Anticipative alpha-stable linear processes for time series analysis : conditional dynamics and estimation

Sébastien Fries

université
PARIS-SACLAY

NNT : 2018SACLG005

Processus linéaires α-stables anticipatifs pour l'analyse des séries temporelles: dynamique conditionnelle et estimation

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'Ecole nationale de la statistique et de l'administration économique

Ecole doctorale n°574 Ecole Doctorale de Mathématiques Hadamard (EDMH)
Spécialité de doctorat : Mathématiques appliquées

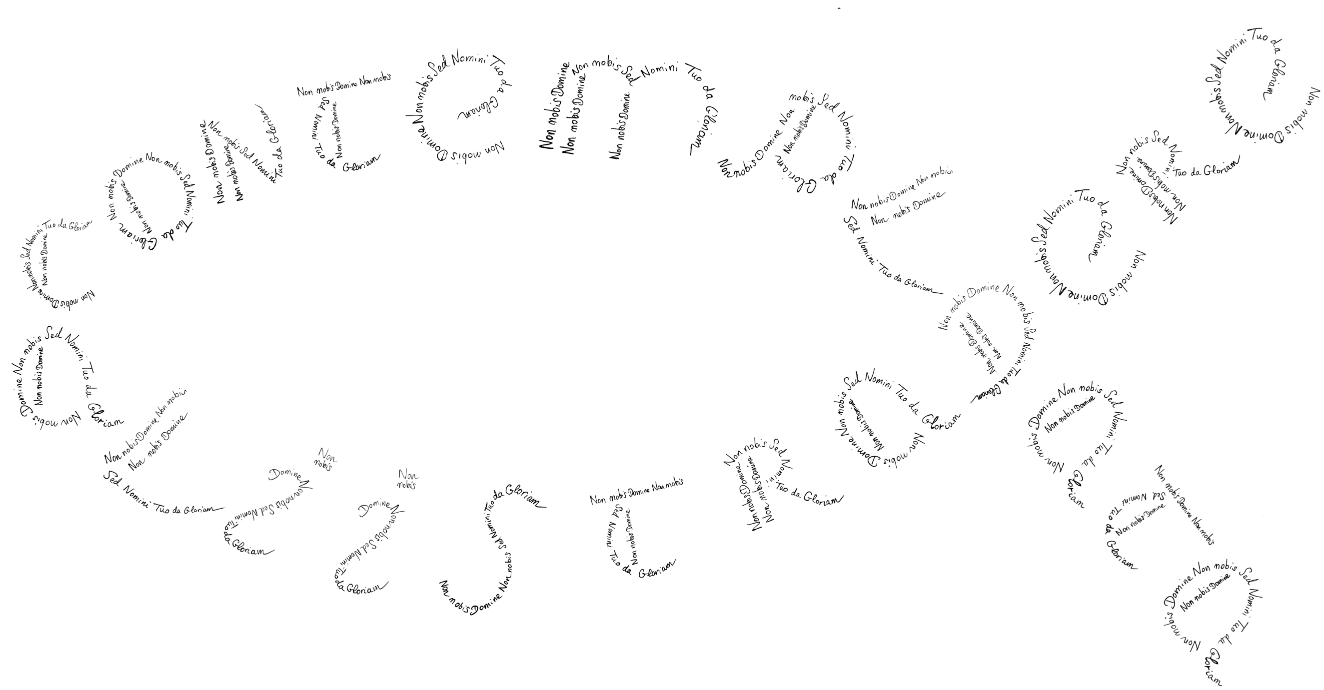
Thèse présentée et soutenue à Palaiseau, le 4 Décembre 2018, par

M. SÉBASTIEN FRIES

Composition du Jury :

M. Jean-Marc Bardet Professeur, Université Paris I - Panthéon-Sorbonne	Rapporteur
Mme. Frédérique Bec Professeur, Université de Cergy-Pontoise	Examinateur
M. Christian Francq Professeur, Université de Lille et ENSAE-CREST	Président
M. Alain Hecq Professeur, Maastricht University	Examinateur
M. Anders Rahbek Professeur, University of Copenhagen	Rapporteur
M. Jean-Michel Zakoïan Professeur, Université de Lille et ENSAE-CREST	Directeur de thèse

Thèse de doctorat


Non nobis Domine Non nobis Sed Nomini Tuo da Gloriam

En quelques sortes qu'arrive ce pressentiment secret des choses futures, on ne saurait voir que ce qui est. Or ce qui est déjà, n'est point à venir, mais présent. Ainsi, lorsqu'on dit que l'on voit les choses futures, ce ne saurait être elles-mêmes, puisqu'elles ne sont pas encore ; mais c'est peut-être leur cause ou leur signe que l'on voit, lesquels sont déjà. [...] *Lorsque j'aperçois l'aurore, je prévois aussitôt que le soleil va se lever: ce que j'aperçois est présent, et ce que je prédis est à venir.* [...] *Cette aurore même, laquelle je vois dans le ciel, n'est pas le lever du soleil, ni cette imagination que je conçois dans mon esprit n'est pas non plus son lever; mais ce sont deux choses, lesquelles sont présentes, qui me font prédire le lever du soleil qui est à venir.*

In whatever manner this secret preconception of future things may be, nothing can be seen except what exists. But what exists now, is not future, but present. When, therefore, they say that future events are seen, it is not the events themselves, for they do not exist as yet, but perhaps, instead, their causes and their signs are seen, which already do exist. [...] *I see the dawn, I predict that the sun is about to rise. What I see is in time present, what I predict is in time future. That dawn which I see in the sky is not the rising of the sun, nor is that imagination in my mind. These two are seen in time present, in order that the event which is in time future may be predicted.*

Saint Augustin, 13 November 354 AD – 28 August 430 AD, bishop of Hippo Regius in the Roman province of Numidia (modern-day Annaba in Algeria), *Confessions*, Book XI, Chapter XVIII.

Combination of translations from
Joseph G. Pilkington and Albert C. Outler (English),
Arnauld d'Andilly and Joseph Trabucco (French).

Acknowledgements

Les trois années de cette thèse arrivant à leur terme, il est temps de rassembler les résultats et de remercier ceux qui ont contribué à rendre cette aventure possible.

Je suis infiniment reconnaissant à mon directeur, Jean-Michel Zakoïan, pour tout le temps qu'il m'a consacré depuis mes débuts dans le monde de la recherche, en thèse de master, jusqu'au bout de ces trois années de thèse de doctorat. Son extraordinaire patience et son sens de la pédagogie auront *in fine* eu raison de mes interminables interrogations et indénombrables tentatives de réécriture. Ses conseils m'auront été d'une valeur inestimable pour apprendre à conduire ma recherche, à en communiquer et valoriser les résultats. Je suis très reconnaissant, en particulier, de la confiance qu'il m'a témoigné en me laissant librement entreprendre le deuxième, puis le troisième article de ma thèse, malgré les obstacles qu'avaient dû surmonter le premier et ceux qui s'annonçaient sur ce chemin.

I am extremely grateful to Professors Anders Rahbek and Jean-Marc Bardet for having accepted and taken the time to review this manuscript, and to Professors Frédérique Bec, Alain Hecq and Christian Francq for having accepted to take part in the jury. I am thankful for their interests in this work and for the fruitful discussions on the topics covered here.

Mes années passées au CREST ont été merveilleuses ; la passion pour la recherche chez les professeurs et les doctorants est palpable et entraînante. Je suis reconnaissant pour tous les échanges que j'ai pu avoir avec eux. Je remercie en particulier les professeurs Christian Gouriéroux, Jean-David Fermanian, Francesco Violante, Elena Dumitrescu, Gilles de Truchis, Yang Lu, Pierre Alquier, Guillaume Lecué, les compagnons de thèse Alexis, Anna, Benjamin, Clara, Ishak, Nicolas, Ophélie, Sean, Yiyi, et le personnel administratif du laboratoire, toujours présent pour aider dans toutes les démarches, en particulier Fanda Traoré et Edith Verger.

Le soutien de ma famille fût essentiel. Je suis infiniment reconnaissant envers mes parents et envers mes sœurs pour tous ces précieux moments à Rueil, en Provence, au Kazakhstan. Leur présence indéfectible m'a permis de traverser sereinement ces années de travail intense, parsemées de démonstrations résistantes, d'impasses et de rebroussements. Je remercie ma mère pour m'avoir inspiré sa témérité naturelle (sans laquelle j'aurais sans doute prudemment évité de me lancer dans une thèse). Il m'aura fallu faire un doctorat pour comprendre enfin le sens de ses invitations à «regarder dans la littérature» qui rythmaient mes études depuis le lycée, émergeant aussitôt qu'un exercice résistât. Je remercie mon père pour m'avoir introduit aux mathématiques, depuis la rue Charles Floquet jusqu'aux classes préparatoires, et pour m'en avoir transmis le goût. La thèse n'aurait pu porter que sur deux domaines: les mathématiques ou la géologie. Ne parvenant pas à ordonner les différents noms d'ammonites et d'époques (Hypacanthoplites jacobi, Duffrenoyia furcata, Deshayesites deshayesi, Triceratopseras duratruve - cherchez l'intrus), le choix s'est rapidement restreint. Je remercie chacune de mes merveilleuses sœurs qui toutes semblent sortir droit d'une mythologie ancienne ou de contes slaves: Isabelle, Astrid, Alexandra, Béatrice et Caroline. Chacune est une inestimable source d'inspiration. Je remercie Sergio pour son soutien actif et permanent au lobby du 7 Wonders, Rails et Citadelles. Je remercie Nikita pour son modèle de persévérance (au-delà de 42km, les courses à pied ne portent plus de noms identifiables par le commun des mortels). Je remercie Antoine et Véra pour leur irremplaçable accompagnement lors de la rédaction finale de la thèse. Je remercie toute ma famille pour m'avoir soutenu tout au long de ce chemin.

Je suis extrêmement reconnaissant aux compagnons plus ou moins lointains, de la première heure ou de plus tard, ~~de ne pas avoir tenté de dérailler~~ de m'avoir soutenu activement dans mes efforts tout au long de ce chemin : Charles-Antoine alias Charly, Corentin, Pierre (avec un «j» et pas un «g»), Pierre (le byzantiniste), Orso (qui aura toujours une partie IV d'avance), Marco, Jeanne, Blandine, Timothée, Raphaël Gomez, Michael N'Dété, Benoît et Andréa Gantelmi d'Ille, Aurélien Duhem.

Je remercie mes co-auteurs de la Banque de France, Jean-Stéphane Mésonnier, Sarah Mouabbi et Jean-Paul Renne, pour la passionante collaboration que nous avons eu autour de l'estimation des taux d'intérêts naturels dans la zone Euro. J'ai beaucoup appris à leur contact et c'est à eux et à leurs efforts que je dois ma première publication.

Je remercie du fond du cœur Anastasia Nikolaevna Dyudkina, de la lignée des tsars à s'y méprendre, pour son soutien en toutes circonstances, dans les joies comme dans les peines, et pour m'avoir insufflé son optimisme capable de déplacer les montagnes.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Résumé substantiel des travaux en français

1.1.1 Contexte

La collection d'un ensemble de mesures au cours du temps, éventuellement dans des conditions contrôlées, permet à un observateur d'entreprendre l'étude quantitative de l'évolution d'un phénomène. Certaines de ces études aboutissent à des schémas quasi-mécaniques prétendant pouvoir déduire avec quasi-certitude l'évolution future dudit phénomène, étant observé un certain état. D'autres en revanche, si elles parviennent à identifier certains comportements typiques au cours du temps, en arrivent à attribuer une place prépondérante à un aléa dont les causes semblent échapper à l'analyse. Dans l'approche probabiliste des phénomènes temporels erratiques, la recherche d'un schéma mécanique cède ainsi le pas à celle d'une «bonne» approximation par un processus aléatoire. Plutôt que de révéler explicitement des relations de cause à effet entre différentes mesures et grandeurs, l'approximation en question se doit seulement de répliquer «au mieux» la dynamique du système observé en vue de fournir des prévisions fiables. Dans une de ses acceptions les plus larges, cette approche suppose que la série de mesures du phénomène, disons (X_t), obéit à une dynamique de la forme

$$X_t = \varphi(X_{t-1}, X_{t-2}, \dots; \varepsilon_t),$$

pour une certaine fonction φ décrivant la dépendance de l'observation présente à l'évolution passée et où (ε_t) est une suite d'erreurs indépendantes et identiquement distribuées (i.i.d.). Le cas où φ est linéaire constitue sans doute le point de départ historique de la formalisation des séries temporelles et a donné lieu à une littérature massive, à la fois théorique et appliquée. Yule (1927) [130] fît le premier usage de ces modèles linéaires, dits *autoregressifs*, pour modéliser la série dénombrant les tâches solaires sur la période 1749-1924 et publiée par Alfred Wolfer en 1925 [126].[1] Les recherches sur les fondements théoriques des processus

[1] Devenue un classique de la discipline et souvent intitulée *Wolfer's sunspot series*, la constitution de cet ensemble de mesures fût cependant initiée et conduite principalement par Johann Rudolph Wolf à partir de 1848 [125], dont Wolfer devint l'assistant

autoregressifs (AR) et des séries temporelles linéaires en général suivirent peu après (voir par exemple Mann et Wald (1943) [98], Kendall (1944) [82], Barlett (1946) [6]). Des applications à divers domaines –économie, télécommunication, géophysique, astronomie entre autres– ont émergé et se poursuivent jusqu'à aujourd'hui, accélérées par la parution d'ouvrages méthodologiques tels ceux de Box et Jenkins (1970) [14] et de Brockwell et Davis (1991) [19] ainsi que par l'automatisation informatique de certaines procédures d'analyse. Dans le cadre des processus AR, on cherche à approximer la dynamique d'une série d'observations par un processus vérifiant une équation de récurrence stochastique linéaire de la forme

$$\varphi(B)X_t = \varepsilon_t, \tag{1.1}$$

où $\varphi(z) := 1 + \sum_{i=1}^{p} \varphi_i z^i$, $z \in \mathbb{C}$, est un polynôme à coefficients réels de degré $p \in \mathbb{N}$, B est l'opérateur retard ($BX_t := X_{t-1}$) et (ε_t) est une séquence i.i.d. Les équations récurrentes stochastiques de la forme (1.1) admettent un unique processus solution strictement stationnaire si et seulement si le polynôme φ n'admet pas de racine sur le cercle unité du plan complexe:

$$\varphi(z) \neq 0, \qquad \text{pour} \quad |z| = 1. \tag{1.2}$$

La variable aléatoire X_t à l'instant $t \in \mathbb{Z}$ dépend alors en général de tous les termes de la suite i.i.d. (ε_t). Très souvent toutefois, le processus approximant est supposé satisfaire (1.1) non pas avec (1.2) mais en supposant que φ n'admet pas de racine sur ou à l'intérieur du cercle unité:

$$\varphi(z) \neq 0, \qquad \text{pour} \quad |z| \leq 1, \tag{1.3}$$

qui est nécessaire et suffisante pour que l'unique solution (X_t) stationnaire ne dépende que des termes «passés» $\{\varepsilon_s : \ s \leq t\}$ de la suite i.i.d., à tout indice $t \in \mathbb{Z}$. La solution, dite alors *non-anticipative*, *causale* ou *de phase minimale*, serait ainsi conforme à une certaine intuition de la causalité lorsque t représente le temps : le présent du processus ne serait déterminé que par les évènements passés. Lorsque la condition (1.3) n'est pas imposée, des solutions dites *anticipatives*, *non-causales* ou *de phase non-minimale*, apparaissent, dépendant des valeurs «futures» de la séquence (ε_t) et de ce fait considérées comme «contre-naturelles». Si les solutions stationnaires anticipatives ont été généralement écartées pour l'analyse des signaux temporels, les processus AR ne satisfaisant pas (1.3) ont été considérés dans d'autres configurations. Ainsi, plutôt que de considérer la solution stationnaire de (1.1), une littérature statistique a étudié les processus partant de conditions initiales en $t = 0$ puis suivant une dynamique dictée par (1.1). Lorsque φ admet des racines à l'intérieur du cercle unité, ce type de processus suit alors des trajectoires non-stationnaires explosives. L'estimation du polynôme φ dans ce cadre a fait l'objet de plusieurs articles (Rubin (1950) [114], White (1958) [123], Anderson (1959) [2], Rao (1961) [108], Stigum (1974) [120], Lai et Wei (1983) [84], Breton et Pham (1989) [17]). Par ailleurs, lorsque l'indice t ne représente pas le temps mais indice un domaine spatial ou fréquentiel, l'interprétation «causale» ou «non-causale» des solutions stationnaires anticipatives est sans objet. Dans ce type de contexte, les AR de phase non-minimale, considérés uniquement en tant que «filtres»,

en 1875 (voir Izenman (1983) [76] et les études qui y sont citées).

ont été au centre de problèmes de déconvolution en géophysique, traitement de signaux vocaux, télécommunication et astronomie (voir par exemple Wiggins (1978)[124], Benveniste, Goursat et Ruget (1980) [11], Donoho (1981) [45], Scargle (1981) [119], Godfrey et Rocca (1981)[58], Lii et Rosenblatt (1982, 1988, 1992, 1996) [90, 91, 92, 93], Giannakis et Mendel (1989) [56], Giannakis et Swami (1990) [57], Gassiat (1990a, 1990b, 1993) [53, 54, 55], Breidt, Davis, Lii et Rosenblatt (1991) [16], Chi et Kung (1995) [29], Chien, Yang et Chi (1997) [28], Andrews, Davis et Breidt (2007)[5]).

Récemment toutefois, des méthodes d'estimation n'imposant pas la contrainte de causalité ont été appliquées à des séries de prix et volumes d'action, de chômage et d'inflation et ont favorisé des modèles linéaires non-causaux au détriment de leurs homologues causaux (Huang et Pawitan (2000) [75], Breidt, Davis et Trindade (2001) [15], Andrews, Calder et Davis (2009) [3], Wu et Davis (2010) [129], Wu (2011) [127], Lanne et Saikkonen (2011) [87], Lanne, Luoto et Saikkonen (2012) [85]). A la croisée des processus anticipatifs et de la théorie des valeurs extrêmes, il a été de plus remarqué que des solutions stationnaires anticipatives de (1.1) démontraient des dynamiques temporelles similaires à celles des bulles spéculatives sur les marchés financiers, vues comme des déviations explosives de court-terme dans les séries de prix par rapport à un niveau stationnaire. La modélisation par des processus anticipatifs de ces phénomènes a été trouvée adéquate pour des séries telles que le taux Bitcoin/USD, des séries de prix du pétrole, l'indice Nasdaq ainsi que des séries de volatilité réalisée (Hencic et Gouriéroux (2015) [70], Hecq, Lieb et Telg (2016) [67], Gouriéroux et Zakoïan (2017) [63], Cavaliere, Nielsen et Rahbek (2018) [24]). Une littérature en économétrie et finance a ainsi émergé, faisant des AR anticipatifs son outil central pour l'analyse temporelle de tels phénomènes.

1.1.2 Motivation

En vue de modéliser les bulles spéculatives, Gouriéroux et Zakoïan (2017) [63] proposent et étudient l'AR(1) anticipatif α-stable. Ce processus strictement stationnaire à variance infinie génère des trajectoires présentant des périodes calmes –proches des valeurs centrales– entrecoupées d'épisodes de croissance explosive («l'inflation de la bulle»), s'achevant sur un abrupte retour aux valeurs centrales (le «crash»). Sans doute l'un des processus les plus élémentaires dans la famille des anticipatifs, l'AR(1) stable anticipatif est défini comme la solution strictement stationnaire de l'équation

$$X_t = \rho X_{t+1} + \varepsilon_t, \qquad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, \beta, \sigma, 0), \tag{1.4}$$

où $0 < |\rho| < 1$ et $\mathcal{S}(\alpha, \beta, \sigma, 0)$ désigne la loi stable de paramètre de queue $\alpha \in]0, 2[$, d'asymmétrie $\beta \in [-1, 1]$ et d'échelle $\sigma > 0$. Outre la pure curiosité mathématique, l'étude de l'AR(1) anticipatif est motivée par la possibilité qu'il offre d'inférer les dates de culmination et d'évanouissement des bulles, qui serait d'un intérêt évident pour les gestionnaires de portefeuilles mais également pour la gestion des risques et le régulateur. Cela requiert cependant de connaître sa dynamique conditionnelle, qui, contrairement à celle de son homologue causal, est non triviale à obtenir du fait de la dépendance entre les observations passées du processus

$\{X_s,\ s \leq t\}$ et les erreurs «futures» $\{\varepsilon_s,\ s \geq t\}$. Cet aspect probabiliste a reçu le moins d'attention de la littérature, ce qui peut s'expliquer par la nature spatiale/fréquentielle des problèmes de déconvolution qui ont motivé la recherche sur les processus non-causaux au cours des années 1980 et 1990. La dualité passé observé/futur à inférer est absente du contexte spatial où l'ensemble du domaine est potentiellement observable. La problématique de prédiction des processus non-causaux a toutefois été abordé par Rosenblatt (1995, 2000) [111, 112] dans le contexte de variance finie. Gouriéroux et Zakoïan (2017) [63] ont obtenu des résultats étonnants sur la dynamique conditionnelle de l'AR(1) anticipatif α-stable. Après avoir établi son caractère markovien, ils montrent l'existence de moments conditionnels d'ordre plus élevé que les moments marginaux:

$$\mathbb{E}[|X_{t+1}|^\gamma | X_t] < +\infty, \quad \text{pour} \quad \gamma < 2\alpha + 1,$$

bien que $\mathbb{E}[|X_t|^\alpha] = +\infty$, car X_t est marginalement α-stable.[2] Dans le cas où la séquence (ε_t) est symmétrique α-stable ($\beta = 0$), ils ont également montré que l'espérance conditionnelle s'écrit

$$\mathbb{E}[X_{t+1}|X_t] = \rho^{<\alpha-1>} X_t, \tag{1.5}$$

où $x^{<b>} := \text{sign}(x)|x|^b$ pour tout $x, b \in \mathbb{R}$, et si (ε_t) est marginalement de loi de Cauchy, ($\alpha = 1$, $\beta = 0$), la variance conditionnelle est alors une fonction quadratique de l'observation présente:

$$\mathbb{V}(X_{t+1}|X_t) = \left(\frac{1}{|\rho|} - 1\right) X_t^2 + \frac{\sigma^2}{|\rho|(1 - |\rho|)}.$$

Ces propriétés tranchent singulièrement avec les résultats connus pour les processus non-anticipatifs, que ce soit dans le contexte de variance finie ou infinie. En particulier, la discrépance entre moments marginaux et conditionnels laisse entrevoir un certain «excès de prédictabilité» des processus anticipatifs par rapport à leurs homologues non-anticipatifs. Au moment d'entreprendre cette thèse, la dynamique conditionnelle de l'AR(1) anticipatif α-stable reste seulement partiellement comprise tandis que celle des processus d'ordres plus élevés est totalement à explorer. A ce premier aspect probabiliste s'ajoute une problématique statistique incontournable en vue d'une utilisation pratique de cette classe de modèles, à savoir l'estimation des paramètres étant donnée une série d'observations.

Cette thèse s'est fixée pour objectif l'étude des processus linéaires α-stable anticipatifs dans le cas général, et en priorité, l'étude de leur dynamique conditionnelle.

