and 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 and 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 and 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. Gouriéroux et al. (2025) bring this regularly varying, spectral tail process machinery to bear on mixed causal-noncausal autoregressive (MAR) processes, deriving CEV-type limits for their extreme trajectories. Our α-stable, seminorm-based approach differs from this line of work, and more broadly from the two standard CEV routes to conditional extremes, multivariate regular variation of the path vector, and the spectral tail process anchored at the single coordinate $\{|X_t| > x\}$, in three connected respects. First, whereas the general regu-

[4] Another strand of this vast literature relates to the extremogram introduced by Davis and 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.

larly varying framework only guarantees the existence of the limit tail measure and, for practical use, requires estimating its tail index and extremal skewness, the α-stable assumption yields an exact and finite expression of the spectral measure $\Gamma^{\|\cdot\|}$, a discrete weighted sum of Dirac masses located at the normalized MA(∞) coefficient vectors (Section 4). Second, the spectral tail process and Gouriéroux et al. (2025) anchor extremal conditioning on a single coordinate, the large-magnitude event $\{|X_t| > x\}$, and update forecasts online as new observations arrive. Our seminorm framework instead conditions directly on the observed shape of the entire past window $(X_{t-m}, \ldots, X_t)$, which is the information actually available for prediction. Third, this shape conditioning yields a sharper limit: for sufficiently anticipative processes (anticipative AR($p \geq 2$) or fractionally integrated), the conditional distribution of the future path degenerates to a single Dirac mass (Proposition 6), an asymptotically deterministic forecast of the path shape under the assumed model. Section 4.4, together with Appendix A of the Supplementary Material, positions the seminorm representation relative to these two routes. Single-coordinate anchoring on $\{|X_t| > x\}$ amounts to a degenerate seminorm that retains only the present coordinate X_t of $\boldsymbol{X_t}$ and discards the rest of the past window. For purely anticipative processes observed over a window of length at least two, the path vector $\boldsymbol{X_t}$ is representable on every past-window seminorm cylinder we use, but not on this coordinate cylinder, and the missing spectral mass is computed in closed form (Proposition 7 in the Supplementary Material). It is carried by an explicit family of post-peak patterns, in which an extreme has already occurred in the recent past while the present coordinate has reverted toward central values, and it is exactly the asymptotic fraction of the extremes of $\boldsymbol{X_t}$ that fail to trigger $\{|X_t| > \epsilon x\}$ however small $\epsilon > 0$. Conditioning on the observed past window captures these patterns. This past-window conditioning can also be reached within the tail-measure calculus of Kulik and Soulier (2020): Proposition 8 in the Supplementary Material re-derives our conditional limit from the tail measure of (X_t), for any norm on the past block and any two-sided stable moving average with $\alpha \neq 1$ or $(\alpha, \beta) = (1, 0)$, and shows that the two routes coincide. The seminorm representation is thus not a competitor to the tail-process framework but an exact, stable-specific route to the same limits, which delivers them in closed form and makes explicit the link between anticipativeness, past-representability and the predictability of extremes.

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, the conditional distribution of the future path degenerates asymptotically to a single pattern, so that, under the assumed model and provided the current pattern is correctly identified, the peak and reversal dates are pinned down in the limit. The same section closes with a dedicated subsection (Section 4.4) that positions our approach relative to the multivariate regular variation and spectral tail process frameworks, establishing that it coincides with them on the past-window conditioning event and quantifying what the single-coordinate anchoring misses. 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] Section 6 concludes the paper. Proofs, the formal comparison with the regular variation and spectral tail process frameworks, additional Monte Carlo simulations and complementary empirical results are provided in the Supplementary Material (de Truchis et al., 2026). All results can be reproduced with a replication package in Python and R and with a web application (`https://marforecast.streamlit.app/`), which also applies our methods to other time series.

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

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

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,

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

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

where $\langle \cdot, \cdot \rangle$ denotes the canonical scalar product, $w(\alpha, s) = \text{tg}\left(\frac{\pi\alpha}{2}\right)$, 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 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 and Taqqu (1994)), and in that case

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

The representations (5) and (6) of a stable random vector involves integration over all directions of $\mathbb{R}^d$,[6] parameterized here by the unit sphere with respect to the Euclidean norm. Proposition 2.3.8 in Samorodnitsky and 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 $\alpha = 1$,

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

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

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

In the case where $\boldsymbol{X}$ is symmetric α-stable ($S\alpha S$), $0 < \alpha < 2$, 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 $(C_d^{\|\cdot\|}, \Gamma^{\|\cdot\|})$.