1.1.3 Synthèse des principaux résultats

Le deuxième chapitre commence par explorer les aspects probabilistes et statistiques des processus AR d'ordres supérieurs, c'est-à-dire dont le polynôme caractéristique comporte des racines à la fois à l'intérieur

[2] Un résultat similaire a été obtenu par Cambanis et Fakhre-Zakeri (1995) [20] pour l'AR(1) α-stable non-anticipatif en temps inversé.

et à l'extérieur du cercle unité. Pour ces processus AR, dits *mixtes* causaux/non-causaux, la dynamique peut s'écrire sous la forme factorisée

$$\psi(F)\phi(B)X_t = \varepsilon_t,$$

où ψ et ϕ sont deux polynômes à coefficients réels d'ordre respectif p et q, vérifiant tous deux (1.3) et où F est l'opérateur avance ($F := B^{-1}$, i.e., $FX_t = X_{t+1}$). L'extension des propriétés probabilistes est étudiée sur la base des résultats et des techniques de Gouriéroux et Zakoïan (2017) [63] d'une part, et d'autre part la décomposition du processus mixte (X_t) en partie causale et partie non-causale de Lanne et Saikkonen (2011) [87] (voir aussi Gouriéroux et Jasiak, 2016 [59]). On montre que (X_t) est markovien d'ordre $p+q$ et que la discrépance entre moments conditionnels et marginaux n'est présente que lorsque $p \geq 1$, à savoir, lorsque la composante anticipative du processus est non-triviale. Des formules fermées pour les moments conditionnels sont obtenues dans des cas particuliers, par exemple lorsque $\psi(F) = 1 - \psi F$ et $\beta = 0$, l'espérance conditionnelle est linéaire et s'écrit pour tout $q \geq 1$

$$\mathbb{E}[X_t|\mathscr{F}_{t-1}] = \psi^{<\alpha-1>}X_{t-1} + (1 - \psi^{<\alpha-1>}B)(\phi_1 X_{t-1} + \ldots + \phi_q X_{t-q}),$$

où l'on note $\phi(z) = 1 - \phi_1 z - \ldots - \phi_q z^q$ et $(\mathscr{F}_t)$ la filtration canonique du processus (X_t). Si de surcroît $\alpha = 1$, on montre que (X_t) admet la représentation causale semi-forte

$$\begin{aligned}(1 - \operatorname{sign}(\psi)B)\phi(B)X_t &= \sigma_t \eta_t,\\ \sigma_t^2 &= \left(\frac{1}{|\psi|} - 1\right)(X_{t-1} - \phi_1 X_{t-2} - \ldots - \phi_q X_{t-q-1})^2 + \frac{\sigma^2}{|\psi|(1-|\psi|)},\end{aligned}$$

où $\mathbb{E}\left[\eta_t | \mathscr{F}_{t-1}\right] = 0$ et $\mathbb{E}\left[\eta_t^2 \middle| \mathscr{F}_{t-1}\right] = 1$.

Du point de vue statistique, ce chapitre aborde l'estimation par les moindres carrés du modèle

$$\psi_0(F)\phi_0(B)X_t = \varepsilon_t,$$

où $\psi_0(z) = 1 - \sum_{i=0}^{p} \psi_{0i} z^i$, $\phi_0(z) = 1 - \sum_{i=0}^{q} \phi_{0i} z^i$, à partir d'observations $X_1, \ldots, X_n$ du processus. Contrairement à la méthode du maximum de vraisemblance, l'approche par les moindres carrés n'exige pas de supposer que les ε_t suivent une distribution précise, ce qui la rend plus robuste aux erreurs de spécification. En lieu d'être distribuée selon une loi α-stable, on suppose seulement que les ε_t appartiennent au domaine d'attraction d'une loi stable, c'est-à-dire que

$$\mathbb{P}(|\varepsilon_0| > x) = x^{-\alpha}L(x), \qquad \text{et} \qquad \lim_{x\to\infty} \frac{\mathbb{P}(\varepsilon_0 > x)}{\mathbb{P}(|\varepsilon_0| > x)} \to c \in [0,1],$$

pour L une fonction à variations lentes à l'infini.[3] Le statisticien fait face à plusieurs inconnues: les ordres p et q des polynômes ψ_0 et ϕ_0, leurs coefficients, et la décomposition causale/non-causale de la dynamique, à savoir, lequel des deux polynômes est associé à l'opérateur avance F, et lequel est associé à l'opérateur

[3] $\lim_{x\to\infty} L(tx)/L(x) = 1, \forall t > 0$.

retard B. Une difficulté que l'on rencontre pour l'estimation consiste en ce que le processus (X_t) admet plusieurs représentations polynomiales avec erreurs dépendantes mais en apparence non-autocorrélées. En effet, pour tout polynôme η_0^* obtenu à partir de $\psi_0(z)\phi_0(z)$ en remplaçant une ou plusieurs racines par leur inverse, on a

$$\eta_0^*(B)X_t = \zeta_t^*,$$

où (ζ_t^*) est un processus dit «*all-pass*» dont les autocorrélations empiriques tendent vers zéro avec la taille de l'échantillon. Il est ainsi attendu que l'on ne pourra pas retrouver la structure causale/non-causale de la dynamique avec les seuls moindres carrés, qui n'exploitent que l'information provenant des moments d'ordre deux. On propose donc une procédure mettant à profit cette limitation des moindres carrés en la couplant avec le phénomène d'«*extreme clustering*».[4] On tire avantage de la limitation des moindres carrés en constatant que l'on a en particulier

$$\psi_0(B)\phi_0(B)X_t = \zeta_t, \tag{1.6}$$

pour un certain processus *all-pass* (ζ_t). La procédure consiste en trois étapes.

1) A l'aide de (1.6) et en supposant connu le degré $p+q$ du polynôme $\eta_0(z) := \psi_0(z)\phi_0(z)$, on montre tout d'abord la convergence en probabilité et en distribution de l'estimateur de Yule-Walker de la régression de X_t sur $X_{t-1}, \dots, X_{t-p-q}$ vers les coefficients de η_0 en utilisant les techniques développées par Davis et Resnick (1986) [41].

2) En pratique, le degré $p+q$ étant inconnu, on introduit un test de type portmanteau afin de détecter une éventuelle «autocorrélation empirique» dans les résidus après estimation. On est ainsi en mesure de rejeter les modèles sous-spécifiés (d'ordre plus petit que $p+q$). En commençant par des ordres petits, on peut estimer le modèle, tester sa validité, et répéter cette étape en incrémentant l'ordre jusqu'à ne plus rejeter l'hypothèse de validité.

3) Une fois l'ordre $p+q$ validé et le polynôme $\eta_0(z)$ estimé, on obtient alors un estimateur consistant des racines de $\psi_0(z)\phi_0(z)$ et il suffit d'identifier quelles racines «appartiennent» à $\psi_0(z)$ et lesquelles «appartiennent» à $\phi_0(z)$ afin d'identifier la structure causale/non-causale de la dynamique. Toute allocation des racines entre composantes causale et non-causale aboutit à une représentation *all-pass* du processus (X_t) et l'on ne peut donc pas déterminer par un test d'autocorrélation des résidus si la «vraie» dynamique a été identifiée. Si toutes les représentations *all-pass* ont des erreurs non-autocorrélées, seules la représentation originelle du processus admet des erreurs indépendantes. Dans ce contexte de variance infinie, on justifie par l'intermédiaire de techniques provenant de la théorie des processus ponctuels que la dépendance des erreurs des représentations *all-pass* induit un phénomène d'*extreme clustering* : les erreurs extrêmes apparaissent en *clusters*, et ces *clusters* sont d'autant plus grand que la dépendance est forte. Ce phénomène ne se manifeste pas dans les erreurs de la représentation originelle, qui sont indépendantes, et les valeurs extrêmes y apparaissent isolées les unes des autres. La troisième étape consiste à mesurer le degré de *clustering* des extrêmes dans les résidus de chaque allocation possible des racines entre composantes causale et non-causale.

[4] On pourrait tenter de traduire cette expression par «phénomène des amas d'extrêmes».

L'allocation correspondant à la «vraie» structure causale/non-causale est, en principe, la seule pour laquelle aucun indice d'amassement ne devrait être détecté.
Des simulations illustrent la validité empirique de chaque étape et le chapitre se conclut sur une application à six séries financières similaires.

L'étroitesse du cadre dans lequel la dynamique conditionnelle a pu être obtenue au deuxième chapitre appelle une nouvelle approche. Le troisième chapitre revisite l'AR(1) anticipatif en mobilisant la théorie des distributions stables multivariées, et en particulier la littérature des années 1990 sur les moments conditionnels des vecteurs stables bivariés (Hardin, Samorodnitsky, Taqqu (1991) [64], Cioszek-Georges, Taqqu (1994, 1995a, 1995b, 1998) [30, 31, 32, 33]). Tout vecteur stable $\boldsymbol{X} = (X_1, \dots, X_d)$ est caractérisé par une unique paire $(\Gamma, \boldsymbol{\mu}^0)$, où Γ est une mesure borélienne finie sur la sphère euclidienne $S_d = \{\boldsymbol{s} \in \mathbb{R}^d : \ \ \|\boldsymbol{s}\|_e = 1\}$, avec $\|\cdot\|_e$ dénotant la norme euclidienne, et $\boldsymbol{\mu}^0 \in \mathbb{R}^d$ un vecteur fixe, telle que

$$\mathbb{E}\Big[e^{i\langle \boldsymbol{u}, \boldsymbol{X}\rangle}\Big] = \exp\left\{ -\int_{S_d} |\langle \boldsymbol{u}, \boldsymbol{s}\rangle|^{\alpha}\Big(1 - i\,\mathrm{sign}(\langle \boldsymbol{u}, \boldsymbol{s}\rangle) w(\alpha, \langle \boldsymbol{u}, \boldsymbol{s}\rangle)\Big)\Gamma(d\boldsymbol{s}) + i\,\langle \boldsymbol{u}, \boldsymbol{\mu}^0\rangle \right\}, \qquad \forall \boldsymbol{u} \in \mathbb{R}^d, \tag{1.7}$$

avec $w(\alpha, s) = \mathrm{tg}\left(\frac{\pi\alpha}{2}\right)$, si $\alpha \neq 1$, $w(1, s) = -\frac{2}{\pi}\ln|s|$, $s \in \mathbb{R}$ (Théorème 2.3.1 de Samorodnitsky, Taqqu, 1994 [117]). La mesure Γ est appelée *mesure spectrale* de $\boldsymbol{X}$ et encode l'information à propos de la dépendance entre les composantes du vecteur, et la paire $(\Gamma, \boldsymbol{\mu}^0)$ est sa *représentation spectrale*. On montre que pour (X_t) solution de (1.4), le vecteur (X_t, X_{t+h}) est bivarié α-stable et sa fonction caractéristique est de la forme (1.7) avec pour mesure spectrale

$$\Gamma_h = \frac{\bar{\sigma}^{\alpha}}{2}\sum_{\vartheta \in S_1}\left[\left(1 - |\rho|^{\alpha h} + \left(1 - \left(\rho^{<\alpha>}\right)^h\right)\vartheta\bar{\beta}\right)\delta_{\{(\vartheta, 0)\}} + \left(1 + |\rho|^{2h}\right)^{\alpha/2}(1 + \vartheta\bar{\beta})\delta_{\{\vartheta \boldsymbol{s}_h\}}\right], \tag{1.8}$$

où $S_1 = \{-1, +1\}$, $\delta_{\{x\}}$ est la masse de Dirac au point $x \in \mathbb{R}^2$, $\bar{\sigma}^{\alpha} = \dfrac{\sigma^{\alpha}}{1 - |\rho|^{\alpha}}$, $\bar{\beta} = \beta\dfrac{1 - |\rho|^{\alpha}}{1 - \rho^{<\alpha>}}$, et $\boldsymbol{s}_h = \dfrac{(\rho^h, 1)}{\sqrt{1 + |\rho|^{2h}}} \in S_2$. Toute distribution de masse finie sur la sphère unité définie une mesure spectrale valide d'un certain vecteur stable. Il est remarquable qu'ici, Γ_h est un objet purement discret chargeant ou bien deux, ou bien quatre atomes. Une fois ce constat acté, plusieurs résultats de la littérature sur les vecteurs stables bivariés sont immédiatement applicables. L'existence des moments conditionnels de X_{t+h} sachant X_t jusqu'à l'ordre $2\alpha + 1$ découle du fait que $\int_{S_2} |s_1|^{-\nu}\Gamma_h < +\infty$ pour tout $\nu \geq 0$ et $h \in \mathbb{N}^*$ (voir le Théorème 5.1.3 de Samorodnitsky, 1994, [117] et les travaux de Cioszek-George et Taqqu). Les Théorèmes 5.2.2 et 5.2.3 de Samorodnitsky et Taqqu (1994) [117] donnent la forme fonctionnelle de l'éspérance conditionnelle en termes d'une mesure spectrale arbitraire et permet d'étendre le résultat (1.5) de Gouriéroux et Zakoïan (2017) [63] à tout $\beta \in [-1, 1]$, révélant la non-linéarité de $x \longmapsto \mathbb{E}[X_{t+h}|X_t = x]$ dans le cas général. La forme de la variance conditionnelle a également fait l'objet d'un article par Cioszek-George et Taqqu (1995) [31], toutefois, la preuve dans le cas asymmétrique $\alpha \neq 1$ est omise, et le cas asymmétrique $\alpha = 1$ n'est pas traité. Les résultats pour la variance conditionnelle sont complétés, et l'on contribue à la littérature sur les vecteurs stables bivariés par l'obtention des formes fonctionnelles des moments d'ordre trois et quatre (à savoir, asymétrie et kurtosis). Les formes des quatre premiers moments conditionnels étant établies pour des vecteurs (bivariés) stables généraux, on déduit alors leurs formes particulières pour (X_t, X_{t+h}), et ce pour toute paramétrisation admissible du processus (X_t), en substituant par la mesure spectrale (1.8). La dépendance entre la réalisation future et l'observation présente se révèle fort complexe, comme on peut le constater au Théorème 3.2.1. Toutefois, la dynamique de l'AR(1) α-stable anticipatif se simplifie drastiquement lors des évènements extrêmes dans le cas $\rho > 0$. Lorsque $sx \to +\infty$, $s = \pm 1$, on montre que les moments conditionnels admettent les équivalents[5]

$$\begin{aligned}
\mu(x,h) &\sim (\rho^{-h}x)\rho^{\alpha h}, && \text{si} \quad \alpha \in (0,2),\\
\sigma^2(x,h) &\sim (\rho^{-h}x)^2\rho^{\alpha h}(1-\rho^{\alpha h}), && \text{si} \quad \alpha \in (1/2,2),\\
\gamma_1(x,h) &\longrightarrow s\frac{1-2\rho^{\alpha h}}{\sqrt{\rho^{\alpha h}(1-\rho^{\alpha h})}}, && \text{si} \quad \alpha \in (1,2),\\
\gamma_2(x,h) &\longrightarrow \frac{1}{\rho^{\alpha h}} + \frac{1}{1-\rho^{\alpha h}} - 6, && \text{si} \quad \alpha \in (3/2,2),
\end{aligned}$$

où $\mu(x,h)$, $\sigma^2(x,h)$, $\gamma_1(x,h)$ et $\gamma_2(x,h)$ désignent respectivement l'espérance, variance, asymétrie et excès de kurtosis de X_{t+h} sachant $X_t = x$. On peut alors remarquer que les expressions apparaissant à droite correspondent aux quatre premiers moments de la variable aléatoire Z définie par

$$\mathbb{P}(Z = \rho^{-h}x) = \rho^{\alpha h}, \qquad \text{et} \qquad \mathbb{P}(Z = 0) = 1 - \rho^{\alpha h}.$$

Une interprétation inédite de la dynamique de (X_t) émerge de la forme asymptotique des moments conditionnels : lors des épisodes de bulle, (X_t) semble suivre une trajectoire exponentielle de taux d'accroissement ρ^{-1}, et la probabilité conditionnelle de survie de la bulle à horizon h serait donnée par $\rho^{\alpha h}$. On montre au quatrième chapitre une convergence en distribution vers ce comportement lors des évènements extrêmes. Avec les outils appropriés pour définir les processus stables en temps continu –à savoir, les mesures aléatoires α-stables et les intégrales stables– le chapitre obtient des résultats similaires pour le processus d'Ornstein-Uhlenbeck anticipatif.

L'approche suivie dans ce troisième chapitre peut être appliquée à tout processus α-stable (Y_t) pour analyser les moments conditionnels de Y_{t+h} sachant l'observation présente Y_t, ce que l'on illustre avec l'agrégation d'AR(1) définie par

$$X_t = \sum_{j=1}^{J} \pi_j X_{j,t}, \qquad X_{j,t} = \rho_j X_{j,t+1} + \varepsilon_{j,t}, \qquad \varepsilon_{j,t} \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, \beta_j, 1, 0), \tag{1.9}$$

pour π_j des réels strictement positifs et $J \geq 1$. Ce processus a été proposé par Gouriéroux et Zakoïan (2017) [63] afin de lever une limitation de l'AR(1) anticipatif simple, à savoir que ce dernier ne semble capable de générer que des bulles de taux de croissance ρ^{-1} identique d'un évènement extrême à l'autre.

[5] Si $|\beta| = 1$, l'une des queues de la distribution marginale de X_t n'est pas à variation régulière et peut même être totalement absente. Dans ce cas, les valeurs extrêmes sont nécessairement toutes de même signe et le comportement asymptotique n'est valable que pour $x \to +\infty$ ou $x \to -\infty$.

Le processus agrégé (1.9) en revanche, semble produire des trajectoires marquées par des bulles de taux croissance différents ρ_j^{-1}, $j = 1, \ldots, J$. On montre que (X_t, X_{t+h}) est également α-stable pour ce processus et que sa mesure spectrale Γ_h est donnée par

$$\Gamma_h = \sum_{j=1}^{J} \pi_j^{\alpha} \Gamma_{j,h}, \tag{1.10}$$

où $\Gamma_{j,h}$ est la mesure spectrale du vecteur $(X_{j,t}, X_{j,t+h})$ lié à l'AR(1) simple de paramètre ρ_j dans (1.9) et qui est donnée *mutatis mutandis* par (1.8). Par-delà l'existence des moments conditionnels au moins jusqu'à l'ordre $2\alpha + 1$, la forme fonctionnelle des quatre premiers moments conditionnels, et malgré le fait que $Y_{t+h}|Y_t$ ne caractérise pas totalement la distribution conditionnelle pour ce processus non-markovien, (1.10) trahit la similarité des structures de dépendance entre processus stables «simples» et processus stables agrégés. On tirera avantage de cette similarité au chapitre suivant pour proposer une approche unifiée de l'étude des processus stables agrégés et non-agrégés, les premiers démontrant des dynamiques bien plus riches que les derniers.

Le quatrième chapitre s'appuie sur deux indices entraperçus au cours du chapitre abordé précédemment. D'une part la nature α-stable multivariée des parcelles de trajectoire $\boldsymbol{X}_t := (X_{t-m}, \ldots, X_t, X_{t+1}, \ldots, X_{t+h})$, $m \geq 0$, $h \geq 1$, et d'autre part la simplification de la dynamique lorsque le processus s'éloigne de ses valeurs centrales. Cette simplification constatée sur la forme des moments conditionnels peut être obtenue comme une conséquence de la variation régulière des queues de distribution du vecteur $\boldsymbol{X}_t$. Si Γ est la mesure spectrale de $\boldsymbol{X}_t$ sur la sphère euclidienne S_{m+h+1}, on a en effet par application du Théorème 4.4.8 de Samorodnitsky et Taqqu (1994) [117] et du Théorème de Bayes que

$$\mathbb{P}\Big(\boldsymbol{X}_t/\|\boldsymbol{X}_t\|_e \in A \;\Big|\; \|\boldsymbol{X}_t\|_e > x \text{ and } \boldsymbol{X}_t/\|\boldsymbol{X}_t\|_e \in B\Big) \underset{x \longrightarrow +\infty}{\longrightarrow} \frac{\Gamma(A \cap B)}{\Gamma(B)}, \tag{1.11}$$

pour tous boréliens A et B vérifiant des conditions de Γ-continuité. La mesure Γ décrit ainsi complètement la distribution conditionnelle des chemins normalisés $\boldsymbol{X}_t/\|\boldsymbol{X}_t\|_e$ –la «forme» de la trajectoire– quand le vecteur $\boldsymbol{X}_t$ est grand au sens de la norme euclidienne. Dans le cas où (X_t) est un processus moyenne mobile infinie de forme générale

$$X_t = \sum_{k=-\infty}^{+\infty} d_k \varepsilon_{t+k}, \qquad t \in \mathbb{Z}, \tag{1.12}$$

la mesure Γ est explicite en termes de la séquence (d_k) et il serait tentant d'invoquer directement (1.11) afin d'étudier la distribution conditionnelle du chemin futur $(X_{t+1}, \ldots, X_{t+h})$ étant donnée la trajectoire passée $(X_{t-m}, \ldots, X_t)$. Dans un contexte de prévision cependant, (1.11) est en tant que telle de peu d'intérêt du fait de la dépendance de l'évènement conditionnant aux réalisations futures, principalement à travers la norme euclidienne de $\boldsymbol{X}_t$. L'idée de ce chapitre est d'obtenir une version de (1.11) où la norme euclidienne serait remplacer par une semi-norme $\|\cdot\|$ vérifiant

$$\|(x_{-m}, \ldots, x_0, x_1, \ldots, x_h)\| = \|(x_{-m}, \ldots, x_0, 0, \ldots, 0)\|, \tag{1.13}$$

pour tout $(x_{-m}, \dots, x_h) \in \mathbb{R}^{m+h+1}$. Dans ce but, une nouvelle représentation spectrale des vecteurs stables sur la «sphère» unité $C_{m+h+1}^{\|\cdot\|} := \{\boldsymbol{s} \in \mathbb{R}^{m+h+1} : \ \|\boldsymbol{s}\| = 1\}$ relative à une telle semi-norme est explorée. Pour des raisons géométriques évidentes, on préfèrera employer le terme de *cylindre* plutôt que celui de «sphère». On dira qu'un vecteur stable $\boldsymbol{X} = (X_1, \dots, X_d)$ est *représentable* sur un cylindre unité $C_d^{\|\cdot\|}$ relatif à une semi-norme $\|\cdot\|$ s'il existe une paire $(\Gamma^{\|\cdot\|}, \boldsymbol{\mu}^0_{\|\cdot\|})$, où $\Gamma^{\|\cdot\|}$, est une mesure borélienne sur $C_d^{\|\cdot\|}$, $\boldsymbol{\mu}^0_{\|\cdot\|} \in \mathbb{R}^d$, telle que la fonction caractéristique de $\boldsymbol{X}$ soit de la forme (1.7) avec $(S_d, \Gamma, \boldsymbol{\mu}^0)$ remplacée par $(C_d^{\|\cdot\|}, \Gamma^{\|\cdot\|}, \boldsymbol{\mu}^0_{\|\cdot\|})$. Pourvu qu'une telle représentation existe, on montre alors que (1.11) peut effectivement être reformulée en remplaçant la norme euclidienne par la semi-norme correspondante, et en substituant Γ par $\Gamma^{\|\cdot\|}$.