[6] 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 $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^d$.

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.[7] 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 and 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 SαS). If $\alpha \neq 1$ 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 $(\Gamma^{\|\cdot\|}, \boldsymbol{\mu}^0_{\|\cdot\|})$ where

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

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

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

[7] 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 $\lambda \in \mathbb{R}$.

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

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 $\Gamma(\{(0, -1), (0, +1)\}) = 0$ 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 and Taqqu (1994, 1998) - see also Theorem 5.1.3 in Samorodnitsky and 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 $(X_{t-m},\dots,X_t,X_{t+1},\dots,X_{t+h})$ 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

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

for any $(x_{-m},\dots,x_h)\in\mathbb{R}^{m+h+1}$, 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

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

for any $(x_{-m},\dots,x_0,x_1,\dots,x_h)\in\mathbb{R}^{m+h+1}$ 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

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

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

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

it follows from Proposition 13.3.1 in Brockwell and 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

$$\begin{aligned} \Gamma &= \sigma^\alpha \sum_{\vartheta \in S_1} \sum_{k \in \mathbb{Z}} w_\vartheta \|\boldsymbol{d}_k\|_e^\alpha \delta_{\left\{ \frac{\vartheta \boldsymbol{d}_k}{\|\boldsymbol{d}_k\|_e} \right\}}, \\ \boldsymbol{\mu}^0 &= -\mathbb{1}_{\{\alpha=1\}} \frac{2}{\pi} \beta \sigma \sum_{k \in \mathbb{Z}} \boldsymbol{d}_k \ln \|\boldsymbol{d}_k\|_e, \end{aligned} \tag{15}$$

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 and 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.[8] 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

[8] 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.

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.[9] 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*

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

(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 $\|\cdot\|$.

[9] 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.

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 C.5 and C.6 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 and 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 $S_{m+1}^{\|\cdot\|} = \{(s_{-m}, \ldots, s_0) \in \mathbb{R}^{m+1} : \|(s_{-m}, \ldots, s_0, 0, \ldots, 0)\| = 1\}$.[10] Then, for any Borel set $V \subset S_{m+1}^{\|\cdot\|}$, define the Borel set

[10] 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.

$B(V) \subset C_{m+h+1}^{\|\cdot\|}$ as

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

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

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

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

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

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 for prediction. We deal with two of them in the rest of the paper and leave the third for future research.*[11]

(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, 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*

[11] 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.

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 the Supplementary Material). 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*

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

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

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

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 $\mathcal{I} := S_1 \times \Big(\{-m+1, \dots, h\} \cup \{(0, -m)\}\Big)$. 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

$$\forall (\vartheta', k') \in S_1 \times \mathbb{Z} \text{ with } \boldsymbol{d}_{k'} \neq \boldsymbol{0}, \qquad \frac{\vartheta' f(\boldsymbol{d}_{k'})}{\|\boldsymbol{d}_{k'}\|} \in V_0 \implies \frac{\vartheta' f(\boldsymbol{d}_{k'})}{\|\boldsymbol{d}_{k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{k_0})}{\|\boldsymbol{d}_{k_0}\|}, \tag{27}$$

In other terms, $V_0 \times \mathbb{R}^d$ is a subset of $C^{\|\cdot\|}_{m+h+1}$ 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:

$$\forall(\vartheta',k') \in S_1 \times \mathbb{Z} \text{ with } \boldsymbol{d}_{k'} \neq \boldsymbol{0}, \qquad \frac{\vartheta'\boldsymbol{d}_{k'}}{\|\boldsymbol{d}_{k'}\|} \in A_{\vartheta,k} \implies (\vartheta',k') = (\vartheta,k). \tag{28}$$

Conditions (27) and (28) are stated for all indices $(\vartheta',k') \in S_1 \times \mathbb{Z}$, and not only for the finitely many indices $\mathcal{I}$ charged by the spectral measure of the anticipative AR(1). They therefore apply without change to any past-representable moving average whose spectral measure charges infinitely many points, such as the anticipative AR(2) and FWN of Proposition 6, and to the general moving averages of Appendix A of the Supplementary Material. Such isolating neighborhoods exist for all the targets considered in Propositions 5 and 6: for the AR(1), the charged past shapes take finitely many distinct values; for the FWN and for the anticipative AR(2) with real characteristic roots, the charged past shapes are pairwise distinct and can only accumulate at the direction associated with the dominant root, which differs from every target shape with finite index k_0 (see the proof of Proposition 6 in Appendix E of the Supplementary Material). For complex conjugate roots $\lambda_1 = \bar{\lambda}_2 = re^{i\phi}$ in (29), by contrast, the normalized past shapes with $k \geq 0$ depend on k only through the phase $(k+1)\phi \bmod 2\pi$: when $\phi \in \pi\mathbb{Q}$ they recur exactly along an arithmetic progression of indices, so that neighborhoods satisfying (27) exist but select a whole recurrence class rather than a single index; when $\phi \notin \pi\mathbb{Q}$ they are dense on a closed curve of $\mathbb{R}^{m+1}$, precluding the existence of such neighborhoods at targets $k_0 \geq -1$. In both regimes the 0–1 conclusion of Proposition 6 fails at targets $k_0 \geq -1$, which is why the proposition is stated for real characteristic roots.

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 $m \geq 0$, $h \geq 1$. 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 $0 \leq k_0 \leq h$:*

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

(*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 $k_0 \in \{0,\dots,h\}$ or $k_0 \in \{-m,\dots,-1\}$. 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),

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

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

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

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 ($0 \leq k < h$), 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 $d \notin \mathbb{Z}$ and $\min(\alpha,1)(1-d) > 1$, i.e. $d < 1 - 1/\alpha$ if $\alpha \leq 1$ and $d < 0$ if $\alpha > 1$, so that (12) holds. 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 processes 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) with real characteristic roots, or the 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 $\vartheta_0 f(\boldsymbol{d}_{k_0})/\|\boldsymbol{d}_{k_0}\|$ 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 conditional law of the future-path shape degenerates to a single atom. This holds even if the peak/collapse of a bubble is not yet present in the observed piece of trajectory. Therefore, within these models, as the magnitude of the observed extreme grows, the conditional limit concentrates on a unique peak/crash date. We stress that this is an asymptotic, within-model statement: at finite magnitude it is an approximation, and it presupposes correct pattern identification (see point (ι) of Remark 2) and estimated parameters. Finally, the case of complex conjugate characteristic roots $\lambda_1 = \bar{\lambda}_2 = re^{i\phi}$ is not treated explicitly here. As discussed after (28), the normalized past shapes then recur (along an arithmetic progression of indices when $\phi \in \pi\mathbb{Q}$, densely on a closed curve otherwise) and the 0-1 property fails at every target $k_0 \geq -1$. This is a limitation of the Dirac specialization only, not of the underlying framework: when $\phi \in \pi\mathbb{Q}$, neighborhoods V_0 satisfying (27) still exist and Proposition 4 applies, the conditional limit distributing its mass over the recurrence class of indices sharing the past shape of the target, with geometrically decaying weights, in analogy with case $(\iota)(a)$ of Proposition 5: an oscillating trajectory whose observed past matches a pre-peak pattern may either reverse durably or rebound one pseudo-period later, with relative probabilities delivered in closed form by (24). The exact identification nonetheless persists at post-crash targets $k_0 \in \{-m, \ldots, -2\}$, whose past shapes end with at least two consecutive zero coordinates, a configuration the recurrent shapes never attain. A complete treatment of the complex-rooted case is left for future research.

Remark 4 A fitted MAR(p, q) process is Markov of order $p + q$ (Fries and Zakoian, 2019). For the anticipative AR(2) ($p = 0$, $q = 2$), this implies that the pair (X_{t-1}, X_t), i.e. $m = 1$, is already a sufficient state vector to identify the current trajectory piece, coinciding with the theoretical minimum $m \geq 1$ required by Proposition 6.