S'il est établi que tout vecteur stable admet une représentation sur la sphère unité relative à toute norme (Théorème 2.3.8, Samorodnitsky et Taqqu, 1994 [117]), on montre qu'il n'en va pas de même sur les cylindres unités relatifs à des semi-normes. Le chapitre s'ouvre donc sur l'étude préliminaire de la *représentabilité* des vecteurs stables en général, et des chemins $\boldsymbol{X}_t$ en particulier, sur de tels cylindres unités. Comme souvent en travaillant avec les distributions stables, le cas $\alpha = 1$ nécessite une considération particulière en présence de lois asymétriques. Ce cas est traité au quatrième chapitre et les résultats sont similaires. Toutefois, afin de ne pas alourdir l'exposition des résultats, il sera supposé dans ce qui suit que $\alpha \neq 1$. Pour une semi-norme quelconque $\|\cdot\|$ sur $\mathbb{R}^d$, on établit qu'un vecteur stable $(X_1, \dots, X_d)$ arbitraire admettra une représentation sur $C_d^{\|\cdot\|}$ si et seulement si[6]

$$\Gamma\Big(\{\boldsymbol{s} \in S_d : \ \|\boldsymbol{s}\| = 0\}\Big) = 0. \tag{1.14}$$

Cette condition peut être comprise de la manière suivante. Intuitivement, compte tenu de (1.11), la mesure spectrale d'un vecteur stable encode l'information à propos de la dépendance extrême de ses composantes en assignant de la masse aux directions de l'espace.[7] Un vecteur stable est totalement caractérisé par la donnée de cette distribution de masse sur les directions de l'espace, qui s'exprime usuellement par l'intermédiaire d'une mesure sur une sphère unité relative à une norme. Puisque les sphères unités sont en bijection avec l'ensemble des directions de l'espace, ces-dernières permettent de caractériser toutes les dépendances extrêmes potentielles d'un vecteur. En revanche, un cylindre unité n'est pas en bijection avec l'ensemble des directions de l'espace et une mesure sur un tel ensemble ne peut pas décrire la dépendance extrême de certains vecteurs stables. La condition (1.14) indique qu'un vecteur stable sera représentable sur le cylindre $C_d^{\|\cdot\|}$ pourvu que les directions problématiques –celles appartenant au noyau de la semi-norme– soient non informatives pour décrire sa dépendance extrême. En dimension 2, un vecteur (X_1, X_2), de mesure spectrale Γ sur la sphère euclidienne, sera représentable sur le cylindre $\{(s_1, s_2) \in \mathbb{R}^2 : \ |s_1| = 1\}$ pourvu que $\Gamma\Big(\{(0,-1), (0,+1)\}\Big) = 0$, c'est-à-dire si le vecteur $(X_1, X_2)/\sqrt{X_1^2 + X_2^2}$ a probabilité zéro de tendre vers

[6] Dans le cas $\alpha = 1$ en présence d'asymétrie, on obtient la condition nécessaire et suffisante légèrement plus forte $\int_{S_d} \Big|\ln \|\boldsymbol{s}\|\Big| \Gamma(d\boldsymbol{s}) < +\infty$. Cette-dernière implique en particulier (1.14) étant donné que $\Big|\ln \|\boldsymbol{s}\|\Big| = +\infty$ pour tout $\boldsymbol{s}$ dans le noyau de la semi-norme.

[7] Par «direction», on entend les classes d'équivalence de $\mathbb{R}^d$ pour la relation «$\equiv$» définie par: $\boldsymbol{u} \equiv \boldsymbol{v}$ si et seulement si il existe $\lambda > 0$ tel que $\boldsymbol{u} = \lambda \boldsymbol{v}$, $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^d$

$(0, \pm 1)$ lorsque la norme $\sqrt{X_1^2 + X_2^2}$ devient grande. En d'autres termes, $\Gamma\Big(\{(0,-1),(0,+1)\}\Big) = 0$ indique que les réalisations de (X_1, X_2) où X_2 est extrême et X_1 est non-extrême, ont zéro probabilité d'occurrence.

Après avoir considéré le cas des vecteurs stables généraux, on considère celui des chemins d'une moyenne mobile (X_t) de forme générale (1.12) pour laquelle on impose *seulement* la condition de sommabilité

$$0 < \sum_{k\in\mathbb{Z}} |d_k|^s < +\infty, \qquad \text{pour un certain} \quad s \in]0,\alpha[\cap [0,1]. \tag{1.15}$$

La moyenne mobile (X_t) n'est donc pas supposée *a priori* anticipative. La propriété (1.11) ne permettra d'étudier la distribution conditionnelle de (X_t) que si l'on peut trouver des chemins de la forme

$$\boldsymbol{X}_t = (\underbrace{X_{t-m}, \dots, X_t}_{\substack{m+1 \\ \text{observations}}}, \underbrace{X_{t+1}, \dots, X_{t+h}}_{\substack{h \\ \text{horizon de prévision}}}),$$

représentables sur $C^{\|\cdot\|}_{m+h+1}$ pour une semi-norme satisfaisant (1.13). Plusieurs cas apparaissent selon la disposition du noyau de la semi-norme. Pour fixer les idées, on considère dans ce chapitre les semi-normes telles que

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

pour tout vecteur $(x_{-m}, \dots, x_h) \in \mathbb{R}^{m+h+1}$, et qui satisfont en particulier (1.13).
En définissant $\mathcal{M} = \{m \geq 1 : \ \exists k \in \mathbb{Z}, \ d_{k+m} = \ldots = d_{k+1} = 0, \ d_k \neq 0\}$, et

$$m_0 = \begin{cases} \sup \mathcal{M}, & \text{si} \quad \mathcal{M} \neq \emptyset, \\ 0, & \text{si} \quad \mathcal{M} = \emptyset, \end{cases}$$

on montre par l'intermédiaire de (1.14) qu'il est nécessaire et suffisant que

$$m_0 < +\infty, \tag{1.16}$$

pour que les chemins $\boldsymbol{X}_t$ soient représentables sur $C^{\|\cdot\|}_{m+h+1}$ pour tout $m \geq m_0$ et $h \geq 1$. En d'autres termes, (X_t) n'induira des chemins représentables sur $C^{\|\cdot\|}_{m+h+1}$ que si toute séquence de zéros consécutifs dans les coefficients (d_k) est soit finie, soit infinie «à gauche». Etonnamment, alors que l'on ne présuppose rien sur la causalité ou non de (X_t), on démontre ainsi que seules les moyennes mobiles anticipatives induisent des chemins $\boldsymbol{X}_t$ représentables sur un cylindre unité approprié au contexte de prévision. Cette condition de représentabilité apporte une lumière nouvelle sur la prédictabilité des extrêmes d'un processus α-stable. Intuitivement, le chemin $\boldsymbol{X}_t$ sera représentable sur $C^{\|\cdot\|}_{m+h+1}$ pourvu que tout évènement extrême affectant les h dernières composantes inobservées ne puissent pas se produire indépendamment d'un évènement extrême sur les $m+1$ premières composantes observées (condition (1.14)). En d'autres termes, tout évènement extrême à venir sur la trajectoire (X_t) doit manifester des signes avant-coureurs. La différence entre processus anticipatifs et non-anticipatifs est évidente de ce point de vue. Pour ces-derniers (l'AR(1) stable non-anticipatif, $Y_t = \rho Y_{t-1} + \eta_t$, par exemple), les évènements extrêmes surgissent soudainement, sans crier gare, sous la forme de saut dont la magnitude suit une loi de puissance de variance infinie. Pour les premiers en revanche, les évènements extrêmes sont identifiables en avance par des tendances ou des signaux trahissant leur survenue prochaine et sont atteints graduellement (l'inflation des bulles de l'AR(1) anticipatif). Pour les processus autoregressifs, qui furent le point de départ de cette thèse, la condition de représentabilité des chemins se simplifie. On considère un ARMA (X_t) donné par

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

où ψ, ϕ, Θ, H sont des polynômes vérifiant (1.3), ψ et Θ (resp. ϕ et H) n'ayant pas de racine commune. Il est alors montré que (X_t) induira des chemins représentables sur un cylindre unité approprié au contexte de prévision si et seulement si

$$\deg(\psi) \geq 1.$$

Dans ce contexte, les AR non-anticipatifs sont donc des processus typiquement pathologiques. Une des forces de cette approche est d'englober naturellement les processus (X_t) résultant de la combinaison linéaire de moyennes mobiles α-stables, appelés *agrégats stables* et définis comme

$$X_t = \sum_{j=1}^{J} \pi_j X_{j,t}, \qquad X_{j,t} = \sum_{k\in\mathbb{Z}} d_{j,k}\varepsilon_{j,t+k}, \qquad \varepsilon_{j,t} \overset{i.i.d.}{\sim} \mathcal{S}(\alpha,\beta_j,1,0), \tag{1.17}$$

où les π_j sont des réels strictement positifs, $J \geq 1$ et où chaque suite de coefficients $(d_{j,k})_k$ vérifie la condition de sommabilité (1.15). Par le biais d'une relation similaire à (1.10) mise au jour dans le troisième chapitre, il est établit que les chemins du processus agrégés (X_t) seront représentables sur un cylindre unité si et seulement si toutes les moyennes mobiles latentes $(X_{j,t})$ sont anticipatives au sens de (1.16). On dira ici par analogie que l'agrégat stable (X_t) est lui-même anticipatif.

Une fois achevée l'étude préliminaire de la représentabilité des chemins de processus α-stable, l'analyse de la distribution conditionnelle peut commencer. Le choix du borélien $B \subset C^{\|\cdot\|}_{m+h+1}$ apparaissant dans le conditionnement de (1.11) doit être adapté au contexte de la prévision, i.e., l'évènement $\{\boldsymbol{X}_t/\|\boldsymbol{X}_t\| \in B\}$ doit être indépendant des h réalisations inobservées. Gardant à l'esprit que $\|\boldsymbol{X}_t\| = \|(X_{t-m},\dots,X_t,0,\dots,0)\|$, supposons que l'on observe

$$\frac{(X_{t-m},\dots,X_t)}{\|\boldsymbol{X}_t\|} \in V,$$

pour un certain borélien V sur la sphère unité $S^{\|\cdot\|}_{m+1} = \{(s_{-m},\dots,s_0) \in \mathbb{R}^{m+1} : \ \|(s_{-m},\dots,s_0,0,\dots,0)\| = 1\}$ de $\mathbb{R}^{m+1}$.[8] Le plus grand borélien dans lequel vit le vecteur complet $\boldsymbol{X}_t/\|\boldsymbol{X}_t\| \in C^{\|\cdot\|}_{m+h+1}$ est alors

$$B(V) := V \times \mathbb{R}^h \subset C^{\|\cdot\|}_{m+h+1}, \tag{1.18}$$

et c'est ce type de boréliens que l'on choisira pour apparaître dans le conditionnement de (1.11). Pour (X_t) un agrégat α-stable anticipatif défini par (1.17), on montre que la mesure spectrale des chemins $\boldsymbol{X}_t =$

[8] L'ensemble $S^{\|\cdot\|}_{m+1}$ n'est autre que la sphère unité de $\mathbb{R}^{m+1}$ relative à la restriction de la semi-norme $\|\cdot\|$ aux $m+1$ premières composantes.

$(X_{t-m},\dots,X_t,X_{t+1},\dots,X_{t+h})$ sur $C^{\|\cdot\|}_{m+h+1}$ s'écrit

$$\Gamma^{\|\cdot\|} = \sum_{j=1}^{J}\sum_{\vartheta=\pm1}\sum_{k\in\mathbb{Z}} w_{j,\vartheta}\pi_j^{\alpha}\|\boldsymbol{d}_{j,k}\|^{\alpha}\delta_{\left\{\frac{\vartheta\boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|}\right\}},$$

avec $\delta_{\{.\}}$ masse de Dirac, $\boldsymbol{d}_{j,k} := (d_{j,k+m},\dots,d_{j,k},d_{j,k-1},\dots,d_{j,k-h})$, $w_{j,\boldsymbol{\vartheta}} := (1+\vartheta\beta_j)/2$, et si $\boldsymbol{d}_{j,k} = (0,\dots,0)$, le terme disparaît de la somme par convention. La propriété (1.11) donne alors

$$\mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t},A\Big|B(V)\Big) := \mathbb{P}\bigg(\frac{\boldsymbol{X}_t}{\|\boldsymbol{X}_t\|}\in A\bigg|\|\boldsymbol{X}_t\|>x,\frac{\boldsymbol{X}_t}{\|\boldsymbol{X}_t\|}\in B(V)\bigg) \underset{x\to+\infty}{\longrightarrow} \frac{\Gamma^{\|\cdot\|}\Big(A\cap B(V)\Big)}{\Gamma^{\|\cdot\|}\Big(B(V)\Big)}, \tag{1.19}$$

pour tous boréliens $A\subset C^{\|\cdot\|}_{m+h+1}$ et «observation» V telle que $\Gamma^{\|\cdot\|}\Big(B(V)\Big)>0$ (et satisfaisant certaines conditions de continuité).

On peut maintenant explorer la distribution conditionnelle lors des évènements extrêmes en évaluant la probabilité asymptotique en différentes localisations de l'espace et pour différents conditionnements. On tentera ici de dégager les lignes d'analyse principales en présentant côte-à-côte le cadre général et l'exemple de l'agrégation d'AR(1) anticipatifs.

On remarque tout d'abord que la mesure $\Gamma^{\|\cdot\|}$ charge uniquement les points de la forme $\pm\boldsymbol{d}_{j,k}/\|\boldsymbol{d}_{j,k}\|$ pour $k\in\mathbb{Z}$ et $j=1,\dots,J$. Ces points correspondent eux-mêmes à des «chemins» déterministes extraits des suites de coefficients des moyennes mobiles latentes $(X_{j,t})$. En choisissant A un voisinage arbitrairement petit autour de tous les points $(\pm\boldsymbol{d}_{j,k}/\|\boldsymbol{d}_{j,k}\|)_{j,k}$, on constate que la probabilité «inconditionnelle»[9] asymptotique vaut

$$\lim_{x\to+\infty}\mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t},A\Big|C^{\|\cdot\|}_{m+h+1}\Big)=1.$$

Par conséquent, lors de tout évènement extrême, le chemin suivi par le processus (X_t) est nécessairement «de la même forme» qu'un des chemins déterministes $\pm\boldsymbol{d}_{j,k}/\|\boldsymbol{d}_{j,k}\|$.

Pour l'AR(1) agrégé (1.9), les suites de coefficients des moyennes mobiles sont de la forme $(\rho_j^k\mathbb{1}_{\{k\geq0\}})_k$, ce qui donne pour tout $j=1,\dots,J$

$$\frac{\boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|} = \begin{cases} \dfrac{(\overbrace{\rho_j^{k+m},\dots,\rho_j,1,0,\dots,0}^{m+1},\overbrace{0,\dots,0}^{h})}{\|(\rho_j^{k+m},\dots,\rho_j,1,0,\dots,0,0,\dots,0)\|}, & \text{pour} \qquad k\in\{-m,\dots,-1\},\\[2em] \dfrac{(\overbrace{\rho_j^{k+m},\dots,\rho_j^{k}}^{m+1},\overbrace{\rho_j^{k-1},\dots,\rho_j,1,0,\dots,0}^{h})}{\|(\rho_j^{k+m},\dots,\rho_j^{k},\rho_j^{k-1},\dots,\rho_j,1,0,\dots,0)\|}, & \text{pour} \qquad k\in\{0,\dots,h-1\},\\[2em] \dfrac{(\overbrace{\rho_j^{m+h},\dots,\rho_j^{h}}^{m+1},\overbrace{\rho_j^{h-1},\dots,\rho_j,1}^{h})}{\|(\rho_j^{m+h},\dots,\rho_j^{h},\rho_j^{h-1},\dots,\rho_j,1)\|}, & \text{pour} \qquad k\geq h,\end{cases}$$

[9] Conditionnelle au borélien non-informatif $B(S^{\|\cdot\|}_{m+1}) = C^{\|\cdot\|}_{m+h+1}$ et au fait d'observer un évènement extrême (i.e., $\|\boldsymbol{X}_t\|>x$).

Les chemins déterministes ci-dessus consistent en des fragments d'exponentielles divergentes, arrêtées à une certaine coordonnée (pour $-m \leq k \leq h-1$), puis demeurant à zéro.[10] Ainsi, lors des évènements extrêmes, la trajectoire de l'AR(1) anticipatif agrégé est nécessairement de la forme d'une exponentielle explosive avec un certain taux d'acroissement ρ_j^{-1}, et s'achevant éventuellement sur un retour à des niveaux non-extrêmes. On formalise ainsi le comportement que Gouriéroux et Zakoïan (2017) [63] avait intuitivement conjecturé, à savoir l'apparition de bulles aux taux d'accroissement différents et l'on peut obtenir les probabilités d'occurrence de chaque type de bulle explicitement. En particulier, on constate que pour $J = 1$, c'est-à-dire pour l'AR(1) anticipatif non-agrégé, un seul taux d'accroissement est possible et le processus ne génère qu'un seul type de bulle. On s'aperçoit ainsi que les processus non-agrégés ont des dynamiques limitées, astreintes à un seul motif apparaissant de façon récurrente d'un évènement extrême à l'autre. Les processus agrégés ne sont pas soumis à une telle restriction et génèrent des trajectoires où peuvent apparaître différents «*patterns*» au cours du temps.

Plus encore que les probabilités inconditionnelles d'occurrence, on peut aussi évaluer la probabilité que le processus suivent certains chemins étant observée une parcelle de trajectoire formée de $m+1$ observations appartenant à un voisinage V sur $S_{m+1}^{\|\cdot\|}$. La condition $\Gamma^{\|\cdot\|}\Big(B(V)\Big) > 0$ indique que l'on ne peut conditionner que par un borélien V contenant au moins une parcelle de trajectoire «plausible» du processus (X_t). Pour un agrégat stable anticipatif général, ces trajectoires «plausibles», ou «observables», de longueur $m+1$ sont nécessairement de la forme

$$\frac{(d_{j,k+m}, \dots, d_{j,k})}{\|\boldsymbol{d}_{j,k}\|}.$$

Dans le cas de l'AR(1) agrégé, on montre que

$$\frac{(d_{j,k+m}, \dots, d_{j,k})}{\|\boldsymbol{d}_{j,k}\|} = \begin{cases} \dfrac{(\overbrace{\rho_j^{k+m}, \dots, \rho_j, 1, 0, \dots, 0}^{m+1})}{\|(\underbrace{\rho_j^{k+m}, \dots, \rho_j, 1, 0, \dots, 0}_{m+1}, \underbrace{0, \dots, 0}_{h})\|}, & \text{for} \qquad k \in \{-m, \dots, -1\}, \\[2em] \dfrac{(\overbrace{\rho_j^{m}, \dots, \rho_j, 1}^{m+1})}{\|(\underbrace{\rho_j^{m}, \dots, \rho_j, 1}_{m+1}, \underbrace{0, \dots, 0, 0, \dots, 0}_{h})\|}, & \text{pour} \qquad k \geq 0. \end{cases}$$

Les trajectoires «observables» consistent ou bien en une tendance exponentielle ininterrompue, ou bien en une tendance exponentielle suivie d'un retour à zéro. Il est aisé, intuitivement, d'identifier l'observation d'une tendance exponentielle ininterrompue avec la période d'«inflation» d'une bulle, c'est-à-dire, pour un certain $j_0 \in \{1, \dots, J\}$, un évènement de la forme

$$\frac{(X_{t-m}, \dots, X_t)}{\|\boldsymbol{X}_t\|} \in V, \quad \text{pour} \quad V \quad \text{un voisinage de} \quad \frac{(\rho_{j_0}^m, \dots, \rho_{j_0}, 1)}{\|(\rho_{j_0}^m, \dots, \rho_{j_0}, 1, 0, \dots, 0)\|}.$$

Conditionnellement à une telle observation, quelles sont les trajectoires futures possibles et leurs probabilités d'occurrence ? Pour répondre à cette question il faut en premier lieu identifier les éléments de $V \times \mathbb{R}^h$

[10] Pour $k < -m$, $\boldsymbol{d}_{j,k} = (0, \dots, 0)$ et ces indices k n'interviennent donc pas dans la mesure spectrale $\Gamma^{\|\cdot\|}$.

auxquels $\Gamma^{\|\cdot\|}$ attribuent une masse strictement positive, à savoir

$$B(V)^+ := \Big\{ \boldsymbol{s} \in V \times \mathbb{R}^h : \quad \Gamma^{\|\cdot\|}(\{\boldsymbol{s}\}) > 0 \Big\}.$$

On montre que

$$B(V)^+ = \left\{ \frac{\boldsymbol{d}_{j_0,k}}{\|\boldsymbol{d}_{j_0,k}\|} : \quad k \geq 0 \right\},$$

et on en déduit à l'aide de (1.19) qu'avec probabilité 1, asymptotiquement, le chemin complet $\boldsymbol{X}_t/\|\boldsymbol{X}_t\|$ appartiendra à un voisinage arbitrairement petit des points $(\boldsymbol{d}_{j_0,k}/\|\boldsymbol{d}_{j_0,k}\|)_{k\geq 0}$, comme données plus haut. Ainsi, le chemin futur est nécessairement une tendance exponentielle de même taux d'accroissement que la trajectoire observée, $\rho_{j_0}^{-1}$, et dont la date d'arrêt peut être atteinte à n'importe quelle date future. Si la date de «crash» de la bulle demeure incertaine, on peut néanmoins évaluer sa probabilité d'occurrence à tout horizon. En considérant A un voisinage arbitrairement petit de

$$\frac{(\overbrace{\rho_{j_0}^{k+m},\ldots,\rho_{j_0}^{k}}^{m+1},\overbrace{\rho_{j_0}^{k-1},\ldots,\rho_{j_0},1,0,\ldots,0}^{h})}{\|(\rho_{j_0}^{k+m},\ldots,\rho_{j_0}^{k},\rho_{j_0}^{k-1},\ldots,\rho_{j_0},1,0,\ldots,0)\|}, \qquad \text{pour un certain} \qquad k \in \{0,\ldots,h-1\},$$

l'évènement sur le vecteur complet $\{\boldsymbol{X}_t/\|\boldsymbol{X}_t\| \in A\}$ correspond à celui pour lequel le sommet de la bulle est atteint dans exactement k périodes, $0 \leq k < h$. On montre pour ce choix de A que

$$\mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t}, A \Big| B(V)\Big) \underset{x\to+\infty}{\longrightarrow} \frac{\Gamma^{\|\cdot\|}\Big(A \cap B(V)\Big)}{\Gamma^{\|\cdot\|}\Big(B(V)\Big)} = \frac{\Gamma^{\|\cdot\|}\Big(A \cap B(V)^+\Big)}{\Gamma^{\|\cdot\|}\Big(B(V)^+\Big)} = |\rho_{j_0}|^{\alpha k}(1-|\rho_{j_0}|^{\alpha}).$$

D'autre part, pour A un voisinage arbitrairement petit de

$$\frac{(\overbrace{\rho_{j_0}^{m+h},\ldots,\rho_{j_0}^{h}}^{m+1},\overbrace{\rho_{j_0}^{h-1},\ldots,\rho_{j_0},1}^{h})}{\|(\rho_{j_0}^{m+h},\ldots,\rho_{j_0}^{h},\rho_{j_0}^{h-1},\ldots,\rho_{j},1)\|},$$

l'évènement $\{\boldsymbol{X}_t/\|\boldsymbol{X}_t\| \in A\}$ correspond à celui pour lequel le sommet sera atteint dans h périodes ou plus. On obtient

$$\mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t}, A \Big| B(V)\Big) \underset{x\to+\infty}{\longrightarrow} \frac{\Gamma^{\|\cdot\|}\Big(A \cap B(V)^+\Big)}{\Gamma^{\|\cdot\|}\Big(B(V)^+\Big)} = |\rho_{j_0}|^{\alpha h}.$$

Ainsi, étant observée la phase d'inflation d'une bulle (disons de taux d'accroissement $\rho_{j_0}^{-1}$), la probabilité que le sommet soit atteint dans exactement k périodes, $0 \leq k < h$ est $|\rho_{j_0}|^{\alpha k}(1-|\rho_{j_0}|^{\alpha})$, tandis que la probabilité que le sommet soit atteint dans h périodes ou plus est $|\rho_{j_0}|^{\alpha h}$. L'interprétation des moments conditionnels du troisième chapitre est retrouvée avec plus de flexibilité.[11]

[11] Au troisième chapitre, la forme des moments conditionnels de X_{t+h} sachant X_t laissait voir l'apparition d'une distribution de type Bernoulli lors des évènements extrêmes : soit la bulle aura éclaté strictement avant la période h, soit la bulle survivra h périodes ou plus. Les probabilités respectives de ces évènements étaient $1-|\rho|^{\alpha h}$ et $|\rho|^{\alpha h}$, qui sont bien compatibles avec la distribution obtenue au quatrième chapitre.

La démarche suivie ici illustre l'utilisation de la propriété (1.19) pour l'agrégation d'AR(1) stables anticipatifs mais peut être déployée pour étudier tout type de processus stable, pourvu qu'il soit anticipatif. On voit émerger une interprétation de la prédiction des chemins futurs en termes de reconnaissance de formes («*pattern identification*») dans les trajectoires. D'autres processus sont considérés dans ce quatrième chapitre, et en particulier un processus bivarié, afin de mettre en exergue le potentiel du cadre d'analyse proposé.

1.2 Introduction in English

1.2.1 Background

After gathering a set of measurements through time from a given phenomenon, an observer can undertake the quantitative study of its evolution. In some cases, such studies lead to quasi-mechanistic explanations from which the future evolution of the phenomenon given any observed state can be deduced almost exactly. In other cases however, even in the presence of some identifiable typical behaviours through time, the evolution seems to feature a significant random component which origin remains unexplained. In the probabilistic approach of erratic time-evolving phenomena, one does not seek a mechanistic scheme anymore but rather a «well approximating» stochastic process. Instead of directly exhibiting the causality relations between measurements and variables of interest, this approximation shall only replicate «as best as possible» the time dynamics of the phenomenon in order to provide accurate predictions of future evolutions. Under one of its most general setting, this approach assumes that the evolution of measurements, say (X_t), follows a dynamics of the form

$$X_t = \varphi(X_{t-1}, X_{t-2}, \ldots; \varepsilon_t),$$

where φ is a function describing the dependence of the present observation to the past evolution and (ε_t) is an independent and identically distributed (i.i.d.) error sequence. The historical starting point of time series analysis lies in the case where φ is assumed linear, which has given rise to a massive theoretical and applied literature. Such linear models, called *autoregressive* (AR), were first used by Yule (1927) [130] to study the time series of sunspot counts spanning the period from 1749 to 1924 that was published by Alfred Wolfer in 1925 [126].[12] The theoretical foundations of autoregressive processes and linear time series analysis in general have been the object of investigations shortly afterwards (see for instance Mann and Wald (1943) [98], Kendall (1944) [82], Bartlett (1946) [6]). Linear time series modelling has been fruitfully applied to domains such as, *inter alia*, economics, telecommunications, geophysics and astronomy and their usefulness stretches until today. Their diffusion have recently greatly benefited from the publications of methodological books such as those of Box and Jenkins (1970) [14] and Brockwell and Davis (1991) [19], and of the computer automatisation of several analysis procedures. In the framework of AR processes, one attempts to approximate the dynamics of some time series data by a process satisfying a linear stochastic recurrence equation of the form

$$\varphi(B)X_t = \varepsilon_t, \tag{1.20}$$

for a polynomial $\varphi(z) := 1 + \sum_{i=1}^{p} \varphi_i z^i$, $z \in \mathbb{C}$, of degree $p \in \mathbb{N}$ with real coefficients, and where B is the lag operator ($BX_t := X_{t-1}$) and (ε_t) an i.i.d. sequence. Stochastic recurrence equations such as (1.20) admit a

[12] Now a textbook example of time series analysis, this series is often named *Wolfer's sunspot series* despite the fact that it was Johann Rudolph Wolf who initiated and mainly conducted the gathering of sunspot counts starting 1848 [125]. Wolfer later became Wolf's assistant, in 1875 (see Izenman (1983) [76] and the references therein).

unique strictly stationary solution if and only if the polynomial φ has no root on the complex unit circle:

$$\varphi(z) \neq 0, \qquad \text{for} \quad |z| = 1. \tag{1.21}$$

At any date $t \in \mathbb{Z}$, the random variable X_t depends in general on all the terms of the i.i.d. sequence (ε_t). The sought-after approximating process is however often assumed to satisfy (1.20) with a polynomial φ that does not have roots neither on nor inside the unit circle:

$$\varphi(z) \neq 0, \qquad \text{for} \quad |z| \leq 1, \tag{1.22}$$

which is necessary and sufficient to guarantee that the stationary solution (X_t) only depends on «past» values $\{\varepsilon_s : \ s \leq t\}$ of the i.i.d. process at any date $t \in \mathbb{Z}$. The solution, called *non-anticipative*, *causal* or *minimum phase*, would then be consistent with a certain intuition of causality when t represents a time-dimension: the present of the process would be only determined by past events. When (1.22) is relaxed, *anticipative* solutions arise, also called *noncausal* or *nonminimum phase*, which depend on «future» values of the sequence (ε_t) and, because of this, are regarded as «unnatural». The anticipative stationary solutions have been generally sidelined when it comes to time series analysis. Nevertheless, AR processes for which (1.22) is not imposed have been considered in other settings. Instead of focusing on the stationary solution of (1.20), a statistical literature has studied processes starting from initial conditions at $t = 0$ and then following a dynamics given by (1.20). When φ has roots inside the unit circle, this type of processes features non-stationary explosive trajectories. The estimation of φ in this context has been the object of several articles (Rubin (1950) [114], White (1958) [123], Anderson (1959) [2], Rao (1961) [108], Stigum (1974) [120], Lai and Wei (1983) [84], Breton and Pham (1989) [17]). Furthermore, when t does not represent a time-dimension but is instead indexing a spatial or a frequency domain, the «causal» and «noncausal» interpretations of the anticipative stationary solutions are immaterial. In such frameworks, nonminimum phase AR processes –only viewed as «filters»– have been at the center of deconvolution problems in geophysics, speech signal processing, telecommunications and astronomy (e.g., Wiggins (1978)[124], Benveniste, Goursat and Ruget (1980) [11], Donoho (1981) [45], Scargle (1981) [119], Godfrey and Rocca (1981)[58], Lii and Rosenblatt (1982, 1988, 1992, 1996) [90, 91, 92, 93], Giannakis and Mendel (1989) [56], Giannakis and Swami (1990) [57], Gassiat (1990a, 1990b, 1993) [53, 54, 55], Breidt, Davis, Lii and Rosenblatt (1991) [16], Chi and Kung (1995) [29], Cheng, Yang and Chi (1997) [28], Andrews, Davis and Breidt (2007)[5]).

Recently however estimation methods that do not impose the causality constraint have been applied to time series of stock prices, trading volumes, unemployment and inflation, and have favoured noncausal linear models over their causal counterparts (Huang and Pawitan (2000) [75], Breidt, Davis and Trindade (2001) [15], Andrews, Calder and Davis (2009) [3], Wu and Davis (2010) [129], Wu (2011) [127], Lanne and Saikkonen (2011) [87], Lanne, Luoto and Saikkonen (2012) [85]). At the intersection of anticipative processes and extreme value theory, it has been noticed in addition that anticipative stationary solutions of (1.20) displayed time dynamics similar to that of speculative bubbles on financial markets, viewed as short-term explosive deviations of prices from a stationary level. The modelling of such phenomena by anticipative processes has been found appropriate for time series such as the Bitcoin/USD rate, oil prices, the Nasdaq index and several series of realised volatilities (Hencic and Gouriéroux (2015) [70], Hecq, Lieb and Telg (2016) [67], Gouriéroux and Zakoïan (2017) [63], Cavaliere, Nielsen and Rahbek (2018) [24]). An econometric and financial literature using noncausal AR processes as its main tool for the analysis of such time series phenomena has thus emerged.

1.2.2 Motivation

For the purpose of speculative bubble modelling, Gouriéroux and Zakoïan (2017) [63] have proposed and studied the anticipative α-stable autoreggression of order 1. This infinite variance strictly stationary process generates trajectories featuring calm periods –close to central values– interspersed by explosive growth episodes (the «bubble inflation») ending on sharp returns to central values («the crash»). The stable anticipative AR(1), arguably one of the most elementary processes within the anticipative family, is defined as the stationary solution of the equation

$$X_t = \rho X_{t+1} + \varepsilon_t, \qquad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, \beta, \sigma, 0), \tag{1.23}$$

where $0 < |\rho| < 1$ and $\mathcal{S}(\alpha, \beta, \sigma, 0)$ stands for the α-stable distribution of tail index $\alpha \in (0, 2)$, asymmetry $\beta \in [-1, 1]$ and scale $\sigma > 0$. Beyond the purely mathematical interest, the study of the stable anticipative AR(1) is motivated by the possibility it offers to infer the peak and crash dates of bubbles, which could be valuable not only for portfolio managers, but also for risk managers and regulators. This nonetheless requires to know its conditional dynamics which, contrary to its non-anticipative counterpart, is non-trivial to obtain because of the dependence of the observed past trajectory $\{X_s : s \leq t\}$ on «future» errors $\{\varepsilon_s : s \geq t\}$. This probabilistic aspect received the least attention from the literature which is explained by the spatial/frequency perspective of the deconvolution problems that motivated the research on noncausal processes across the 1980s and 1990s. The duality between an observed past and a to-be-predicted future is absent from such contexts where the whole domain is potentially observable. Prediction aspects of noncausal processes were nevertheless addressed by Rosenblatt (1995, 2000) [111, 112] in the finite variance framework. Gouriéroux and Zakoïan (2017) [63] have obtained surprising results about the conditional dynamics of the α-stable anticipative AR(1) process. For instance, after establishing its markovian nature, they have shown the existence of conditional moments of higher order than marginal moments:

$$\mathbb{E}[|X_{t+1}|^{\gamma} | X_t] < +\infty, \quad \text{for} \quad \gamma < 2\alpha + 1,$$

despite the fact that $\mathbb{E}[|X_t|^{\alpha}] = +\infty$, because X_t is marginally α-stable.[13] In the case where the sequence (ε_t) is symmetric α-stable ($\beta = 0$), they have furthermore shown that the conditional expectation writes

$$\mathbb{E}[X_{t+1} | X_t] = \rho^{<\alpha - 1>} X_t, \tag{1.24}$$

[13] A similar result has been obtained by Cambanis and Fakhre-Zakeri (1995) [20] for the non-anticipative α-stable AR(1) in reversed time.

where $x^{<b>} := \text{sign}(x)|x|^b$ for $x, b \in \mathbb{R}$, and if (ε_t) is marginally Cauchy-distributed, ($\alpha = 1$, $\beta = 0$), the conditional variance is then a quadratic function of the observed present value:

$$\mathbb{V}(X_{t+1}|X_t) = \left(\frac{1}{|\rho|} - 1\right) X_t^2 + \frac{\sigma^2}{|\rho|(1-|\rho|)}.$$

These results contrast sharply with properties known for non-anticipative processes, either in finite or infinite variance. In particular, the discrepancy between marginal and conditional moments suggests a certain «excess of predictability» of anticipative process compared to non-anticipative ones. When undertaking this thesis, the conditional dynamics of the anticipative α-stable AR(1) is only partially understood, while that of higher-order processes is completely unexplored. In addition to this probabilistic aspect, a statistical dimension also arises as the estimation of the model parameters given a series of observations constitutes an unavoidable step before any practical use of this class of processes.

The purpose of this thesis is to study the class of linear anticipative α-stable processes in the general case, with special attention given to their conditional dynamics.

1.2.3 Main results

The second chapter starts with the study of probabilistic and statistical aspects of higher-order AR processes for which the characteristic polynomial admits roots both inside and outside the unit circle. For these processes, called *mixed* causal/noncausal, the dynamics can be written under the factorised form

$$\psi(F)\phi(B)X_t = \varepsilon_t,$$

for two polynomials ψ and ϕ of respective orders p and q satisfying (1.22), and where F is the lead operator ($F := B^{-1}$, i.e., $FX_t = X_{t+1}$). The extension of the probabilistic properties is studied on the basis of Gouriéroux and Zakoïan's results (2017) [63] on the one hand, and on Lanne and Saikkonen's (2011) [87] decomposition of the mixed process (X_t) into its purely causal and noncausal components on the other hand (see also Gouriéroux and Jasiak (2016) [59]). It is shown that (X_t) is markovian of order $p+q$ and that the discrepancy between marginal and conditional moments is present only if $p \geq 1$, that is, when the noncausal component of the process is non-trivial. Closed formulae are obtained for the conditional moments in particular cases. For instance when $\psi(F) = 1 - \psi F$ and $\beta = 0$, the conditional expectation is linear and writes for all $q \geq 0$

$$\mathbb{E}[X_t|\mathscr{F}_{t-1}] = \psi^{<\alpha-1>} X_{t-1} + (1 - \psi^{<\alpha-1>} B)(\phi_1 X_{t-1} + \ldots + \phi_q X_{t-q}),$$

where $\phi(z) = 1 - \phi_1 z - \ldots - \phi_q z^q$ and $(\mathscr{F}_t)$ denotes the canonical filtration of the process (X_t). If in addition $\alpha = 1$, it is shown that (X_t) admit the semi-strong causal representation

$$(1 - \text{sign}(\psi)B)\phi(B)X_t = \sigma_t \eta_t,$$

$$\sigma_t^2 = \left(\frac{1}{|\psi|} - 1\right)(X_{t-1} - \phi_1 X_{t-2} - \ldots - \phi_q X_{t-q-1})^2 + \frac{\sigma^2}{|\psi|(1-|\psi|)},$$

where $\mathbb{E}\left[\eta_t \middle| \mathscr{F}_{t-1}\right] = 0$ and $\mathbb{E}\left[\eta_t^2 \middle| \mathscr{F}_{t-1}\right] = 1$.

On statistical aspects, this chapter addresses the least squares estimation of the model

$$\psi_0(F)\phi_0(B)X_t = \varepsilon_t,$$

from observations $X_1, \ldots, X_n$, where $\psi_0(z) = 1 - \sum_{i=0}^{p} \psi_{0i} z^i$, $\phi_0(z) = 1 - \sum_{i=0}^{q} \phi_{0i} z^i$. Contrary to the maximum likelihood method, the least squares approach does not require a fully parametric distributional assumption on the error sequence (ε_t), which makes it more robust to misspecifications. Instead of requiring the ε_t's to be α-stable distributed, it is only assumed that they belong to the domain of attraction of an α-stable distribution, i.e.,

$$\mathbb{P}(|\varepsilon_0| > x) = x^{-\alpha} L(x), \qquad \text{and} \qquad \lim_{x\to\infty} \frac{\mathbb{P}(\varepsilon_0 > x)}{\mathbb{P}(|\varepsilon_0| > x)} \to c \in [0,1],$$

for a slowly varying function L at infinity.[14] The statistician faces several unknowns: the orders p and q of the polynomials ψ_0 and ϕ_0, their coefficients, and the causal/noncausal decomposition of the dynamics, that is, which of the two polynomials is associated to the lead operator F, and which is associated to the operator B. One of the difficulty encountered for the estimation is that the process (X_t) admits several polynomial representations with dependent yet non-autocorrelated errors. Indeed, for any polynomial η_0^* obtained from $\psi_0(z)\phi_0(z)$ by replacing one or several roots by their reciprocals, we have that

$$\eta_0^*(B)X_t = \zeta_t^*,$$

where ζ_t^* is a so-called «all-pass» process which empirical autocorrelations tend to zero with the sample size. It is thus expected that one will not be able to retrieve the causal/noncausal structure of the dynamics from the least squares alone given that they only exploit the information from second-order moments. A procedure is hence proposed to alleviate this limitation of the least squares by coupling them with the extreme clustering phenomenon. One can actually take advantage of this limitation by noticing that it yields in particular

$$\psi_0(B)\phi_0(B)X_t = \zeta_t, \tag{1.25}$$

for a certain all-pass process (ζ_t). The procedure consists of three steps.

1) Relying on (1.25) and assuming the order $p+q$ of the polynomial $\eta_0(z) := \psi_0(z)\phi_0(z)$ to be known, it is shown using techniques from Davis and Resnick (1986) [41] that the Yule-Walker estimator of the regression of X_t on $X_{t-1}, \ldots, X_{t-p-q}$ converges in probability and distribution towards the coefficients of η_0.

2) In practice, the degree $p+q$ being unknown, a portmanteau-type tests is introduced to detect any «empirical autocorrelation» in the residuals after estimation. One is thus able to reject under-specified models (of order lower than $p+q$). Starting from low orders, one can estimate the model, test its validity, and increment the order until the validity hypothesis is not rejected.

[14] $\lim_{x\to\infty} L(tx)/L(x) = 1, \forall t > 0$.

3) Having validated the order $p+q$ and estimated the polynomial $\eta_0(z)$, a consistent estimator of the roots of $\psi_0(z)\phi_0(z)$ is available and it is sufficient to identify which roots «belong» to $\psi_0(z)$ and which «belong» to $\phi_0(z)$ in order to recover the causal/noncausal structure of the dynamics. Any allocation of the roots to the causal and noncausal components leads to an all-pass representation of the process (X_t) and it is therefore impossible to determine whether the «true» one has been found on the basis of a residuals autocorrelation test. If every all-pass representations admit non-autocorrelated errors, only those of the original representation are actually independent. In this infinite variance framework, it is justified based on point processes arguments that the dependence of all-pass representations errors gives rise to a phenomenon of extreme clustering: the extreme errors occur in clusters, and the stronger the dependence, the larger the clusters. This phenomenon is absent from the errors of the original representation of the process, because of their independence, and extreme values appear isolated from one another. The third step hence consists in measuring the intensity of extreme clustering in the residuals of each possible allocation of the roots between causal and noncausal components. The allocation corresponding to the «true» causal/noncausal structure is, in principle, the only one for which no evidence of extreme clustering should be found.
The empirical validity of each step is illustrated by simulations and the chapter concludes on an application to six financial time series.

The narrowness of the conditions under which the conditional dynamics has been obtained in the second chapter calls for a new approach. The third chapter revisits the stable anticipative AR(1) from the perspective of multivariate stable random vectors, and leverages in particular the literature on the conditional moments of bivariate stable vectors from the 90s (Hardin, Samorodnitsky, Taqqu (1991) [64], Cioszek-Georges, Taqqu (1994, 1995a, 1995b, 1998) [30, 31, 32, 33]). Any stable vector $(X_1,\dots,X_d)$ is characterised by a unique pair $(\Gamma,\boldsymbol{\mu}^0)$, where Γ is a finite Borel measure on the Euclidean unit sphere $S_d=\{\boldsymbol{s}\in\mathbb{R}^d:\ \|\boldsymbol{s}\|_e=1\}$, $\|\cdot\|_e$ denoting the Euclidean norm, $\boldsymbol{\mu}^0\in\mathbb{R}^d$ a fixed vector, such that

$$\mathbb{E}\Big[e^{i\langle\boldsymbol{u},\boldsymbol{X}\rangle}\Big]=\exp\bigg\{-\int_{S_d}|\langle\boldsymbol{u},\boldsymbol{s}\rangle|^\alpha\Big(1-i\,\mathrm{sign}(\langle\boldsymbol{u},\boldsymbol{s}\rangle)w(\alpha,\langle\boldsymbol{u},\boldsymbol{s}\rangle)\Big)\Gamma(d\boldsymbol{s})+i\,\langle\boldsymbol{u},\boldsymbol{\mu}^0\rangle\bigg\},\qquad\forall\boldsymbol{u}\in\mathbb{R}^d,\tag{1.26}$$

with $w(\alpha,s)=\mathrm{tg}\left(\frac{\pi\alpha}{2}\right)$, if $\alpha\neq1$, and $w(1,s)=-\frac{2}{\pi}\ln|s|$, $s\in\mathbb{R}$ (Theorem 2.3.1 by Samorodnitsky, Taqqu, 1994 [117]). The measure Γ is called the *spectral measure* of $\boldsymbol{X}$ and encodes the information regarding the dependence between the components of the vector, and the pair $(\Gamma,\boldsymbol{\mu}^0)$ is called its *spectral representation.* It is shown that for (X_t) solution of (1.23), the vector (X_t,X_{t+h}) is bivariate α-stable and that its characteristic function is of the form (1.26) with spectral measure Γ_h given by

$$\Gamma_h=\frac{\bar\sigma^\alpha}{2}\sum_{\vartheta\in S_1}\left[\left(1-|\rho|^{\alpha h}+\left(1-\left(\rho^{<\alpha>}\right)^h\right)\vartheta\bar\beta\right)\delta_{\{(\vartheta,0)\}}+\left(1+|\rho|^{2h}\right)^{\alpha/2}(1+\vartheta\bar\beta)\delta_{\{\vartheta\boldsymbol{s}_h\}}\right],\tag{1.27}$$

where $S_1=\{-1,+1\}$, $\delta_{\{x\}}$ is the Dirac mass at point $x\in\mathbb{R}^2$, $\bar\sigma^\alpha=\dfrac{\sigma^\alpha}{1-|\rho|^\alpha}$, $\bar\beta=\beta\dfrac{1-|\rho|^\alpha}{1-\rho^{<\alpha>}}$, and $\boldsymbol{s}_h=\dfrac{(\rho^h,1)}{\sqrt{1+|\rho|^{2h}}}\in S_2$. Any finite distribution of mass on the unit sphere defines a proper spectral measure of a certain stable vector and it is remarkable that, here, Γ_h has a very specific form as a purely discrete object charging either two or four atoms.