4.4 Why a seminorm representation?

Two more standard routes to conditional extremes of time series are available: the multivariate regular variation of the path vector $\boldsymbol{X_t}$, and the spectral tail process of Basrak and Segers (2009), anchored at the single coordinate $\{|X_t| > x\}$ (see Kulik and Soulier, 2020; Mikosch and Wintenberger, 2024). Appendix A of the Supplementary Material positions the seminorm representation relative to both routes, through two propositions whose content is summarized below.

First, single-coordinate anchoring amounts to the degenerate seminorm $\boldsymbol{s} \mapsto |s_0|$, which does

not satisfy (10) when $m \geq 1$. For a purely anticipative stable moving average with nonvanishing coefficients, $\boldsymbol{X_t}$ is representable on every unit cylinder associated with a seminorm satisfying (10), but not on the coordinate cylinder (Proposition 7 in the Supplementary Material). The missing spectral mass is computed in closed form. It is carried by post-peak patterns, in which an extreme has already occurred in the observed window while the present coordinate X_t has reverted toward central values, and it equals the asymptotic fraction of the extremes of $\boldsymbol{X_t}$ that fail to trigger $\{|X_t| > \epsilon x\}$ however small $\epsilon > 0$. The same construction, applied to any seminorm satisfying (10), yields the asymptotic fraction π_m of the extremes of $\boldsymbol{X_t}$ that are not announced by the observed window of length $m+1$: π_m vanishes if and only if $\boldsymbol{X_t}$ is representable on $C^{\|\cdot\|}_{m+h+1}$ (for $\alpha \neq 1$ or $(\alpha, \beta) = (1, 0)$), i.e. if and only if $m \geq m_0$, and it is positive for every m for purely non-anticipative processes, which gives a quantitative content to the link between anticipativeness and the predictability of extremes (Corollary 1).

Second, the conditional limit of Proposition 4 is also accessible within the tail-measure calculus of Kulik and Soulier (2020). For every norm N on the past block and every two-sided α-stable moving average with $\alpha \neq 1$ or $(\alpha, \beta) = (1, 0)$, the limit given $\{N(X_{t-m}, \ldots, X_t) > x\}$ and the observed past shape can be derived from the tail measure of (X_t), through the identity (5.5.7) of Kulik and Soulier (2020) applied to their first-exceedance anchor (Example 5.5.2 there), an anchoring map in the sense of Basrak and Planinić (2021), combined with the single-big-jump representation of the tail process, without any representability condition; it coincides with (24) whenever (X_t) is past-representable (Proposition 8 in the Supplementary Material). The difference with the spectral tail process is thus one of conditioning events rather than of frameworks, and the Dirac collapse of Proposition 6 is a property of the coefficient sequence, which can also be recovered from the spectral tail process for patterns whose peak is still ahead. What the seminorm representation contributes is an exact, stable-specific route to these limits: it delivers them in a single closed-form statement, uniformly across past-representable specifications and seminorms, under a normalization computable from the observations whatever the position of the peak; and its representation theorem, which holds at every finite scale, identifies past-representability with the explicit condition (16) on the coefficients. The tail-measure route, which uses stability only through the single-big-jump representation, is the natural one for extending these conditional limits beyond the stable class, which we leave for future research.

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).[12] This section discusses the performance of our forecasting procedure 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). The specification retained for the in-sample SOI data is a purely anticipative AR(2), estimated using two methods from the literature: the Generalized Covariance (GCoV) estimator of Gourieroux and Jasiak (2023) and the approximate maximum likelihood (AML) method of Lanne and Saikkonen (2011), both adapted for α-stable laws, which agree closely on the estimated coefficients. We refer the reader to Appendix D of the Supplementary Material for the specification search, the estimation details, and the associated robustness checks.[13]

5.1 A formal forecasting algorithm for the anticipative AR(2)

This subsection formalizes the pattern-matching forecasting device used to predict reversal dates. We present it here for the anticipative AR(2), as this is the specification retained for the SOI. The same algorithm applies verbatim to the anticipative fractionally integrated white noise because Proposition 6 covers both processes; the only difference is the closed-form expression of the coefficient sequence (d_k). The AR(1) case, which requires a probabilistic reformulation because Proposition 5 does not yield a $\{0,1\}$-valued limit, is treated in Appendix B.1 of the Supplementary Material. The general MAR(r,s) case, which relies on the closed-form MA(∞) coefficients derived in de Truchis and Thomas (2026), is treated in Appendix B.2 of the Sup-

[12] 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.