From this representation, several results from the literature on bivariate stable vectors can be readily applied. The existence of conditional moments of X_{t+h} given X_t up to order $2\alpha+1$ is obtained from the fact that $\int_{S_2} |s_1|^{-\nu}\Gamma_h(d\boldsymbol{s}) < +\infty$ for any $\nu \geq 0$ and $h \geq 1$ (see Theorem 5.1.3 by Samorodnitsky, 1994, [117] and the articles of Cioszek-George and Taqqu). Theorems 5.2.2 and 5.2.3 by Samorodnitsky and Taqqu (1994) [117] provide the functional forms of the conditional expectation in terms of an arbitrary spectral measure and enable to extend Gouriéroux and Zakoïan's (2017) [63] result (1.24) to any $\beta \in [-1, 1]$, revealing the non-linearity of $x \longmapsto \mathbb{E}[X_{t+h}|X_t = x]$ in the general case. The form of the conditional variance has also been studied in an article by Cioszek-Georges and Taqqu (1995) [31]. In presence of asymmetry, the proof in the case $\alpha \neq 1$ is however missing and the case $\alpha = 1$ is not considered. The results for the conditional variance are completed and this chapter also contributes to the literature on bivariate stable random vectors by providing the functional forms of the conditional moments of order 3 and 4 (i.e., skewness and kurtosis). Having obtained the forms of the first four conditional moments for general (bivariate) stable vector, we deduce their forms for the particular vector (X_t, X_{t+h}) for any admissible parameterisation of the process (X_t) by substituting by the spectral measure (1.27). The dependence between the future realisation and the present observation appears complicated in general, as can be noticed in Theorem 3.2.1. The dynamics of the anticipative α-stable AR(1) however drastically simplifies during extreme events for $\rho > 0$. As $sx \to +\infty$, $s = \pm 1$, it is shown that the conditional moments admit the following asymptotic equivalents[15]

$$
\begin{aligned}
\mu(x,h) &\sim (\rho^{-h}x)\rho^{\alpha h}, && \text{if} \quad \alpha \in (0,2),\\
\sigma^2(x,h) &\sim (\rho^{-h}x)^2\rho^{\alpha h}(1-\rho^{\alpha h}), && \text{if} \quad \alpha \in (1/2,2),\\
\gamma_1(x,h) &\longrightarrow s\frac{1-2\rho^{\alpha h}}{\sqrt{\rho^{\alpha h}(1-\rho^{\alpha h})}}, && \text{if} \quad \alpha \in (1,2),\\
\gamma_2(x,h) &\longrightarrow \frac{1}{\rho^{\alpha h}} + \frac{1}{1-\rho^{\alpha h}} - 6, && \text{if} \quad \alpha \in (3/2,2),
\end{aligned}
$$

where $\mu(x,h)$, $\sigma^2(x,h)$, $\gamma_1(x,h)$ and $\gamma_2(x,h)$ denote respectively the conditional expectation, variance, skewness and excess kurtosis of X_{t+h} given $X_t = x$. It can then be noticed that the expressions appearing on the right-hand side correspond to the first four moments of the random variable Z defined by

$$\mathbb{P}(Z = \rho^{-h}x) = \rho^{\alpha h}, \qquad \text{and} \qquad \mathbb{P}(Z = 0) = 1 - \rho^{\alpha h}.$$

An enlightening interpretation of the dynamics emerges from the asymptotic forms of the conditional moments: during bubble episodes, (X_t) appears to follow an exponential trajectory with growth rate ρ^{-1}, and the conditional probability that the bubble lasts at least h more periods would be given by $\rho^{\alpha h}$. A convergence in distribution towards this behaviour during extreme events is shown in the fourth chapter. With the appropriate tools to define stable processes in continuous time –such as stable random measures and stable

[15] If $|\beta| = 1$, one of the tails of the marginal distribution of X_t is not regularly varying and can even be completely absent. In such case, the extreme values are necessarily all of the same sign and the asymptotics is valid only for $x \to +\infty$ ou $x \to -\infty$.

integrals– parallel results are obtained for the anticipative α-stable Ornstein-Uhlenbeck process.

The approach followed in this chapter can be applied to any stable process (Y_t) to analyse the conditional moments of Y_{t+h} given the present observation Y_t, which is illustrated on the aggregation of anticipative AR(1) defined by

$$X_t = \sum_{j=1}^{J} \pi_j X_{j,t}, \qquad X_{j,t} = \rho_j X_{j,t+1} + \varepsilon_{j,t}, \quad \varepsilon_{j,t} \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, \beta_j, 1, 0), \tag{1.28}$$

for positive real numbers π_j and $J \geq 1$. This process has been proposed by Gouriéroux and Zakoïan (2017) [63] to alleviate a limitation of the simple anticipative AR(1), namely, that it appears only able to generate one type of bubble characterised by a single growth rate ρ^{-1}, necessarily identical from one extreme episode to another. The aggregated process (1.28), on the contrary, features trajectories where bubbles of different growth rates ρ_j^{-1}, $j = 1, \dots, J$ appear. It is shown that (X_t, X_{t+h}) is also α-stable in that case and that its spectral measure Γ_h writes

$$\Gamma_h = \sum_{j=1}^{J} \pi_j^{\alpha} \Gamma_{j,h}, \tag{1.29}$$

where the $\Gamma_{j,h}$'s are given *mutatis mutandis* by (1.27) as the spectral measures of the vectors $(X_{j,t}, X_{j,t+h})$ associated to the simple AR(1) processes with parameter ρ_j, $j = 1, \dots, J$ in (1.28). Beyond the existence of conditional moments up to order $2\alpha + 1$, the functional forms of the first four conditional moments, and despite the fact that $Y_{t+h}|Y_t$ does not fully characterise the conditional distribution for such a non-Markov process, (1.29) betrays how similar the dependence structures of «simple» and aggregated processes are. This similarity is leveraged in the next chapter to study aggregated and non-aggregated stable processes alike, the former featuring much richer dynamics than the latter.

The fourth chapter builds on two indications that emerged from the third one. First, the fact that any piece of trajectory $\boldsymbol{X}_t = (X_{t-m}, \dots, X_t, X_{t+1}, \dots, X_{t+h})$, $m \geq 0$, $h \geq 1$ is multivariate α-stable, and second, the fact that the dynamics simplifies when the process is far from its central values. This simplification, noticed on the forms of the conditional moments, can be obtained as a consequence of the regularly varying tails property of the stable vector $\boldsymbol{X}_t$. Denoting Γ the spectral measure of $\boldsymbol{X}_t$ on the Euclidean unit sphere S_{m+h+1}, we have by Theorem 4.4.8 by Samorodnitsky and Taqqu (1994) [117] and the Bayes Theorem that

$$\mathbb{P}\Big(\boldsymbol{X}_t / \|\boldsymbol{X}_t\|_e \in A \;\Big|\; \|\boldsymbol{X}_t\|_e > x \;\text{ and }\; \boldsymbol{X}_t / \|\boldsymbol{X}_t\|_e \in B\Big) \underset{x \longrightarrow +\infty}{\longrightarrow} \frac{\Gamma(A \cap B)}{\Gamma(B)}, \tag{1.30}$$

for any Borel sets A and B satisfying some Γ-continuity conditions. The spectral measure Γ thus completely describes the conditional distribution of normalised paths $\boldsymbol{X}_t / \|\boldsymbol{X}_t\|_e$ –the «shape» of the trajectory– when the vector $\boldsymbol{X}_t$ is large according to the Euclidean norm. In the case where (X_t) is an infinite moving average process of general form

$$X_t = \sum_{k=-\infty}^{+\infty} d_k \varepsilon_{t+k}, \qquad t \in \mathbb{Z}, \tag{1.31}$$

the measure Γ is explicit and it would be tempting to directly invoke (1.30) in order to study the conditional distribution of the future path $(X_{t+1}, \ldots, X_{t+h})$ given the observed past trajectory $(X_{t-m}, \ldots, X_t)$. For prediction purposes however, (1.30) is of little interest because the conditioning event depends on future realisations of the process, mainly through the Euclidean norm of $\boldsymbol{X}_t$. The idea in this chapter is to obtain a version of (1.30) where the Euclidean norm would be replaced by a semi-norm $\|\cdot\|$ satisfying

$$\|(x_{-m}, \ldots, x_0, x_1, \ldots, x_h)\| = \|(x_{-m}, \ldots, x_0, 0, \ldots, 0)\|, \tag{1.32}$$

for any vector $(x_{-m}, \ldots, x_h) \in \mathbb{R}^{m+h+1}$. In this view, a new spectral representation of stable vectors on the unit «sphere» $C_{m+h+1}^{\|\cdot\|} := \{\boldsymbol{s} \in \mathbb{R}^{m+h+1} : \ \|\boldsymbol{s}\| = 1\}$ relative to such semi-norm is explored. For obvious geometrical reasons, the term *cylinder* is preferred to that of «sphere». It will be said that a stable random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ is *representable* on a unit cylinder relative to a semi-norm $\|\cdot\|$ if there exists a pair $(\Gamma^{\|\cdot\|}, \boldsymbol{\mu}^0_{\|\cdot\|})$, $\Gamma^{\|\cdot\|}$ a Borel measure on $C_d^{\|\cdot\|}$, $\boldsymbol{\mu}^0_{\|\cdot\|} \in \mathbb{R}^d$, such that the characteristic function of $\boldsymbol{X}$ is of the form (1.26) with $(S_d, \Gamma, \boldsymbol{\mu}^0)$ replaced by $(C_d^{\|\cdot\|}, \Gamma^{\|\cdot\|}, \boldsymbol{\mu}^0_{\|\cdot\|})$. Provided such a representation exists, it is then shown that (1.30) can indeed be restated with the Euclidean norm replaced by the corresponding semi-norm and Γ by $\Gamma^{\|\cdot\|}$. If all stable vectors admit representations on unit spheres relative to norms (Theorem 2.3.8 by Samorodnitsky and Taqqu (1994) [117]), it is shown here that this not the case when it comes to representations on unit cylinders relative to semi-norms. This chapter hence starts with the preliminary study of the *representability* of general stable vectors, and of paths $\boldsymbol{X}_t$ in particular, on such unit cylinders. As often arises when working with stable distributions, the case $\alpha = 1$ requires special attention in the presence of asymmetry. This case is considered as well in the fourth chapter and the results are similar, however for expository purposes, we will assume in the following that $\alpha \neq 1$. It is shown that for any semi-norm $\|\cdot\|$ on $\mathbb{R}^d$, an arbitrary stable vector $(X_1, \ldots, X_d)$ will admit a representation on $C_d^{\|\cdot\|}$ if and only if[16]

$$\Gamma\Big(\{\boldsymbol{s} \in S_d : \ \|\boldsymbol{s}\| = 0\}\Big) = 0. \tag{1.33}$$

This condition can be understood as follows. Intuitively, in view of (1.30), the spectral measure of a stable vector encodes the information about the tail dependence of its components by assigning mass to the directions of the space.[17] A stable vector is completely characterised by this distribution of mass on the directions of the space, which is usually expressed in terms of a measure on a unit sphere relative to a norm. As unit spheres are in bijection with the set of all directions of the space, measures thereon are able to characterise any potential tail dependence of a stable vector. On the contrary, unit cylinders are not in bijection with the set of all directions of the space, and measures thereon cannot describe the tail dependence of certain stable vectors. Condition (1.33) indicates that a stable vector will be representable on $C_d^{\|\cdot\|}$ provided

[16] In the asymmetric $\alpha = 1$ case, a slightly stronger necessary and sufficient condition holds, namely that $\int_{S_d} \Big|\ln\|\boldsymbol{s}\|\Big|\Gamma(d\boldsymbol{s}) < +\infty$. The latter in particular implies (1.33) given that $\Big|\ln\|\boldsymbol{s}\|\Big| = +\infty$ for any $\boldsymbol{s}$ in the kernel of the semi-norm.

[17] By «direction», it is meant the equivalence class of $\mathbb{R}^d$ for the relation «$\equiv$» defined by: $\boldsymbol{u} \equiv \boldsymbol{v}$ if and only if there exists $\lambda > 0$ such that $\boldsymbol{u} = \lambda\boldsymbol{v}$, $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^d$

the pathological directions –the ones belonging to the kernel of the semi-norm– are uninformative to describe its tail dependence. In dimension 2, a vector (X_1, X_2) with spectral measure Γ on the Euclidean unit sphere will be representable on the cylinder $\{(s_1, s_2) \in \mathbb{R}^2 : \ |s_1| = 1\}$ provided that $\Gamma\Big(\{(0,-1),(0,+1)\}\Big) = 0$, that is, if the vector $(X_1, X_2)/\sqrt{X_1^2 + X_2^2}$ has probability zero of tending towards $(0, \pm 1)$ when its norm $\sqrt{X_1^2 + X_2^2}$ grows infinitely large. In other terms, $\Gamma\Big(\{(0,-1),(0,+1)\}\Big) = 0$ indicates that realisations (X_1, X_2) where X_2 is extreme and X_1 is non-extreme occur with probability zero.

After considering the case of general stable vectors, we focus on paths of a stable moving average (X_t) of general form (1.31) for which we *only* require the summability condition

$$0 < \sum_{k \in \mathbb{Z}} |d_k|^s < +\infty, \qquad \text{for some} \quad s \in (0, \alpha) \cap [0, 1]. \tag{1.34}$$

The moving average is hence not *a priori* assumed anticipative. Property (1.30) will be applicable to study the conditional distribution of (X_t) only if there are paths of the form

$$\boldsymbol{X}_t = (\underbrace{X_{t-m}, \ldots, X_t}_{\substack{m+1 \\ \text{observations}}}, \underbrace{X_{t+1}, \ldots, X_{t+h}}_{\substack{h \\ \text{prediction horizons}}}),$$

which are representable on $C^{\|\cdot\|}_{m+h+1}$ for a semi-norm satisfying (1.32). Several cases arise according to the kernel of the semi-norm. To fix ideas, we consider in this chapter semi-norms such that

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

for any $(x_{-m}, \ldots, x_h) \in \mathbb{R}^{m+h+1}$, which in particular satisfy (1.32).
Letting $\mathcal{M} = \{m \geq 1 : \ \exists k \in \mathbb{Z}, \ d_{k+m} = \ldots = d_{k+1} = 0, \ d_k \neq 0\}$, and

$$m_0 = \begin{cases} \sup \mathcal{M}, & \text{si} \quad \mathcal{M} \neq \emptyset, \\ 0, & \text{si} \quad \mathcal{M} = \emptyset, \end{cases}$$

it is shown using (1.33) that the paths $\boldsymbol{X}_t$ will be representable on $C^{\|\cdot\|}_{m+h+1}$ for all $m \geq m_0$, $h \geq 1$ if and only if

$$m_0 < +\infty, \tag{1.35}$$

Hence, (X_t) will induce paths representable on $C^{\|\cdot\|}_{m+h+1}$ only if any sequence of consecutive zero values in the coefficients (d_k) is either finite, or infinite «to the left». Astonishingly, although no assumption were imposed regarding the causality or noncausality of (X_t), it is thus shown that only anticipative moving averages induce paths $\boldsymbol{X}_t$ representable on unit cylinders that are appropriate for prediction purposes. This representability condition sheds a new light on the predictability of extremes events of an α-stable process. Intuitively, the path $\boldsymbol{X}_t$ will be representable on $C^{\|\cdot\|}_{m+h+1}$ provided any extreme event affecting the h last unobserved components cannot occur independently of an extreme event on the $m+1$ first observed ones (condition (1.33)). In other terms, all incoming extreme event must manifest early visible signs. The difference between anticipative and non-anticipative processes is obvious from this perspective. For the latter (e.g., the non-anticipative stable AR(1) $Y_t = \rho Y_{t-1} + \eta_t$), extreme events erupt suddenly, without warning, under the form of jumps which magnitudes follow an infinite variance power-law. For the former on the contrary, extreme events are reached gradually and are identifiable in advance from trends and patterns that precede and betray their incoming occurrence (the inflation of bubbles in the anticipative AR(1) model). For autoregressive processes, which were the starting point of this thesis, the representability condition of paths simplifies. Consider (X_t) an ARMA process defined as the 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 satisfying (1.22) with ψ and Θ (resp. ϕ and H) sharing no common root. It is then shown that (X_t) will induce paths representable on unit cylinders that are appropriate for prediction purposes if and only if

$$\deg(\psi) \geq 1.$$

In this framework, non-anticipative AR processes are hence typically pathological. A strength of this approach is to naturally encompass processes resulting from the linear combination of α-stable moving averages, coined *stable aggregates* and defined as

$$X_t = \sum_{j=1}^{J} \pi_j X_{j,t}, \qquad X_{j,t} = \sum_{k\in\mathbb{Z}} d_{j,k}\varepsilon_{j,t+k}, \qquad \varepsilon_{j,t} \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, \beta_j, 1, 0), \tag{1.36}$$

where the π_j's are positive real numbers, $J \geq 1$, and where each coefficients sequence $(d_{j,k})_k$ satisfies the summability condition (1.34). Exploiting a relation similar to that of (1.29) obtained in the third chapter, it is established that paths of aggregated processes (X_t) will be representable on unit cylinders provided all the latent moving averages $(X_{j,t})$ are anticipative in the sense of (1.35). By analogy, we will refer here to such stable aggregates as anticipative.

Once conducted the preliminary study of paths representability, the analysis of the conditional distribution can start. The choice of Borel set $B \subset C^{\|\cdot\|}_{m+h+1}$ appearing in the conditioning of (1.30) has to be adapted to the prediction framework, i.e., the event $\{\boldsymbol{X}_t/\|\boldsymbol{X}_t\| \in B\}$ has to be independent of the h unobserved realisations. Keeping in mind that $\|\boldsymbol{X}_t\| = \|(X_{t-m}, \dots, X_t, 0, \dots, 0)\|$, assume that for a certain Borel set V on the unit sphere $S^{\|\cdot\|}_{m+1} = \{(s_{-m}, \dots, s_0) \in \mathbb{R}^{m+1} : \ \|(s_{-m}, \dots, s_0, 0, \dots, 0)\| = 1\}$ of $\mathbb{R}^{m+1}$,[18] the following event is observed:

$$\frac{(X_{t-m}, \dots, X_t)}{\|\boldsymbol{X}_t\|} \in V.$$

The largest Borel set in which the complete vector $\boldsymbol{X}_t/\|\boldsymbol{X}_t\| \in C^{\|\cdot\|}_{m+h+1}$ lives is then

$$B(V) := V \times \mathbb{R}^h \subset C^{\|\cdot\|}_{m+h+1}. \tag{1.37}$$

[18] The set $S^{\|\cdot\|}_{m+1}$ is the unit sphere of $\mathbb{R}^{m+1}$ relative to the restriction of the semi-norm $\|\cdot\|$ to the $m+1$ first components.

This is the type of Borel sets that will be used to appear in the conditioning of (1.30). For an α-stable aggregate (X_t) defined by (1.36), it is shown that the spectral measure of paths $\boldsymbol{X}_t = (X_{t-m}, \ldots, X_t, X_{t+1}, \ldots, X_{t+h})$ on $C^{\|\cdot\|}_{m+h+1}$ writes

$$\Gamma^{\|\cdot\|} = \sum_{j=1}^{J} \sum_{\vartheta=\pm 1} \sum_{k\in\mathbb{Z}} w_{j,\vartheta} \pi_j^{\alpha} \|\boldsymbol{d}_{j,k}\|^{\alpha} \delta_{\left\{ \frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|} \right\}},$$

with $\delta_{\{\cdot\}}$ the Dirac mass, $\boldsymbol{d}_{j,k} := (d_{j,k+m}, \ldots, d_{j,k}, d_{j,k-1}, \ldots, d_{j,k-h})$, $w_{j,\boldsymbol{\vartheta}} := (1+\vartheta\beta_j)/2$, and if $\boldsymbol{d}_{j,k} = (0,\ldots,0)$, the term vanishes by convention from the sum. Property (1.30) then yields

$$\mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t}, A \Big| B(V)\Big) := \mathbb{P}\Big(\frac{\boldsymbol{X}_t}{\|\boldsymbol{X}_t\|} \in A \Big| \|\boldsymbol{X}_t\| > x, \frac{\boldsymbol{X}_t}{\|\boldsymbol{X}_t\|} \in B(V)\Big) \underset{x\to+\infty}{\longrightarrow} \frac{\Gamma^{\|\cdot\|}\Big(A \cap B(V)\Big)}{\Gamma^{\|\cdot\|}\Big(B(V)\Big)}, \tag{1.38}$$

for all Borel sets $A \subset C^{\|\cdot\|}_{m+h+1}$ and «observation» V such that $\Gamma^{\|\cdot\|}\Big(B(V)\Big) > 0$ (and satisfying certain $\Gamma^{\|\cdot\|}$-continuity conditions).

It is now possible to explore the conditional distribution during extreme events by evaluating the asymptotic probability in different regions of the space and for different conditionings. Only the main lines of the analysis will be drawn here, and the general case will be illustrated with the help of the aggregation of anticipative AR(1) with positive coefficients ρ_j's.

One notices first and foremost that the spectral measure $\Gamma^{\|\cdot\|}$ only charges the points of the form $\pm\boldsymbol{d}_{j,k}/\|\boldsymbol{d}_{j,k}\|$ for $k \in \mathbb{Z}$ and $j = 1, \ldots, J$. These points themselves correspond to deterministic «paths» extracted from the coefficients sequences of the latent moving averages $(X_{j,t})$. By setting A to be an arbitrarily small neighbourhood around all the points $(\pm\boldsymbol{d}_{j,k}/\|\boldsymbol{d}_{j,k}\|)_{j,k}$, one can see that the asymptotic «unconditional» probability[19] equals 1:

$$\lim_{x\to+\infty} \mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t}, A \Big| C^{\|\cdot\|}_{m+h+1}\Big) = 1.$$

Therefore, during an extreme event, the path followed by the process (X_t) is necessarily «of the same shape» as one of the deterministic paths $\pm\boldsymbol{d}_{j,k}/\|\boldsymbol{d}_{j,k}\|$.

For the aggregated AR(1) (1.28), the coefficients sequences of the latent moving averages are of the form

[19] Conditional on the uninformative Borel set $B(S^{\|\cdot\|}_{m+1}) = C^{\|\cdot\|}_{m+h+1}$ and on the fact that an extreme event is observed (i.e., $\|\boldsymbol{X}_t\| > x$).

$(\rho_j^k \mathbb{1}_{\{k\geq 0\}})_k$, which yields for all $j=1,\dots,J$ that

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

The above deterministic paths correspond to pieces of diverging exponentials stopped at a certain coordinate (for $-m\leq k\leq h-1$) and remaining at zero afterwards.[20] Thus, during an extreme event, the trajectory of the aggregated anticipative AR(1) necessarily follows an explosive exponential path of a certain growth rate ρ_j^{-1}, possibly ending on a return to non-extreme levels. The behaviour that Gouriéroux et Zakoïan (2017) [63] intuitively conjectured, that is, that bubbles of different growth rates would appear on the trajectories, is hence formalised and the occurrence probability of each type of bubble can be obtained explicitly. In particular, on notices that for $J=1$, that is, for the non-aggregated anticipative AR(1), a single growth rate is possible and the process only generates a single type of bubble. This makes apparent that non-aggregated processes have dynamics limited to a single pattern, recurrently appearing from one extreme event to another. Aggregated processes do not suffer this limitation and generate trajectories which can feature different patterns through time.

More than unconditional occurrence probabilities, one can evaluate the probability that the process will follow a certain path given that the past trajectory, consisting of $m+1$ observations, is observed to be in a neighourhood V on $S_{m+1}^{\|\cdot\|}$. The condition $\Gamma^{\|\cdot\|}\Big(B(V)\Big)>0$ indicates that one can only condition on Borel sets V that contain at least one «plausible» piece of trajectory of the process (X_t). For general anticipative stable aggregates, these «plausible», or «observable» trajectories of length $m+1$ are necessarily of the form

$$\pm\frac{(d_{j,k+m},\dots,d_{j,k})}{\|\boldsymbol{d}_{j,k}\|}.$$

[20] For $k<-m$, $\boldsymbol{d}_{j,k}=(0,\dots,0)$ and these indexes k do not intervene in the spectral measure $\Gamma^{\|\cdot\|}$.