[13] A Hill estimator on the in-sample SOI, together with external evidence (Ausloos and Ivanova, 2001) on SOI increments, suggests heavy tails relative to the Gaussian but plausibly finite variance, rather than a literally α-stable law (see Appendix D.1 of the Supplementary Material). We nonetheless retain the α-stable assumption because, unlike competing heavy-tailed forecasting approaches (Gourieroux and Jasiak, 2026; Gouriéroux et al., 2025), it yields an exact closed-form forecasting formula, which is the object of this paper.

plementary Material.

Inputs and rationale

Fix an observed trajectory $(X_1, \dots, X_t)$ and assume the estimated anticipative AR(2) roots $(\hat{\lambda}_1, \hat{\lambda}_2)$ and the estimated tail index $\hat{\alpha}$ are available. Pick a lag-window length $m \geq 1$, a forecast horizon $h \geq 1$ and an upper search bound $\bar{k} \geq 0$.[14] From the observed extreme sign $\vartheta_0 \in \{-1, +1\}$ (positive vs. negative extreme event) and the closed-form AR(2) coefficients in (30), we compute for every integer $k \in \{-m, \dots, \bar{k}\}$ the candidate reference vector

$$\boldsymbol{d}_k = (d_{k+m},\, d_{k+m-1},\, \dots,\, d_k), \tag{33}$$

where d_j is given by (30). The quantity $\boldsymbol{d}_k$ is the "past-only" piece of the deterministic pattern that the normalized observed trajectory must asymptotically match, by Proposition 6. The algorithm below identifies the unique index k_0 for which this match holds, then extrapolates the deterministic pattern forward to produce the point forecasts $(\hat{X}_{t+1}, \dots, \hat{X}_{t+h})$.

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

Require: observed path $(X_{t-m}, \dots, X_t)$, sign $\vartheta_0 \in \{-1, +1\}$, window m, horizon h, upper bound $\bar{k}$, seminorm $\|\cdot\|$[15]

compute the normalized observed past $\boldsymbol{v}_t := (X_{t-m}, \dots, X_t)/\|(X_{t-m}, \dots, X_t)\|$
for $k = -m, -m+1, \dots, \bar{k}$ **do**
    compute $\boldsymbol{d}_k$ as in (33) using the AR(2) formula (30)
    store the candidate reference $\boldsymbol{u}_k := \vartheta_0\, \boldsymbol{d}_k/\|\boldsymbol{d}_k\|$
    store the distance $e(k) := \|\boldsymbol{v}_t - \boldsymbol{u}_k\|_2$
end for
define $\hat{k}_0 := \arg\min_k e(k)$ (tie-breaking rule: pick the smallest $|k|$; see Section 5.4)
build the extended reference $\boldsymbol{d}^{\text{ext}}_{\hat{k}_0} := (d_{\hat{k}_0+m}, \dots, d_{\hat{k}_0}, \dots, d_{\hat{k}_0-h})$
set the rescaling factor $\hat{c} := \|(X_{t-m}, \dots, X_t)\|/\|(d_{\hat{k}_0+m}, \dots, d_{\hat{k}_0})\|$
for $j = 1, \dots, h$ **do**
    $\hat{X}_{t+j} := \vartheta_0\, \hat{c}\; d_{\hat{k}_0-j}$
end for
the forecasted reversal (peak) date is $\hat{t}_{\text{peak}} := t+\hat{k}_0$; the forecasted end date is $\hat{t}_{\text{end}} := t+\hat{k}_0+1$

Intuitively, the first block of Algorithm 1 identifies the position $\hat{k}_0$ of the observed piece of