In the case of the AR(1), one obtains

$$\frac{(d_{j,k+m},\dots,d_{j,k})}{\|\boldsymbol{d}_{j,k}\|}=\begin{cases}\dfrac{(\overbrace{\rho_j^{k+m},\dots,\rho_j,1,0,\dots,0}^{m+1})}{\|(\underbrace{\rho_j^{k+m},\dots,\rho_j,1,0,\dots,0}_{m+1},\underbrace{0,\dots,0}_{h})\|}, & \text{for} \quad k\in\{-m,\dots,-1\},\\[2em] \dfrac{(\overbrace{\rho_j^{m},\dots,\rho_j,1}^{m+1})}{\|(\underbrace{\rho_j^{m},\dots,\rho_j,1}_{m+1},\underbrace{0,\dots,0,0,\dots,0}_{h})\|}, & \text{for} \quad k\geq 0.\end{cases}$$

The «observable» trajectories consist either in an uninterrupted exponential trend or in an exponential trend followed by a return at zero. Intuitively, the uninterrupted exponential trend is easily identified with an observed «inflation phase» of a bubble, that is, for a certain $j_0\in\{1,\dots,J\}$, with an event of the form

$$\frac{(X_{t-m},\dots,X_t)}{\|\boldsymbol{X}_t\|}\in V,\quad \text{for} \quad V \quad \text{some neighbourhood of} \quad \frac{(\rho_{j_0}^m,\dots,\rho_{j_0},1)}{\|(\rho_{j_0}^m,\dots,\rho_{j_0},1,0,\dots,0)\|}.$$

Conditionally on such an observation, what are the potential future paths and their likelihoods of occurrences? To answer this question, one first needs to identify the elements of $V\times\mathbb{R}^h$ to which $\Gamma^{\|\cdot\|}$ assigns a positive mass, that is

$$B(V)^+:=\left\{\boldsymbol{s}\in V\times\mathbb{R}^h:\quad \Gamma^{\|\cdot\|}(\{\boldsymbol{s}\})>0\right\}.$$

It is shown in this case that

$$B(V)^+=\left\{\frac{\boldsymbol{d}_{j_0,k}}{\|\boldsymbol{d}_{j_0,k}\|}:\quad k\geq 0\right\},$$

and one deduces from (1.38) that with probability 1, asymptotically, the complete path $\boldsymbol{X}_t/\|\boldsymbol{X}_t\|$ will belong to an arbitrarily small neighbourhood of the points $(\boldsymbol{d}_{j_0,k}/\|\boldsymbol{d}_{j_0,k}\|)_{k\geq 0}$, which were explicited above. Thus, the future path will necessarily follow an exponential trend of same growth rate $\rho_{j_0}^{-1}$ as that of the observed trajectory, eventually followed by return to non-extreme levels at any date in the future. If the «crash» date of the bubble remains uncertain, one can nonetheless evaluate the probability that it occurs at any given future horizon. On the one hand, considering A an arbitrarily small neighbourhood of

$$\frac{(\overbrace{\rho_{j_0}^{k+m},\dots,\rho_{j_0}^{k}}^{m+1},\overbrace{\rho_{j_0}^{k-1},\dots,\rho_{j_0},1,0,\dots,0}^{h})}{\|(\rho_{j_0}^{k+m},\dots,\rho_{j_0}^{k},\rho_{j_0}^{k-1},\dots,\rho_{j_0},1,0,\dots,0)\|},\quad \text{for some} \quad k\in\{0,\dots,h-1\},$$

the event on the complete vector $\{\boldsymbol{X}_t/\|\boldsymbol{X}_t\|\in A\}$ corresponds to that for which the peak of the bubble is reached in exactly k periods, $0\leq k<h$. It is shown for this choice of A that

$$\mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t},A\Big|B(V)\Big)\underset{x\to+\infty}{\longrightarrow}\frac{\Gamma^{\|\cdot\|}\Big(A\cap B(V)\Big)}{\Gamma^{\|\cdot\|}\Big(B(V)\Big)}=\frac{\Gamma^{\|\cdot\|}\Big(A\cap B(V)^+\Big)}{\Gamma^{\|\cdot\|}\Big(B(V)^+\Big)}=|\rho_{j_0}|^{\alpha k}(1-|\rho_{j_0}|^{\alpha}).$$

On the other hand, for A an arbitrarily small neighbourhood of

$$\frac{(\overbrace{\rho_{j_0}^{m+h},\dots,\rho_{j_0}^{h}}^{m+1},\overbrace{\rho_{j_0}^{h-1},\dots,\rho_{j_0},1}^{h})}{\|(\rho_{j_0}^{m+h},\dots,\rho_{j_0}^{h},\rho_{j_0}^{h-1},\dots,\rho_{j},1)\|},$$

the event $\{\boldsymbol{X}_t/\|\boldsymbol{X}_t\| \in A\}$ corresponds to the peak of the bubble being reached in h periods or more. One gets

$$\mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t}, A \Big| B(V)\Big) \underset{x\to+\infty}{\longrightarrow} \frac{\Gamma^{\|\cdot\|}\Big(A \cap B(V)^+\Big)}{\Gamma^{\|\cdot\|}\Big(B(V)^+\Big)} = |\rho_{j_0}|^{\alpha h}.$$

Therefore, conditionally on the inflation phase of a bubble being observed, say, of growth rate $\rho_{j_0}^{-1}$, the probability that the peak will be reached in exactly k periods, $0 \leq k < h$ is given by $|\rho_{j_0}|^{\alpha k}(1 - |\rho_{j_0}|^{\alpha})$, whereas the probability that it will be reached in h periods or more is given by $|\rho_{j_0}|^{\alpha h}$. The interpretation of the conditional moments from the third chapter is retrieved with more flexibility.[21]

The approach followed here illustrates the use of Property (1.38) in the case of the aggregation of anticipative AR(1) but can be deployed to study any type of stable process, provided it is anticipative. An interpretation of path prediction in terms of pattern identification in trajectories emerges. Other processes are considered in the fourth chapter, and in particular a bivariate process, to highlight the potential of the proposed framework.

[21] In the third chapter, the forms of the conditional moments of X_{t+h} given X_t were similar to that a two-point distribution: either the bubble will have collapsed strictly before h periods, either the bubble will survive h periods or more. The respective probability of both events were $1 - |\rho|^{\alpha h}$ et $|\rho|^{\alpha h}$ which are indeed compatible with the distribution obtained in the fourth chapter.

Chapter 2

Mixed Causal-Noncausal AR Processes and the Modelling of Explosive Bubbles

Sébastien Fries and Jean-Michel Zakoïan

Abstract Noncausal autoregressive models with heavy-tailed errors generate locally explosive processes and therefore provide a natural framework for modelling bubbles in economic and financial time series. We investigate the probability properties of mixed causal-noncausal autoregressive processes, assuming the errors follow a stable non-Gaussian distribution. Extending the study of the noncausal AR(1) model by Gouriéroux and Zakoian (2017), we show that the conditional distribution in direct time is lighter-tailed than the errors distribution, and we emphasize the presence of ARCH effects in a causal representation of the process. Under the assumption that the errors belong to the domain of attraction of a stable distribution, we show that a causal AR representation with non-i.i.d. errors can be consistently estimated by classical least-squares. We derive a portmanteau test to check the validity of the estimated AR representation and propose a method based on extreme residuals clustering to determine whether the AR generating process is causal, noncausal or mixed. An empirical study on simulated and real data illustrates the potential usefulness of the results.

Keywords: Noncausal process, Stable process, Extreme clustering, Explosive bubble, Portmanteau test.

2.1 Introduction

In the analysis of prices of financial assets such as stocks, it is common to observe phases of locally explosive behaviours, together with heavy-tailed marginal distributions and volatility clustering. Such features seem incompatible with classical *linear* models (namely the class of autoregressive-moving average (ARMA) models) which rely on the second-order properties of a time series. On the other hand, nonlinear models such as ARCH or stochastic volatility models are designed to capture volatility clustering, not to produce locally explosive sample paths mimicking bubbles in financial markets. However, the dynamic limitations of ARMA models are reduced if noncausal components (i.e. AR or MA polynomials with roots inside the unit disk) are introduced. For instance, all-pass models[1] are linear time series with nonlinear behaviours, in particular ARCH effects (see [15] and the references therein). More recently, Gouriéroux and Zakoian (2017 [63], GZ hereafter) showed that a simple noncausal AR(1) process with heavy-tailed errors is able to produce typical nonlinear behavior observed in the prices of financial assets.

Noncausal processes or random fields have been thoroughly studied in the statistical literature [3, 112], and have been applied in various areas, including deconvolution of seismic signals [45, 74, 124] and analysis of astronomical data [119]. Recent years have witnessed the emergence of a significant line of research on noncausal models in the econometric literature (see e.g., [24, 26, 42, 67, 68, 69, 70, 86, 87, 122]). The distinction between causal and noncausal processes is only meaningful in a non-Gaussian framework, and the increasing interest in Mixed causal-noncausal AR processes (MAR) parallels the widespread use of non-Gaussian heavy-tailed processes in economic or financial applications. Besides, rational expectations models in economics have been shown to admit solutions with noncausal components when departing from the finite variance assumption (see [61]).

One important reason for introducing noncausal components in AR processes is to provide a mechanism for generating financial bubbles. GZ showed that the sample paths of a stationary noncausal AR(1) process with heavy-tailed errors may have locally explosive phases. Other recent researches have focused on data generating processes that are able to produce explosive behaviours and model bubbles in financial markets. For example Phillips , Wu and Yu (2011) [106], Phillips , Shi and Yu (2015) [105] and more recently, in a continuous time framework, Chen, Phillips and Yu (2017) [27] investigated mildly explosive processes. Apart from the generation of bubbles, noncausal AR(1) processes with stable distributed errors exhibit surprising features such as a predictive distribution with lighter tails than the marginal distribution, a martingale property in the causal representation when the errors follow a Cauchy distribution, or the presence of GARCH effects. It is of interest to know whether these structural properties extend to higher-order models. Indeed, first-order models are clearly not sufficient to capture complex behaviours of economic series, such as the occurrence of locally explosive behaviours with different rates of explosion, or different types of asymmetries in the growth and downturn phases of the bubbles.

The aim of this paper is to analyze the class of mixed causal-noncausal AR processes with heavy-tailed

[1] All-pass are ARMA models in which all roots of the AR polynomial are reciprocal of the roots of the MA polynomial.

errors. The probability structure is studied under the assumption that the errors follow stable non-Gaussian distributions. Properties of the Least-Squares (LS) estimator are derived under the less stringent assumption that the noise distribution is in the domain of attraction of a non-Gaussian stable law. The paper is organised as follows. Section 2.2 studies the sample paths and the marginal distribution of MAR processes with stable errors. Sections 2.3 analyzes the conditional distributions through conditional moments. Conditional heteroscedasticity effects are depicted and causal representations are exhibited. Section 2.4 derives the asymptotic properties of the LS estimator, deduces a portmanteau test, and studies identification of the strong representation based on the analysis of extreme residuals clustering. Sections 2.5 and 2.6 propose numerical illustrations based on simulated and real data, respectively. Section 2.7 concludes. Proofs are collected in Section 2.8 and complementary results are provided in an Appendix.

2.2 Stable MAR(p, q) processes

MAR processes have been considered, among others, by Lanne and Saikkonen (2011) [87], Gouriéroux and Jasiak (2016) [59], Hecq, Issler, and Telg (2017) [66].[2] A MAR(p, q) process (X_t) is the strictly stationary solution of the difference equation

$$\psi(F)\phi(B)X_t = \varepsilon_t, \qquad \text{where} \quad \psi(F) = 1 - \sum_{i=1}^{p} \psi_i F^i, \quad \phi(B) = 1 - \sum_{i=1}^{q} \phi_i B^i, \tag{2.1}$$

B and F are the usual lag and forward operators ($B^k X_t = X_{t-k}, F^k X_t = X_{t+k}, k \in \mathbb{Z}$), (ε_t) is an independent and identically distributed (i.i.d.) sequence, the polynomials ψ and ϕ have all their roots outside the unit circle and are such that $\psi_p \neq 0$ and $\phi_q \neq 0$. When $q = 0$ (resp. $p = 0$), the model is called purely noncausal (resp. causal).

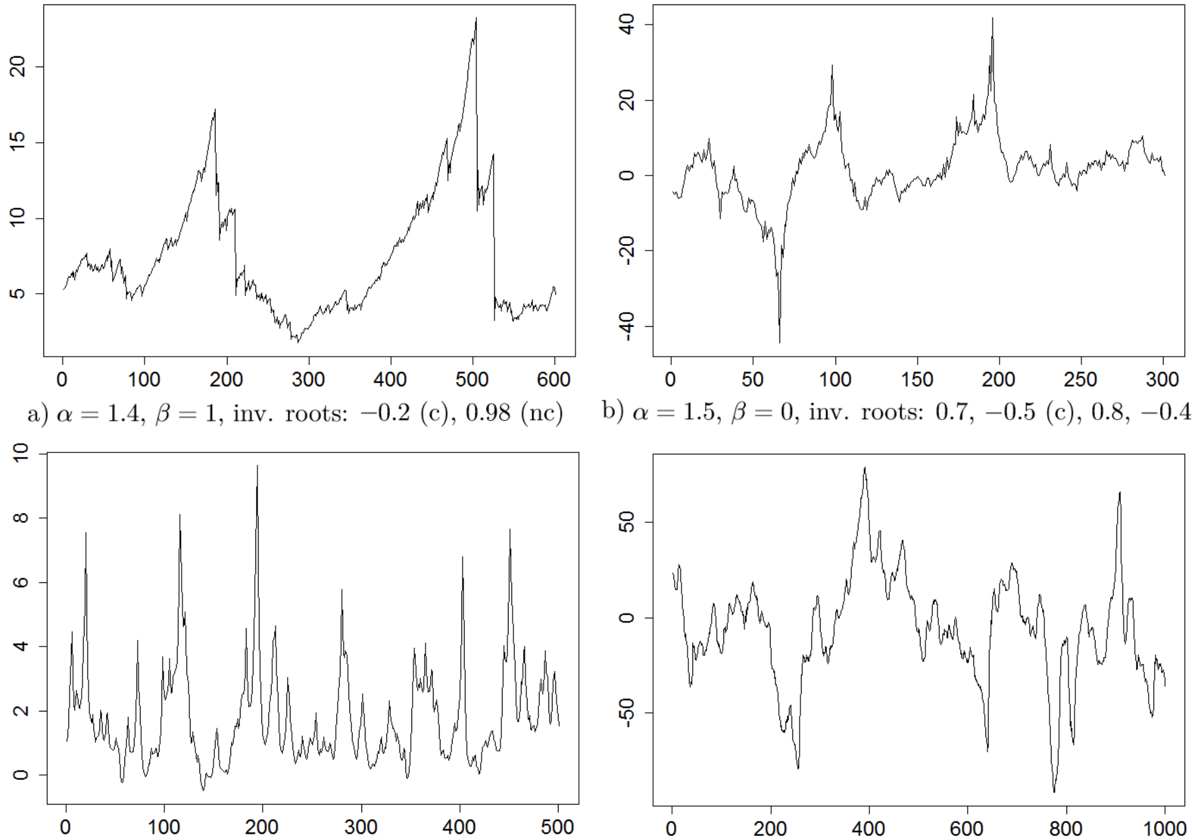
We assume that the errors ε_t follow a stable non-Gaussian distribution but the assumption will be relaxed for the statistical inference. The generality and convenience of this class of distributions is now well established.[3] Stable laws are easily characterised through their characteristic function: ε_t is said to follow a stable distribution with parameters $\alpha \in]0, 2[, \beta \in [-1, 1], \sigma > 0$, $\mu \in \mathbb{R}$, denoted $\varepsilon_t \sim \mathcal{S}(\alpha, \beta, \sigma, \mu)$, if

$$\forall s \in \mathbb{R}, \quad \mathbb{E}(e^{is\varepsilon_t}) = \exp\Big\{ -\sigma^\alpha |s|^\alpha \left(1 - i\,\beta\, \mathrm{sign}(s) w(\alpha, s)\right) + is\mu \Big\}, \tag{2.2}$$

where $w(\alpha, s) = \mathrm{tg}\left(\frac{\pi\alpha}{2}\right)$, if $\alpha \neq 1$, and $w(1, s) = -\frac{2}{\pi} \ln|s|$, otherwise. A stable random variable X has regularly varying tails in the sense that $\mathbb{P}(X < -x) \sim c_\alpha(1 - \beta)x^{-\alpha}$ and $\mathbb{P}(X > x) \sim c_\alpha(1 + \beta)x^{-\alpha}$ as $x \to \infty$, with $c_\alpha > 0$ and $\beta \in (-1, 1)$.

[2] See the latter reference for additional motivations on the use of MAR processes in time series econometrics. The first two references develop forecasting procedures for noncausal MAR processes.

[3] See for instance [47, 117] for the main properties of stable distributions. A major justification for using stable distributions rather than other classes of heavy-tailed distributions (such as the Student's t, the hyperbolic distributions) is that they are the only possible limit distributions for properly normalized and centered sums of i.i.d. random variables (giving rise to generalized Central Limit Theorems). Moreover, they are sufficiently flexible to accommodate asymmetry as well as fat tails. Finally, moving average processes based on stable variables also follow stable distributions, as will be detailed below.



a) $\alpha = 1.4$, $\beta = 1$, inv. roots: -0.2 (c), 0.98 (nc)

b) $\alpha = 1.5$, $\beta = 0$, inv. roots: 0.7, -0.5 (c), 0.8, -0.4 (nc)

c) $\alpha = 1.3$, $\beta = 1$, inv. roots: 0.65 (c), 0.7 (nc)

d) $\alpha = 1.6$, $\beta = 0$, inv. roots: -0.8, 0.7 (c), 0.95, 0.5 (nc)

Figure 2.1: Examples of trajectories of MAR(1,1) (left panel) and MAR(2,2) (right panel) processes with different parameters (nc: inverse of noncausal roots; c: inverse of causal root).

2.2.1 Sample paths

Examples of trajectories of four noncausal MAR processes are displayed in Figure 2.1. It can be seen that the trajectories feature locally explosive trends which are suited for the modelling of bubbles and positive feedback loop phenomena. Bubbles can be trending either upward or downward depending on the value of β. When $\beta = 1$, the density of the errors is maximally skewed towards positive values, yielding trajectories like (a) and (c) which could be suited to model prices or volatilities. In particular, trajectory (a) displays bubble patterns similar to those of real prices (see for instance Figure 2.4 below). The influence of a smaller tail parameter α is visible when comparing trajectories (c) and (d): the extreme events of the former ($\alpha = 1.3$) are more recurrent and further away from the central values than those of the latter ($\alpha = 1.6$).

Under the assumptions made on the AR polynomial, (X_t) admits a two-sided MA(∞) representation[4]

$$X_t = \sum_{k=-\infty}^{+\infty} d_k \varepsilon_{t+k}. \tag{2.3}$$

[4] It follows from Proposition 13.3.1 in Brockwell and Davis (1991) that the infinite sum in (2.3) is well defined under the stable law assumption, which ensures the existence of $\mathbb{E}|\varepsilon_t|^s$ for $s < \alpha$.

A simple index change $X_t = \sum_{\tau \in \mathbb{Z}} \varepsilon_\tau d_{\tau-t}$ allows to interpret the sample path of X_t as a linear combination of *baseline paths*, $t \longmapsto d_{\tau-t}$, weighted by stochastic i.i.d. coefficients ε_τ. Figure 2.2 depicts such baseline paths for four different MAR processes. The first panel illustrates the well-known impulse response function of a classical causal AR(1). The second panel displays an explosive exponential trend followed by a downward, faster decay and corresponds to the baseline path of a MAR(1,1) process. The remaining panels show more complex trajectories: the third one depicts the baseline path of a MAR(2,2) with dented upward and downward trends whereas the last one, corresponding to a noncausal AR(4) with two real and two conjugated complex roots, shows an upward trend with oscillations of increasing amplitudes and fixed pseudo-periods.

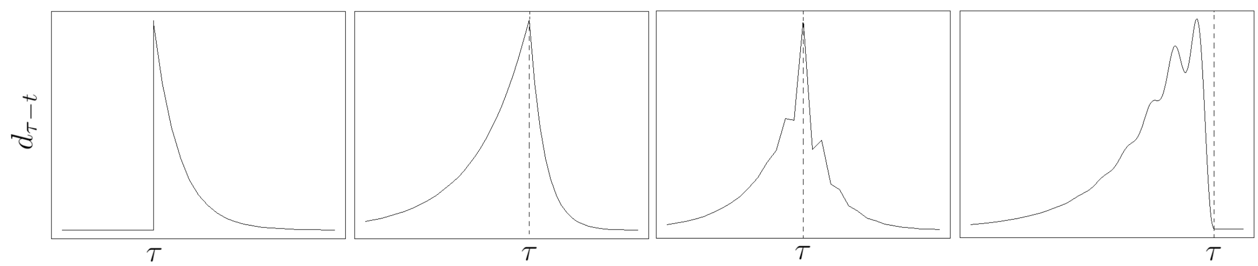


Figure 2.2: Examples of baseline paths $t \longmapsto d_{\tau-t}$ of MAR processes with characteristic polynomials, from left to right: $1-0.7B$; $(1-0.9F)(1-0.7B)$; $(1-0.8F)(1+0.4F)(1-0.7B)(1+0.5B)$; $(1-0.99F)(1-965F)(1-0.98e^{i0.045\pi}F)(1-0.98e^{-i0.045\pi}F)$.

2.2.2 Marginal distribution

Our first result characterises the marginal distribution of the stable MAR(p,q).

Proposition 2.2.1 *Let (X_t) the strictly stationary solution of the MAR(p,q) Model (2.1) where the roots of the polynomials ψ and ϕ are outside the unit disk and $\varepsilon_t \sim \mathcal{S}(\alpha,\beta,\sigma,\mu)$. Then X_t has a stable stationary distribution, $X_t \sim \mathcal{S}(\tilde{\alpha},\tilde{\beta},\tilde{\sigma},\tilde{\mu})$ where*

$$\tilde{\alpha} = \alpha, \qquad \tilde{\beta} = \beta \, \frac{\sum_{k=-\infty}^{+\infty} |d_k|^\alpha \, sign(d_k)}{\sum_{k=-\infty}^{+\infty} |d_k|^\alpha},$$

$$\tilde{\sigma} = \sigma \left(\sum_{k=-\infty}^{+\infty} |d_k|^\alpha \right)^{\frac{1}{\alpha}}, \qquad \tilde{\mu} = \frac{\mu}{\phi(1)\psi(1)} - \mathbb{1}_{\{\alpha=1\}} \frac{2}{\pi} \beta\sigma \sum_{k=-\infty}^{+\infty} d_k \ln |d_k|.$$

It is worth noting that the tail index α of X_t is that of the error term. In particular, $\mathbb{E}|X_t|^s < +\infty$ for $s < \alpha$ and $\mathbb{E}|X_t|^\alpha = +\infty$.

2.3 Predictive distributions

In the presence of a noncausal component in the AR polynomial, the predictive density of a future observation given a sample of consecutive observations is generally not available in closed form. We start by showing that the Markov property holds whatever the error distribution.

Proposition 2.3.1 *Let (X_t) the strictly stationary solution of the MAR(p, q) Model* (2.1) *where the roots of the polynomials ψ and ϕ are outside the unit disk and (ε_t) is an i.i.d. sequence (not necessarily stable). Then (X_t) is a homogeneous Markov chain of order $p+q$.*

In the rest of the section, we will derive properties of the conditional distribution of X_t in direct time when the errors are stable-distributed. We will focus on (i) the existence of conditional moments; (ii) explicit derivation of predictive formulas for X_t; and (iii) the presence of ARCH effects in the case of the MAR(1,q) process. More specific results will be detailed for the MAR(1,1) process.

2.3.1 Existence of moments of the conditional distribution

It follows from Proposition 2.2.1 that $\mathbb{E}|X_t|^s = \infty$ for $s \geq \alpha$. The next result shows a different behaviour for the conditional moments, generalising the result obtained for the AR(1) by GZ.

Theorem 2.3.1 *If (X_t) is the MAR(p, q) solution of Model* (2.1) *with $\varepsilon_t \sim \mathcal{S}(\alpha, \beta, \sigma, \mu)$, we have*

$$\mathbb{E}\left[\,|X_t|^{\gamma}|\, X_{t-1}, X_{t-2}, \ldots\right] < \infty, \quad a.s., \quad whenever \quad 0 < \gamma < 2\alpha + 1.$$

The conditional distribution in direct time, that is with respect to the past observations, thus has lighter tails than both the marginal distribution and the distribution conditional on the future. In particular, whatever the heaviness of the tails of ε_t, the conditional expectation of X_t always exists. The conditional variance in direct time also exists provided that the tails of the errors distribution are not too fat ($\alpha > 1/2$).[5]

2.3.2 Prediction of future values for the MAR$(1, q)$ processes.

Prediction at any horizon can be fully characterised for the symmetric MAR$(1, q)$ process. The next proposition extends in a non trivial way the prediction formula obtained by GZ for the noncausal AR(1), i.e. for the MAR(1,0). Let $\mathscr{F}_t = \sigma(X_t, X_{t-1}, \ldots)$ the canonical filtration of process (X_t). For $x \neq 0$ and $r \in \mathbb{R}$, let $x^{<r>} = \text{sign}(x)|x|^r$.

Proposition 2.3.2 *Let the MAR$(1, q)$ process $(1-\psi F)\phi(B)X_t = \varepsilon_t$, under the assumptions of Model* (2.1), *with $\varepsilon_t \sim \mathcal{S}(\alpha, 0, \sigma, 0)$. Then there exists for any $h \geq 0$ a polynomial $\mathscr{P}_h$ of degree q such that*

$$\mathbb{E}\left[\,X_{t+h}|\, \mathscr{F}_{t-1}\right] = \mathscr{P}_h(B)X_{t-1}.$$

For $h = 0$, the above formula holds with

$$\mathscr{P}_0(B)X_{t-1} = \psi^{<\alpha-1>}X_{t-1} + (1 - \psi^{<\alpha-1>}B)(\phi_1 X_{t-1} + \ldots + \phi_q X_{t-q}),$$

[5] A discrepancy between conditions of existence for marginal and conditional moments also holds for many nonlinear causal models: for instance GARCH (see e.g. Francq and Zakoian (2011), Chapter 2), or models for time series of counts (Davis and Liu, 2012).

and we have the semi-strong causal representation

$$(1-\psi^{<\alpha-1>}B)\phi(B)X_t = \eta_t, \tag{2.4}$$

with $\mathbb{E}\left[\eta_t \middle| \mathscr{F}_{t-1}\right] = 0$.

The proof is based on: i) disentangling pure causal and noncausal components of the MAR process (in the spirit of Lanne and Saikkonen (2011) [87], Gouriéroux and Jasiak (2016) [59]); ii) using the closed-form expression of the conditional expectation of the pure noncausal component, and (iii) invoking the Markov property.[6]

It is worth noting that the conditional expectation is linear in the past and can be explicitly computed. By comparison with finite variance AR processes, the semi-strong representation (2.4) is surprising. Indeed, in the L^2 framework, if (X_t) is mixed causal-noncausal satisfying $\psi(F)\phi(B)X_t = \varepsilon_t$, then there exists a causal version of (X_t) given by $\psi(B)\phi(B)X_t = Z_t$, where (Z_t) is uncorrelated with zero mean and finite variance (see for instance [19], Section 4.4).[7] In our framework, the noncausal component $(1-\psi F)$, with $|\psi| < 1$, is transformed into the causal component $(1-\psi^{<\alpha-1>}B)$.

In the Cauchy case ($\alpha = 1$) we get, when $\psi > 0$,

$$\mathbb{E}\Big[X_t \Big| \mathscr{F}_{t-1}\Big] = X_{t-1} + (1-B)(\phi_1 X_{t-1} + \ldots + \phi_q X_{t-q}), \tag{2.5}$$

with by convention $\phi_1 = \ldots = \phi_q = 0$ when $q = 0$. Hence, the martingale property established by GZ (Proposition 3.3), $\mathbb{E}\big[X_t \big| \mathscr{F}_{t-1}\big] = X_{t-1}$, only holds for the noncausal AR(1) (i.e. when $q = 0$).

The asymptotic behaviour of the conditional expectation -when the horizon h tends to infinity- is highly dependent on the tail index α. Proposition 2.3.2 allows us to distinguish different behaviours summarised in the following Corollary.

Corollary 2.3.1 *Under the assumptions of Proposition 2.3.2, we have almost surely*

$$\Big|\mathbb{E}\left[X_{t+h} \middle| \mathscr{F}_{t-1}\right]\Big| \underset{h\to\infty}{\longrightarrow} \begin{cases} 0 & if \quad \alpha \in (1,2), \\ \ell_{t-1} & if \quad \alpha = 1, \end{cases}$$

where ℓ_{t-1} is an $\mathscr{F}_{t-1}$-measurable random variable. Moreover, when $\alpha \in (0,1)$ and $q = 1$,

$$\Big|\mathbb{E}\left[X_{t+h} \middle| \mathscr{F}_{t-1}\right]\Big| \underset{h\to\infty}{\longrightarrow} \infty.$$

If $\alpha \in (1,2)$, that is for lighter tails within the stable family, the conditional expectation always tends to 0 which is the unconditional expectation. This is consistent with the L^2 framework ([19], p.189). For $\alpha = 1$, the absolute value of the conditional expectation tends to a finite limit whereas the unconditional expectation does not exist. The general case when $\alpha \in (0,1)$ is more intricate and is detailed in Appendix.

[6] The inherent complexity of the pure noncausal component when $p > 1$, for which no such closed-form expression exists, does not allow us to go beyond $p = 1$ for the results of this section.

[7] The equality $\psi(F)Z_t = \psi(B)\varepsilon_t$ indeed implies that (Z_t) has a spectral density given by $f_Z(\lambda) = \frac{\sigma^2}{2\pi}\frac{|\psi(e^{-i\lambda})|^2}{|\psi(e^{i\lambda})|^2} = \frac{\sigma^2}{2\pi}$ where $\sigma^2 = \mathrm{Var}(\varepsilon_t)$. Therefore, (Z_t) is also a white noise with the same variance σ^2 as ε_t.

2.3.3 Conditional heteroskedasticity of the Cauchy MAR$(1, q)$

All-pass models are well known examples of strong linear models displaying ARCH effects (namely the correlation of the squares). However, such effects are difficult to characterise without an explicit specification of the errors specification. The following result provides an explicit characterization of ARCH effects through the conditional variance of MAR processes with Cauchy innovation, extending again the results obtained by GZ for the noncausal AR(1).

Proposition 2.3.3 *Let X_t be a MAR$(1,q)$ process $(1-\psi F)\phi(B)X_t = \varepsilon_t$ with $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(1,0,\sigma,0)$. Then, for any $h \geq 0$, there exists a polynomial $Q_h(z) = \sum_{i=0}^{h} q_{i,h} z^i$ such that*

$$\mathbb{V}\Big(X_{t+h}\Big|\mathscr{F}_{t-1}\Big) = \left((\phi(B)X_{t-1})^2 + \frac{\sigma^2}{(1-|\psi|)^2}\right)\left(c_h - \Big(Q_h(sign\ \psi)\Big)^2\right),$$

with $c_h = \sum_{i=0}^{h}\sum_{j=0}^{h} q_{i,h}q_{j,h}(sign\ \psi)^{i+j}|\psi|^{-\min(i,j)-1}$.

Polynomials $Q_h(z)$, for $h \geq 0$, are defined in the Appendix. The causal representation (2.4) can then be completed and reveals quadratic ARCH effects in the Cauchy MAR$(1,q)$ process.

Corollary 2.3.2 *Under the assumptions of Proposition 2.3.3, there exists a sequence (η_t) of random variables such that,*

$$(1 - sign\,(\psi)B)\phi(B)X_t = \sigma_t\eta_t,$$

$$\sigma_t^2 = \left(\frac{1}{|\psi|} - 1\right)(X_{t-1} - \phi_1 X_{t-2} - \ldots - \phi_q X_{t-q-1})^2 + \frac{\sigma^2}{|\psi|(1-|\psi|)},$$

where $\mathbb{E}\left[\eta_t|\,\mathscr{F}_{t-1}\right] = 0$, $\mathbb{E}\left[\eta_t^2\big|\,\mathscr{F}_{t-1}\right] = 1$.

The process $e_t = \sigma_t\eta_t$ is however not an ARCH in the strict sense: first, because the errors η_t are not i.i.d., and second, because the volatility is a function of the X_{t-i} (not of the e_{t-i}). This representation is actually closer to the Double Autoregressive model studied by Ling (2007) [96] (see also [104] for a multivariate extension).

2.3.4 The MAR$(1,1)$ process.

The results of this section can be made completely explicit for the MAR(1,1) model defined by

$$(1-\psi F)(1-\phi B)X_t = \varepsilon_t, \quad \text{with} \quad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha,\beta,\sigma,\mu), \tag{2.6}$$

with $|\phi| < 1$ and $0 < |\psi| < 1$. The coefficients of the MA(∞) representation (2.3) are given by: $d_k = \dfrac{\psi^k}{1-\phi\psi}$, for any $k \geq 0$, and $d_k = \dfrac{\phi^{-k}}{1-\phi\psi}$, for any $k \leq 0$. Then $X_t \sim \mathcal{S}\left(\alpha,\tilde{\beta},\tilde{\sigma},\tilde{\mu}\right)$ with

$$\tilde{\beta} = \beta\left(\frac{1-\text{sign}(\phi)|\phi\psi|^\alpha}{1-|\phi\psi|^\alpha}\right)\left(\frac{1-\text{sign}(\psi)|\psi|^\alpha}{1-|\psi|^\alpha}\right)\left(\frac{1-\text{sign}(\phi)|\phi|^\alpha}{1-|\phi|^\alpha}\right),$$

$$\tilde{\sigma} = \frac{\sigma}{1-\phi\psi}\left(\frac{1-|\phi\psi|^\alpha}{(1-|\psi|^\alpha)(1-|\phi|^\alpha)}\right)^{\frac{1}{\alpha}},$$

$$\tilde{\mu} = \frac{\mu}{(1-\psi)(1-\phi)} - \mathbb{1}_{\{\alpha=1\}}\frac{2\beta\sigma}{\pi(1-\phi\psi)}\left[\frac{\psi\ln|\psi|}{(1-\psi)^2} + \frac{\phi\ln|\phi|}{(1-\phi)^2} - \frac{(1-\phi\psi)\ln|1-\phi\psi|}{(1-\psi)(1-\phi)}\right].$$

In particular, when $\psi, \phi > 0$ and the errors are Cauchy distributed, that is when $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(1, 0, \sigma, 0)$, then the above formulae simplify and $X_t \sim \mathcal{S}\left(1, 0, \frac{\sigma}{(1-\psi)(1-\phi)}, 0\right)$.

We now derive an explicit prediction formula for the MAR(1,1) process when $\beta = \mu = 0$. Proposition 2.3.2 yields for any $h \geq 0$,

$$\mathbb{E}\Big[X_{t+h}\Big|\mathscr{F}_{t-1}\Big] = \phi^{h+1}X_{t-1} + (X_{t-1} - \phi X_{t-2})(\psi^{<\alpha-1>})^{h+1}\sum_{i=0}^{h}\left(\phi\psi^{<1-\alpha>}\right)^i,$$

$$= \begin{cases} \phi^{h+1}X_{t-1} + \dfrac{(\psi^{<\alpha-1>})^{h+1} - \phi^{h+1}}{1-\phi\psi^{<1-\alpha>}}(X_{t-1} - \phi X_{t-2}), & \text{if} \quad \phi\psi^{<1-\alpha>} \neq 1, \\ \phi^{h+1}\left[X_{t-1} + (h+1)(X_{t-1} - \phi X_{t-2})\right], & \text{if} \quad \phi\psi^{<1-\alpha>} = 1. \end{cases}$$

When $\psi > 0$ and $\alpha = 1$, Corollary 2.3.2 yields

$$(1-B)(1-\phi B)X_t = \eta_t\sqrt{(\psi^{-1}-1)(X_{t-1} - \phi X_{t-2})^2 + \frac{\sigma^2}{\psi(1-\psi)}},$$

where $\mathbb{E}\left[\eta_t \middle| \mathscr{F}_{t-1}\right] = 0$ and $\mathbb{E}\left[\eta_t^2 \middle| \mathscr{F}_{t-1}\right] = 1$. The conditional variance at horizon h in Proposition 2.3.3 takes the more explicit form, for any $h \geq 0$,

$$\mathbb{V}\Big(X_{t+h}\Big|\mathscr{F}_{t-1}\Big) = \left[c_h - \left(\frac{1-\phi^{h+1}}{1-\phi}\right)^2\right]\left((X_{t-1} - \phi X_{t-2})^2 + \frac{\sigma^2}{(1-\psi)^2}\right),$$

where

$$c_h = \frac{(1+\phi\psi)\psi^{-h-1}}{(1-\phi\psi)(1-\phi^2\psi)} - \frac{2\phi^{h+1}}{(1-\phi)(1-\phi\psi)} + \frac{(1+\phi)\phi^{2(h+1)}}{(1-\phi)(1-\phi^2\psi)}.$$

2.4 Statistical Inference

This section is devoted to the LS estimation of the MAR(p, q) model

$$\psi_0(F)\phi_0(B)X_t = \varepsilon_t, \tag{2.7}$$

where $\psi_0(z) = 1 - \sum_{i=1}^{p}\psi_{0i}z^i, \phi_0(z) = 1 - \sum_{i=1}^{q}\phi_{0i}z^i$, with $\psi_0(z) \neq 0$ and $\phi_0(z) \neq 0$ for $|z| \leq 1$.

Contrary to other estimation methods such as Maximum Likelihood (ML)[8], LS do not require full specification of the errors distribution. We relax the assumption that (ε_t) is an α-stable sequence and rather assume that the law of ε_t belongs to the domain of attraction of a stable distribution. Specifically, we assume that there exists a function L which is slowly varying at infinity[9] be such that

$$\mathbb{P}(|\varepsilon_0| > x) = x^{-\alpha}L(x), \qquad \text{and} \qquad \lim_{x\to\infty}\frac{\mathbb{P}(\varepsilon_0 > x)}{\mathbb{P}(|\varepsilon_0| > x)} \to c \in [0, 1]. \tag{2.8}$$

[8] See [3] for asymptotic properties of the ML estimator of both causal and noncausal AR processes with non-Gaussian α-stable distribution. In the finite variance setting, ML estimation of MAR models based on Student's t distribution was studied by Hecq et al. (2016) [67].

[9] i.e. $\lim_{x\to\infty} L(tx)/L(x) = 1, \forall t > 0$.

This more general assumption on the errors distribution encompasses in particular the fully parametric α-stable framework under which the properties of the previous section were derived. Replacing α-stable laws by their domain of attraction alleviates the risk of misspecification.[10]

We will first derive the asymptotic properties of estimators of an "all-pass causal representation" of the MAR(p, q) process. Then, we will develop a portmanteau test for checking the validity of the estimated representation. Finally, we will consider selecting the true model, among the different specifications admitting the same all-pass representation, based on properties of extreme clustering.

2.4.1 All-pass causal representation

A difficulty in the inference of mixed causal-noncausal AR processes, is that many representations with seemingly uncorrelated errors hold. Breidt, Davis and Trindade (Section 4.3, 2001) [15] showed that if (X_t) is the strictly stationary solution of Model (2.7)-(2.8), then for any polynomial $\eta_0^*(z)$ obtained from $\psi_0(z)\phi_0(z)$ by replacing one or several roots by their inverses, we have

$$\eta_0^*(B)X_t = \zeta_t^*, \tag{2.9}$$

where (ζ_t^*) is an all-pass process.[11] Such representations (2.9) will be called all-pass in the following. In the set of all-pass representations, one is characterized by a polynomial η_0 having all its roots outside the unit disk

$$\eta_0(B)X_t = \zeta_t, \qquad \text{where} \quad \eta_0(B) = \psi_0(B)\phi_0(B) = 1 - \sum_{i=1}^{p+q} \eta_{0i}B^i, \tag{2.10}$$

and (ζ_t) is an all-pass process. In the sequel, we call (2.10) the *all-pass causal* representation of (X_t).

Now, let $\rho(h) = \left(\sum_{k=-\infty}^{\infty} d_k d_{k-h}\right) / \left(\sum_{k=-\infty}^{\infty} d_k^2\right)$ for $h \in \mathbb{Z}$, where the d_k's are the MA(∞) coefficients in (2.3).

Proposition 2.4.1 *Let (X_t) be the strictly stationary solution of model (2.7) under (2.8). Then, the $\rho(h)$'s satisfy the recursion*

$$\rho(h) = \sum_{i=1}^{p+q} \eta_{0i}\rho(h-i), \quad \forall h > 0, \tag{2.11}$$

where the coefficients η_{0i} are obtained from (2.10).

It is worth noting that, although the autocorrelations of X_t do not exist, the empirical autocorrelations can be computed and converge to the coefficients $\rho(h)$, which satisfy the usual Yule-Walker equations. Such equations explain why the coefficients of the all-pass causal representation of (X_t) can be consistently estimated by LS.

[10] The same assumption was considered in the context of causal AR processes for the study of least-absolute deviation (LAD) estimators by An and Chen (1982) [1], and for M-estimators by Davis, Knight and Liu (1992) [38].

[11] When the second-order moments are finite, all-pass processes are uncorrelated. Andrews and Davis (2013) showed that this property continues to hold "empirically" in the infinite variance case, in the sense that the sample autocorrelations converge to zero as the sample size goes to infinity.

2.4.2 Least-squares estimation

We consider LS parameter estimation of the all-pass causal representation (2.10), based on observations $X_1, \dots, X_n$ of the MAR(p,q) model (2.7). A LS estimator of $\boldsymbol{\eta}_0 = (\eta_{01}, \dots, \eta_{0,p+q})'$ is

$$\hat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta} \in \mathbb{R}^{p+q}} \mathcal{L}_n^*(\boldsymbol{\eta}), \tag{2.12}$$

where

$$\mathcal{L}_n^*(\boldsymbol{\eta}) = \sum_{t=p+q+1}^{n} \left(X_t - \sum_{i=1}^{p+q} \eta_i X_{t-i} \right)^2. \tag{2.13}$$

For $h \geq 0$, let $\hat{\gamma}(h) = \sum_{t=0}^{n-h} X_t X_{t+h}$ and denote $\hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0)$ the mean-unadjusted sample autocorrelation of order h. The LS estimator of $\boldsymbol{\eta}_0$ coincides, up to negligible terms, with the Yule-Walker estimator and is given by

$$\hat{\boldsymbol{\eta}} = \hat{\boldsymbol{\Gamma}}_n^{-1} \hat{\boldsymbol{\gamma}}_n, \quad \hat{\boldsymbol{\Gamma}}_n = [\hat{\gamma}(i-j)]_{i,j=1,\dots,p+q}, \quad \hat{\boldsymbol{\gamma}}_n = [\hat{\gamma}(i)]_{i=1,\dots,p+q}. \tag{2.14}$$

Proposition 2.4.2 *Let (X_t) be the strictly stationary solution of model (2.7)-(2.8). Then the LS estimator $\hat{\boldsymbol{\eta}}$ is consistent: $\hat{\boldsymbol{\eta}} \to \boldsymbol{\eta}_0$ in probability, as $n \to \infty$.*

To derive the asymptotic distribution of the LS estimator of $\boldsymbol{\eta}_0$, we introduce the sequences

$$a_n = \inf\{x : \mathbb{P}(|\varepsilon_0| > x) \leq n^{-1}\}, \quad \text{and} \quad \tilde{a}_n = \inf\{x : \mathbb{P}(|\varepsilon_0 \varepsilon_1| > x) \leq n^{-1}\}, \tag{2.15}$$

defined by Davis and Resnik (1986). Let $\boldsymbol{J}$ the $(p+q)\times(p+q)$ shift matrix, with ones on the superdiagonal and zeros elsewhere. For $\ell = 1, \dots, p+q$ let $\boldsymbol{K}^{(\ell)} = \boldsymbol{J}^\ell + {}^{\mathrm{t}}\boldsymbol{J}^\ell$ (with $\boldsymbol{K}^{(p+q)} = \mathbf{0}$). Let $\boldsymbol{L} = [\boldsymbol{K} \quad \boldsymbol{K}^{(2)} \quad \dots \; \boldsymbol{K}^{(p+q)}]$. We start by providing the asymptotic behaviour of the LS estimator under the simplifying assumption that the distribution of ε_t is symmetric. This assumption will be relaxed in the next section. The following result is a consequence of Davis and Resnik (1986) [41].

Proposition 2.4.3 *Let (X_t) be the strictly stationary solution of Model (2.7) with symmetric i.i.d. errors (ε_t) satisfying (2.8) and $\mathbb{E}|\varepsilon_t|^\alpha = \infty$.*
Then, letting $\boldsymbol{\rho} = [\rho(i)]_{i=1,\dots,p+q}$, $\boldsymbol{R} = [\rho(i-j)]_{i,j=1,\dots,p+q}$,

$$\frac{a_n^2}{\tilde{a}_n}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) \xrightarrow{d} \boldsymbol{R}^{-1}\{\boldsymbol{I}_{p+q} - \boldsymbol{L}(\boldsymbol{I}_{p+q} \otimes \boldsymbol{R}^{-1}\boldsymbol{\rho})\}\mathbf{Z}, \quad \textit{where} \quad \boldsymbol{Z} = (Z_1, \dots, Z_{p+q})', \tag{2.16}$$

$Z_k = \sum_{l=1}^{+\infty} \{\rho(k+l) + \rho(k-l) - 2\rho(l)\rho(k)\} S_l/S_0$, for $k = 1, \dots, p+q$, and $S_0, S_1, S_2, \dots$ are independent stable random variables; S_0 is positive with index $\alpha/2$ and S_j, for $j \geq 1$, has index α. If the law of $|\varepsilon_t|$ is asymptotically equivalent to a Pareto, (2.16) holds with $a_n^2/\tilde{a}_n = (n/\ln n)^{1/\alpha}$.