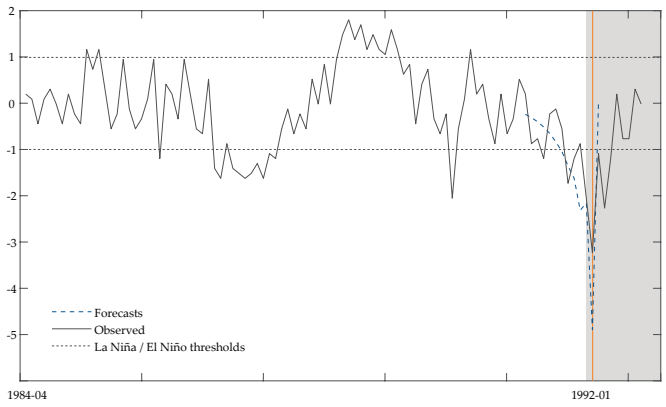
[14] In our implementation we take $\bar{k} = 30$, which largely exceeds m and h. Since the normalized candidate patterns $\boldsymbol{d}_k/\|\boldsymbol{d}_k\|$ stabilize as k grows, converging to the direction associated with the dominant root, the argmin of Algorithm 1 is unaffected by the choice of $\bar{k}$ as soon as it is taken large enough.

trajectory within the deterministic pattern (d_k); the second block uses the identity $\boldsymbol{X_t}/\|\boldsymbol{X_t}\| \to \vartheta_0 \boldsymbol{d}_{k_0}/\|\boldsymbol{d}_{k_0}\|$ from Proposition 6 to extrapolate the remaining h coordinates deterministically.

5.2 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 apply Algorithm 1 to predict the reversal date of the El Niño occurrence that presumably starts at the end of the in-sample period. As the last in-sample observation is below -1, the sign $\vartheta_0 = -1$ and the magnitude $\|\boldsymbol{X_t}\|$ is large. We then run Algorithm 1 for $m \in \{1, \ldots, 10\}$. Our findings offer some robustness in this light-tailed case: for $m \in \{1, 2\}$ and $m \in \{5, \ldots, 10\}$ the algorithm returns $\hat{k}_0 = 1$; for $m = 3$ and $m = 4$ we obtain $\hat{k}_0 = 5$ and $\hat{k}_0 = 3$ respectively. Following the rule detailed in Section 5.3 below, we retain $\hat{k}_0 = 1$ and $m = 10$, therefore implying an imminent reversal date as we are close to the last piece of the trajectory described by $\vartheta_0 \boldsymbol{d}_{\hat{k}_0}$. The selected piece is represented in Figure 1. 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 realization, the reversal date appears accurately predicted; however, the magnitude of the peak is overestimated as $x_{t+1} = -3.04$.

Figure 1: El Niño reversal forecast



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

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 nand 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 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 11 El Niño and La Niña occurrences in the out-of-sample dataset, our procedure leads to a mean absolute error of 0.55 months in finding the peak date and 0.73 months in finding the end date compared to the true peak and end dates. We also report in Table 1 the selected $\hat{k}_0$ and m from Algorithm 1.

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	01/2022	03/2022	10/2022	12/2022
End date	04/1992	10/1994	03/2008	03/2010	11/2010	04/2011	11/2015	03/2022	05/2022	11/2022	02/2023
Forecasted Peak	01/1992	09/1994	02/2008	03/2010	08/2010	01/2011	09/2015	01/2022	04/2022	10/2022	01/2023
Forecasted End	02/1992	10/1994	03/2008	04/2010	09/2010	02/2011	10/2015	02/2022	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
$\hat{k}_0$	1	2	3	3	1	2	2	2	2	2	2
m	10	10	10	9	10	10	10	10	10	10	10

5.3 Selection of m: theoretical pinning and empirical validation

By Remark 4, the Markov property of the anticipative AR(2) pins the theoretical minimum window length at $m = 1$. In finite samples, however, this lower bound turns out to be a poor guide to practice. Monte Carlo evidence for the MAR(0,2) design (Appendix C.1 of the Supplementary Material, 1000 replications per cell) shows that the bias of the crash-date predictor instead increases with m for every tail index α we consider, except in the regime closest to the SOI's estimated tail index ($\hat{\alpha} \approx 1.9$), where it remains close to flat for $m \geq 3$. Concretely, in the Moderately High regime (the most relevant for the SOI in-sample series), the bias for $\alpha = 1.8$ ranges from 8.355 ($m = 3$) to 8.470 ($m = 11$), a span of about 0.12 relative to a level of 8.4, essentially flat, whereas for $\alpha = 0.9$ it rises from 6.850 ($m = 1$) to 7.542 ($m = 11$), a span of about 0.7.

As a robustness check of our empirical approach, we therefore structure the forecast analysis around four complementary benchmarks. (ι) Theoretical benchmark ($m = 1$): the minimal value. ($\iota\iota$) Medium benchmark ($m = 5$): a middle-ground choice. ($\iota\iota\iota$) Baseline ($m = 10$): the value retained in the main analysis of Table 1. ($\iota\iota\iota\iota$) Data-driven benchmark ($m = \hat{m}$): the value in $\{2, 3, \ldots, 10\}$ minimising the pattern-match distance $e_m(\hat{k}_0)$.

Table 19 in Appendix D.5 of the Supplementary Material reports, for each of the 11 out-of-sample episodes, the reversal-date forecast under the four benchmarks above. The baseline $m = 10$ achieves the lowest mean absolute error (0.64 months), against 1.55 for the theoretical minimum $m = 1$, 0.91 for the medium benchmark $m = 5$, and 0.91 for the data-driven $\hat{m}$

(the AML-based counterpart in Table 20 gives near-identical results: 1.36, 0.91, 0.64 and 0.73, respectively). This supports the choice of $m = 10$ made in the main analysis, but Appendix C.2 of the Supplementary Material shows that, at this window length, the tie-breaking rule of the next subsection is invoked in the vast majority of cases for a tail index close to the SOI's; the good performance of $m = 10$ should therefore be read as a property of the pattern-matching step combined with the tie-breaking rule, not of an unambiguous identification of k_0.

5.4 Selection of k_0 and handling of ties

Algorithm 1 returns $\hat{k}_0 = \arg\min_k \|\boldsymbol{v}_t - \vartheta_0 \boldsymbol{d}_k / \|\boldsymbol{d}_k\|\|_2$, i.e. the nearest candidate to the observed normalized trajectory in Euclidean distance. Equivalently, this amounts to choosing the smallest closed neighborhood A of $\boldsymbol{v}_t$ in the sense of Proposition 6 such that exactly one point $\vartheta_0 \boldsymbol{d}_k / \|\boldsymbol{d}_k\|$ belongs to A. We stress that this tie-breaking mechanism has no counterpart in the asymptotic statement of Proposition 6: the Dirac collapse therein is established for a neighborhood A that already isolates a unique candidate $\vartheta_0 \boldsymbol{d}_{k_0} / \|\boldsymbol{d}_{k_0}\|$; when no such isolating neighborhood is available at finite x, the proposition simply does not apply, and the choice made by Algorithm 1 is a practical, not a theoretical, device.

Ties or near-ties occur when two candidates produce almost indistinguishable normalized shapes. Formally, let $e(k) = \|\boldsymbol{v}_t - \vartheta_0 \boldsymbol{d}_k / \|\boldsymbol{d}_k\|\|_2$, and declare a tie whenever the relative gap $(e_{(2)} - e_{(1)})/e_{(1)}$ between the first and second order statistics of $\{e(k)\}_k$ falls below $\eta = 0.05$. We assessed this mechanism by simulation (Appendix C.2 of the Supplementary Material, same MAR(0,2) design and (α, m, N) grid as above, 1000 replications per cell). Two findings depart from what a low-dimensional-collinearity intuition would suggest. First, at the theoretical minimum $m = 1$, the tie frequency is already non-negligible and grows with α: from about 4–16% across regimes at $\alpha \leq 1.5$, up to 35% in the Moderately High regime at $\alpha = 1.8$. Second, and more importantly, the tie frequency increases sharply with m rather than decreasing: for $\alpha = 1.8$ in the Moderately High regime it rises from 0.349 at $m = 1$ to 0.807 at $m = 5$ and 0.954 at $m = 11$. This reflects a geometric feature of the candidate vectors $\boldsymbol{d}_k = (d_{k+m}, \ldots, d_{k-h})$: as m grows, two consecutive candidates $\boldsymbol{d}_k$ and $\boldsymbol{d}_{k+1}$ overlap in m of their $m + 1$ coordinates, so their normalized shapes become increasingly collinear regardless of sample size. Enlarging m therefore does not make the pattern-matching step more discriminating; it makes near-ties more frequent. Since the baseline $m = 10$ retained for the SOI falls between the $m = 9$ and $m = 11$ columns of the simulation grid, where the tie frequency is already 0.922 and 0.954 respectively for $\alpha = 1.8$, the tie-breaking rule below is, at the window length actually used in the empirical application, the mechanism that resolves the pattern-matching step in the large

majority of replications, rather than a safeguard for a marginal edge case. We do not attempt a formal statistical treatment of this step, e.g. deriving the sampling distribution of $\hat{k}_0$ or an optimal tie-breaking rule from first principles is left to future work, and could likely be refined. Our purpose here is more modest: to illustrate, on the SOI application, that the theoretical probability results derived above are applicable in practice, not to deliver a statistically optimal pattern-matching procedure.

When a tie is declared, we apply step 7 of Algorithm 1: among the tied candidates, we keep the one with smallest $|k|$, because the smallest $|k|$ corresponds to the extreme piece of the trajectory closest to the current date, hence the most informative under the seminorm conditioning. As with the tie-declaration threshold above, we do not claim this choice is statistically optimal, comparing it, for $\alpha = 1.8$, against the alternative rule that retains the second-best candidate instead (Appendix C.3 of the Supplementary Material), both rules carry a substantial and comparable-magnitude bias in absolute terms, and "smallest $|k|$" is not uniformly closer to zero bias than the alternative. This is consistent with our overall stance: a proper statistical treatment of the tie-breaking rule is left for future work, and here we retain "smallest $|k|$" as a reasonable, qualitatively motivated default rather than a data-driven or empirically validated choice.

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 an asymptotically deterministic prediction of the future-path shape under the assumed model. 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.

Acknowledgement

We thank Guillaume Chevillon, Elena Dumistrescu, Christian Francq, Christian Gourieroux, Alain Hecq, Joann Jasiak, Sebastien Laurent, Aryan Manafi Neyazi, 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. We are grateful to the Agence Nationale de la Recherche (ANR), which supported this work via Project MultiRisk (ANR CE26 2016 - CR)

Corresponding author and use of AI tools

The corresponding author is Arthur Thomas (arthur.thomas@dauphine.psl.eu). The authors used an AI assistant (Claude, Anthropic) for language editing and consistency checks of the manuscript and Supplementary Material. The AI assistant was not involved in the conception of the research ideas, results or proofs. The authors take full responsibility for the content.

Supplementary Material

Supplement to "Forecasting extreme trajectories using seminorm representations". The supplement contains all the proofs, a formal comparison of the seminorm representation with the regular variation and spectral tail process frameworks, additional simulation results, visualization of unit cylinders, and complementary empirical analyses including estimation procedures and robustness checks.

References

Alley, R.B., Emanuel, K.A., Zhang, F., 2019. Advances in weather prediction. Science 363, 342–344.

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

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

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

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

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

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, 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, 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, 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, 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, 55–67.

Chahrour, R., Jurado, K., 2021. Recoverability and expectations-driven fluctuations. The Review of Economic Studies 89, 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 , 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, 879 – 917.

Davis, R.A., Mikosch, T., 2009. The extremogram: A correlogram for extreme events. Bernoulli 15, 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, 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, 191 – 234.

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

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

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

Giancaterini, F., Hecq, A., Jasiak, J., Neyazi, A.M., 2025. Bubble detection with application to green bubbles: A noncausal approach. arXiv:2505.14911.

Giancaterini, F., Hecq, A., Morana, C., 2022. Is climate change time-reversible? Econometrics 10.

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

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

Gourieroux, C., Jasiak, J., Monfort, A., 2020. Stationary bubble equilibria in rational expectation models. Journal of Econometrics 218, 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, 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. arXiv:2506.04046.

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

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, 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., 2017a. Do seasonal adjustments induce noncausal dynamics in inflation rates? Econometrics 5, 48.

Hecq, A., Telg, S., Lieb, L., 2017b. 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, 325–352.

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

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

Janßen, A., Segers, J., 2014. Markov tail chains. Journal of Applied Probability 51, 1133–1153.

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

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, 1992 – 2035.

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

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

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

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. doi:10.1007/978-3-031-59156-3.

Planinić, H., Soulier, P., 2017. The tail process revisited. Extremes , 1–29.

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

de Truchis, G., Fries, S., Thomas, A., 2026. Supplement to "forecasting extreme trajectories using seminorm representations".

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

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