The α-stable domain of attraction assumption on (ε_t) impacts the asymptotic behaviour of the LS estimator in two important aspects: ι) the limiting distribution depends on α and $\iota\iota$) the convergence rate is $\frac{a_n^2}{\tilde{a}_n} \sim n^{1/\alpha}\tilde{L}(n)$, for some slowly varying function $\tilde{L}$ (see p.551, Davis and Resnick, 1986 [41]). Requiring $\mathbb{E}|\varepsilon_t|^\alpha = \infty$ ensures that the law of $\varepsilon_0\varepsilon_1$ belongs to the α-stable domain of attraction (see [34], Theorem 3.3 iv) p. 80).

Example 2.4.1 (MAR(1,1) process (continued)) For the MAR(1,1) process, Proposition 2.4.3 allows to compute the asymptotic distribution of the LS estimator of $(\phi_0+\psi_0, \phi_0\psi_0)$, using

$$\boldsymbol{R}^{-1}\{\boldsymbol{I}_2-\boldsymbol{L}(\boldsymbol{I}_2\otimes\boldsymbol{R}^{-1}\boldsymbol{\rho})\}=\frac{1+\phi_0\psi_0}{(1-\psi_0^2)(1-\phi_0^2)}\begin{pmatrix}(1+\psi_0\phi_0)^2+(\psi_0+\phi_0)^2 & -(\psi_0+\phi_0)\\ -2(\psi_0+\phi_0)(1+\psi_0\phi_0) & 1+\psi_0\phi_0\end{pmatrix}.$$

This matrix can be straightforwardly estimated by plugging LS estimators of $\phi_0+\psi_0$ and $\phi_0\psi_0$. From the estimated asymptotic distribution, at a given confidence level $\underline{\alpha}\in(0,1)$, we can deduce an asymptotic confidence region, $\mathcal{A}_{n,\underline{\alpha}}$ say, such that $P[(\phi_0+\psi_0,\phi_0\psi_0)\in\mathcal{A}_{n,\underline{\alpha}}]=1-\underline{\alpha}$. Denote by $r_1<r_2$ the inverses of the roots of $\eta_0(z)$ (thus $r_i\in\{\phi_0,\psi_0\}$ for $i=1,2$), that is $(r_1,r_2)=\left(\eta_{01}-\sqrt{\eta_{01}^2+4\eta_{02}},\eta_{01}+\sqrt{\eta_{01}^2+4\eta_{02}}\right)/2:=f(\eta_{01},\eta_{02})$. By the delta method, an asymptotic confidence region can be deduced for (r_1,r_2): let $\mathcal{R}_{n,\underline{\alpha}}$ such that $P[(r_1,r_2)\in\mathcal{R}_{n,\underline{\alpha}}]=1-\underline{\alpha}$. Thus, $P[\{(\phi_0,\psi_0),(\psi_0,\phi_0)\}\subset\mathcal{R}_{n,\alpha}]=1-\underline{\alpha}$. Finally, letting $\mathcal{R}^*_{n,\underline{\alpha}}$ the symmetric set of $\mathcal{R}_{n,\underline{\alpha}}$ around the line $r_1=r_2$, we get an asymptotic confidence region for (ϕ_0,ψ_0): $P[(\phi_0,\psi_0)\in\mathcal{R}_{n,\underline{\alpha}}\cup\mathcal{R}^*_{n,\underline{\alpha}}]\geq 1-\underline{\alpha}$.

The knowledge of index α is not required for the computation of the LS estimator, but the asymptotic distribution, as well as the normalizing constants a_n and $\tilde{a}_n$, depend on α. The presence of this nuisance parameter renders inference difficult for this class of model. Having estimated the AR coefficients, one could overcome this hurdle by using a standard estimator for the tail index α. For instance, for a random sample $(X_1,\dots,X_n)$, the so-called Hill (1975) [71] estimator of $1/\alpha$ based on $m+1$ upper order statistics is defined as:

$$\hat{\alpha}_m^{-1}=\frac{1}{m}\sum_{i=1}^{m}\log\left(\frac{X_{(i)}}{X_{(m+1)}}\right),$$

where $X_{(i)}>0$ is the ith order statistic in decreasing order $(X_{(1)}\geq X_{(2)}\geq\dots X_{(n)})$. Mason (1982) [100] proved that the Hill estimator is a consistent estimator of $1/\alpha$, provided $n\to\infty$, $m\to\infty$ and $m/n\to 0$, in the case of i.i.d. variables. Consistency and asymptotic normality under serial dependence conditions - including ℓ-dependence, β-mixing, ARCH - were established by various authors (see e.g. [43, 72] and the references therein). An alternative to the estimation of the asymptotic distribution is to base inference on bootstrap. Recently Cavaliere, Nielsen and Rahbek (2018) [24] proposed bootstrap schemes for noncausal AR models with infinite variance, and showed their usefulness for hypothesis testing. Extension of this approach to mixed AR models remains an open issue.

2.4.3 Relaxing the symmetry assumption

In the previous section, we derived the asymptotic behaviour of the LS estimator of η_0 assuming the errors (ε_t) were symmetrically distributed. We here relax the symmetry assumption and only require (ε_t) to satisfy (2.8). The asymptotic behaviour of the LS estimator remains unchanged in the case $0<\alpha<1$, and holds for $1<\alpha<2$ after a mean-adjustment.[12]

[12] A bias term appears in the case $\alpha=1$ when departing from the symmetry assumption (See Davis and Resnik (1986), Theorem 4.4).

Let $\tilde{\gamma}(h) = \sum_{t=0}^{n-h}(X_t - \bar{X})(X_{t+h} - \bar{X})$ where $\bar{X} = 1/n \sum_{t=0}^{n} X_t$, and denote $\tilde{\rho}(h) = \tilde{\gamma}(h)/\tilde{\gamma}(0)$ the mean-adjusted sample autocorrelation of order h. Similarly to (2.14), define the mean-adjusted Yule-Walker estimator $\tilde{\boldsymbol{\eta}}$ by

$$\tilde{\boldsymbol{\eta}} = \tilde{\boldsymbol{\Gamma}}_n^{-1} \tilde{\boldsymbol{\gamma}}_n, \quad \tilde{\boldsymbol{\Gamma}}_n = [\tilde{\gamma}(i-j)]_{i,j=1,\dots,p+q}, \quad \tilde{\boldsymbol{\gamma}}_n = [\tilde{\gamma}(i)]_{i=1,\dots,p+q}. \tag{2.17}$$

Proposition 2.4.4 *Let (X_t) be the strictly stationary solution of Model (2.7), where (ε_t) is an i.i.d. sequence satisfying (2.8) and $\mathbb{E}|\varepsilon_t|^\alpha = \infty$.*

- *If $0 < \alpha < 1$, then (2.16) holds.*
- *If $1 < \alpha < 2$, then (2.16) holds with $\hat{\boldsymbol{\eta}}$ replaced by the mean-adjusted estimator $\tilde{\boldsymbol{\eta}}$.*

2.4.4 Diagnostic checking

Validity of the estimated model can be assessed by studying the sample autocorrelations of the residuals. Once the parameters of the all-pass representation (2.10) have been estimated by LS, with $\hat{\boldsymbol{\eta}} = (\hat{\eta}_i)_{i=1,\dots,p+q}$, the corresponding residuals are defined by

$$\hat{\zeta}_t = X_t - \sum_{i=1}^{p+q} \hat{\eta}_i X_{t-i}, \qquad t = p+q+1, \dots, n. \tag{2.18}$$

Let, for $h \geq 0$, $\hat{\rho}_{\hat{\zeta}}(h) = \hat{\rho}_{\hat{\zeta}}(-h) = \frac{\hat{\gamma}_{\hat{\zeta}}(h)}{\hat{\gamma}_{\hat{\zeta}}(0)}$ where $\hat{\gamma}_{\hat{\zeta}}(h) = \sum_{t=h+p+q+1}^{n} \hat{\zeta}_t \hat{\zeta}_{t-h}$ and $\hat{\gamma}_{\hat{\zeta}}(-h) = \hat{\gamma}_{\hat{\zeta}}(h)$. For a fixed integer $H \geq 1$, let $\hat{\boldsymbol{\rho}}_{\hat{\zeta}} = [\hat{\rho}_{\hat{\zeta}}(1), \dots, \hat{\rho}_{\hat{\zeta}}(H)]'$.

Proposition 2.4.5 *Under the assumptions of Proposition 2.4.3, the vector of residual empirical autocorrelations satisfies*

$$\frac{a_n^2}{\tilde{a}_n} \hat{\boldsymbol{\rho}}_{\hat{\zeta}} \xrightarrow{d} \gamma(0) \boldsymbol{A_H Z}, \quad \textit{where} \quad \boldsymbol{Z} = (Z_1, \dots, Z_{H+p+q})',$$

where $\gamma(0) = \sum_{k=-\infty}^{\infty} d_k^2$, the Z_i's are as in Proposition 2.4.3, and A_H is a non random $H \times (p+q+H)$ matrix function of the sole AR coefficients (not of the error distribution).

Details regarding matrix A_H are available in the proof. Note that the symmetry assumption can be relaxed as in Section 2.4.3. It is now possible to propose a portmanteau test to check for residuals autocorrelations based for instance on the statistic

$$T_H = a_n^2 \tilde{a}_n^{-1} \sum_{i=1}^{H} |\hat{\rho}_{\hat{\zeta}}(i)| \underset{n \to +\infty}{\xrightarrow{d}} \|\gamma(0) \boldsymbol{A_H Z}\|_1, \tag{2.19}$$

with $\|\boldsymbol{x}\|_1 = \sum |x_i|$ for any vector $\boldsymbol{x} = (x_i)$. Practical implementation of the test finally requires simulating/bootstrapping the estimated asymptotic distribution in (2.19) (see Section 2.4.2).

2.4.5 Model selection based on extremes clustering

The all-pass causal representation (2.10) is compatible with all MAR(p', q') models of the form (2.9) (with $p' + q' = p + q$). Such models could have generated the observations, and it is important to detect which one is the true model. A distinctive feature of the latter is that the errors are i.i.d., not only "asymptotically empirically non-autocorrelated" as in (2.9) (see footnote [11]). Having estimated the coefficients of the polynomial η_0, a natural strategy for assessing the validity of a candidate model, with polynomial η_0^*, is to test the independence of the ζ_t^* in (2.9). We propose an approach based on extreme clustering of the residuals.[13]

Point process of exceedances

This dependence materialises here in an important feature known as extreme clustering (see e.g. [73, 99], and [25] for a literature review) which yields a way to identify the strong representation among the all-pass alternatives. Let us introduce a linear process (Y_t) with two-sided MA(∞) representation $Y_t = \sum_{k\in\mathbb{Z}} c_k \varepsilon_{t+k}$, where (ε_t) is an i.i.d. sequence satisfying (2.8), $\sum_{k\in\mathbb{Z}} |c_k|^s < +\infty$ for some $0 < s < \alpha$, $s \leq 1$, and assume $\max|c_k| = 1$ for convenience. In our context, Y_t will typically be substituted for the errors ε_t of the strong representation and the errors ζ_t^* of competing all-pass representations.
We can study the time indices for which $a_n^{-1}Y_k$ falls outside the interval $(-x, x)$, for $x > 0$. The corresponding point process converges as the number of observations n grows to infinity (see Davis and Resnick, Section 3.D, 1985 [40]):

$$\sum_{k=1}^{n} \delta_{\left(k/n, a_n^{-1}Y_k\right)}\Big(\,\cdot\, \cap B_x\Big) \xrightarrow{d} \sum_{k=1}^{+\infty} \xi_k \delta_{\Gamma_k}, \tag{2.20}$$

where δ is the Dirac measure, $B_x = (0, +\infty) \times \Big((-\infty, -x) \cup (x, +\infty)\Big)$, $\{\Gamma_k, k \geq 1\}$ are the points of a homogeneous Poisson Random Measure (PRM) on $(0, +\infty)$ with rate $x^{-\alpha}$,[14] and $\xi_k = \mathrm{Card}\{i \in \mathbb{Z} : J_k|c_i| > 1\}$ where $\{J_k, k \geq 1\}$ are i.i.d. on $(1, +\infty)$, independent of $\{\Gamma_k\}$, with common density:

$$f(z) = \alpha z^{-\alpha-1}\mathbb{1}_{(1,+\infty)}(z). \tag{2.21}$$

The sequences $\{\Gamma_k\}$ and $\{\xi_k\}$ are interpreted (see for instance [89]) as describing respectively the occurrence dates of clusters of extreme events and the size of these clusters (i.e. the number of co-occurring extreme events). We now outline some reasons for analysing extreme event clustering and the potential for model selection amongst competing representations.

Analysing the error processes of competing models

Let (X_t) be the MAR(p, q) strictly stationary solution of Model (2.8)-(2.7) and assume the order $p + q$ and the roots of η_0 are known. There is a finite number of competing representations among which the strong

[13] Alternative approaches based on non-parametric and rank-based tests have been proposed for testing iid-ness of innovations (see for instance [18, 46]). However, the validity of these tests requires $\sqrt{n}$-consistent estimators of the model parameters, which is not the case in the α-domain of attraction framework as shown in Proposition 2.4.3.

[14] See [37]: $\{\Gamma_k, k \geq 1\}$ are the points of a homogeneous PRM on $(0, +\infty)$ with rate $x^{-\alpha}$ if and only if, for any $\ell \geq 1$, nonnegative integers $a_1, \ldots, a_\ell$ and $b_1, \ldots, b_\ell$ such that $a_i < b_i \leq a_{i+1}$, $i = 1, \ldots, \ell$, and any nonnegative integers $n_1, \ldots, n_\ell$:

$$\mathbb{P}\Big(N(a_i, b_i] = n_i, i = 1, \ldots, \ell\Big) = \prod_{i=1}^{\ell} \frac{[x^{-\alpha}(b_i - a_i)]^{n_i}}{n_i!} \exp\Big\{-x^{-\alpha}(b_i - a_i)\Big\},$$

where $N(a_i, b_i]$ denotes the number of terms of $\{\Gamma_k, k \geq 1\}$ falling in the half-open interval $(a_i, b_i]$, $i = 1, \ldots, \ell$.

one lies. Denoting (ε_t) of the errors of the strong representation and generically (ζ_t^*) the errors of any specific all-pass representation, we can analyse their extreme clustering behaviour using (2.20).

Error process of the strong representation The i.i.d. errors (ε_t) of the strong representation admit the trivial MA form $\varepsilon_t = \sum_{k\in\mathbb{Z}} c_k \varepsilon_{t+k}$ with $c_0 = 1$ and $c_k = 0$ for $k \neq 0$. Thus, substituting Y_t for ε_t, (2.20) holds with $\xi_k = \text{Card}\{i \in \mathbb{Z} : J_k|c_i| > 1\} = \mathbb{1}_{\{J_k>1\}} = 1$. The random variables (ξ_k) describing the size of the cluster of extremes are degenerate in this case: the extreme errors (ε_t) tend to appear isolated from each other.

Error process of an all-pass representation The (rescaled) errors ζ_t^* of an all-pass representation always admit an infinite one-sided or two-sided MA form, say $\zeta_t^* = \sum\limits_{k\in\mathbb{Z}} \frac{c_k}{\max_j |c_j|}\varepsilon_{t+k}$. Denote $(c_{(k)})_{k\geq 1}$ the sequence obtained by sorting $(|c_k|)_{k\in\mathbb{Z}}$ in descending order. Substituting Y_t for ζ_t^*, (2.20) holds with $\xi_k = \text{Card}\left\{i \geq 1 : J_k \frac{c_{(i)}}{c_{(1)}} > 1\right\} = \arg\max_{i\geq 1}\{J_k > c_{(i)}^{-1} c_{(1)}\}$ and from (2.21), we deduce that for any $\ell \geq 1$:

$$\mathbb{P}\Big(\xi_k \geq \ell\Big) = \mathbb{P}\Big(J_k > c_{(\ell)}^{-1} c_{(1)}\Big) = c_{(\ell)}^{\alpha} c_{(1)}^{-\alpha}. \tag{2.22}$$

In this case, the ξ_k's can take arbitrarily high values with non-zero probability, indicating as expected that the extremes of (ζ_t^*) tend to cluster.

Errors at higher horizons Considering the extreme clustering of errors at further horizons can provide additional discriminating information. For simplicity, consider the noncausal AR(1) model. There are two competing models, yielding the same all-pass causal representation (2.10):

$$X_t = \psi_0 X_{t+1} + \varepsilon_t, \quad \text{and} \quad X_t = \psi_0 X_{t-1} + \zeta_t. \tag{2.23}$$

For any $h \geq 1$, expansions of these equations at horizons h read:

$$\varepsilon_{t+h|t} := X_t - \psi_0^h X_{t+h} = \varepsilon_t + \psi_0 \varepsilon_{t+1} + \ldots + \psi_0^{h-1} \varepsilon_{t+h-1}, \tag{2.24}$$

$$\zeta_{t+h|t} := X_{t+h} - \psi_0^h X_t = (\psi_0^{-h} - \psi_0^h) \sum_{k\geq h} \psi_0^k \varepsilon_{t+k} - \psi_0^h \sum_{k=0}^{h-1} \psi_0^k \varepsilon_{t+k}. \tag{2.25}$$

We can deduce that the point processes of excedances of the errors $\varepsilon_{t+h|t}$ and $\zeta_{t+h|t}$ at horizon h will exhibit clusters of random sizes $\xi_k = \text{Card}\left\{i \in \mathbb{Z} : J_k \frac{|c_i|}{\max_j |c_j|} > 1\right\}$ where $c_i = \psi_0^i$ if $0 \leq i \leq h-1$ for the strong model, whereas for the all-pass model, the sequence $(|c_i|)$ reads: $|\psi_0|^h, \ldots, |\psi_0|^{2h-1}, 1-\psi_0^{2h}, |\psi_0|(1-\psi_0^{2h}), |\psi_0|^2(1-\psi_0^{2h}), \ldots$. Thus, the extreme realisations of the errors (2.24) will appear by clusters of at most h consecutive observations, whereas the errors (2.25) will likely appear by larger clusters (see Appendix for illustration). This analysis can be extended to general MAR processes by disentangling the pure causal and noncausal components of each competiting model (as in the proof of Proposition 2.3.1).

Application to model selection

The previous section highlights that the extreme errors of all-pass representations are likely to appear in large clusters, contrary to the extreme errors of the strong representation that tend to appear isolated. Selecting the strong MAR(p, q) representation, assuming only $p+q$ known, can thus be achieved by looking for evidence of extreme clustering in the errors of all competing representations. In principle, such evidence shall be found in the errors of all representations but the strong one.

2.5 A Monte Carlo study

We conducted three types of experiments in order to gauge the sample properties of the LS procedure applied to the all-pass causal representation. On synthetic data generated from a MAR(1,1) process, we assessed ι) the consistency of the estimators of the roots and the convergence in distribution of the LS estimators of the backward AR(2) specification, $\iota\iota$) the empirical size of the portmanteau-type statistic, and $\iota\iota\iota$) the extreme clustering in the residuals of the four competing models that the LS estimation implies.

2.5.1 LS estimation

We simulated 100,000 paths with lengths 500, 2000 and 5,000 observations of α-stable MAR(1,1) processes solution of $(1-\psi_0 F)(1-\phi_0 B)X_t = \varepsilon_t$ with $\psi_0 = 0.7$, $\phi_0 = 0.9$ and tail indices $\alpha = 1.5$, 1 and 0.5. We computed the LS estimator $(\hat{\eta}_1, \hat{\eta}_2)$ and deduced estimators $(\hat{\psi}, \hat{\phi})$ by taking the inverses of the zeros of $1-\hat{\eta}_1 X - \hat{\eta}_2 X^2$ (we imposed $|\hat{\psi}| \leq |\hat{\phi}|$ for the sake of identifiability when the roots are real). For each model, Table 2.1 reports the empirical frequencies of estimators that are sufficiently close to the actual values of the roots. As expected, the accuracy increases with n but, more strikingly, it increases sharply as α approaches zero.

		$\alpha = 1.5$			$\alpha = 1$			$\alpha = 0.5$		
n		$a = 0.1$	$a = 0.05$	$a = 0.01$	$a = 0.1$	$a = 0.05$	$a = 0.01$	$a = 0.1$	$a = 0.05$	$a = 0.01$
500	$\hat{p}_a(\phi)$	99.8%	94.6%	33.3%	99.7%	96.4%	48.5%	99.1%	97.5%	71.4%
	$\hat{p}_a(\psi)$	78.2%	55.2%	18.7%	83.8%	69.7%	33.0%	86.2%	79.9%	58.6%
2000	$\hat{p}_a(\phi)$	99.9%	98.9%	54.3%	99.9%	99.2%	74.3%	99.8%	99.4%	90.3%
	$\hat{p}_a(\psi)$	96.3%	87.2%	34.6%	96.0%	91.5%	60.4%	96.4%	94.5%	84.6%
5000	$\hat{p}_a(\phi)$	99.9%	99.8%	74.4%	99.9%	99.7%	88.4%	99.9%	99.7%	95.8%
	$\hat{p}_a(\psi)$	98.7%	96.3%	53.6%	98.5%	96.9%	78.9%	98.6%	97.8%	93.2%

Table 2.1: Accuracy of the roots-estimation through backward LS: $\hat{p}_a(\theta)$ denotes the frequency of estimations $\hat{\theta}$ belonging to the set $\left\{\left|\hat{\theta}-\theta_0\right| < a\right\} \cap \left\{\hat{\theta} \in \mathbb{R}\right\}$, for $\theta = \phi$ or ψ, for $a = 0.01$, 0.05, 0.1 and over 100,000 simulated paths of the α-stable MAR(1,1) process (X_t) solution of $(1-\psi_0 F)(1-\phi_0 B)X_t = \varepsilon_t$, with $\psi_0 = 0.7$ and $\phi_0 = 0.9$.

Turning to the asymptotic distribution of $(\hat{\eta}_1, \hat{\eta}_2)$, results reported in Appendix file show that the finite sample distribution approaches its asymptotic behaviour much slower for lower values of α. In the same line, a direct implementation of the portmanteau test using the statistics (2.19) also showed heavy distortions in finite sample. These distortions were expected as they were already reported by Lin and McLeod (2008) [94] in the pure causal AR framework. To alleviate the problem, they suggested a Monte Carlo test which relies on simulations of the estimated causal AR model to approximate the distribution of the portmanteau statistics under the null hypothesis of correct specification. An important difference in our framework is that, under the null, the estimated causal AR is only an all-pass representation of the process and we use its estimated coefficients to simulate paths of the corresponding pure causal AR as if it were the strong representation. Given that the residuals autocorrelations of this pure causal AR have the same asymptotic distribution as those of the all-pass causal representation, the procedure remains valid.[15] We therefore proceeded with this methodology (see Appendix). The empirical sizes of the 1, 5 and 10% nominal tests for lags $H = 1, \dots, 10$ are reported in Table 2.2. It can be seen that using the Monte Carlo procedure, the portmanteau test is much better behaved in finite sample, especially for $\alpha = 1.5$, which is a realistic value for financial series.

	$\alpha = 1.5$			$\alpha = 1$			$\alpha = 0.5$		
H	1%	5%	10%	1%	5%	10%	1%	5%	10%
1	1.30	5.80	10.5	1.25	5.40	10.4	1.45	4.10	7.35
2	1.55	5.65	10.9	1.60	5.25	9.65	1.35	3.90	7.05
3	1.40	5.35	10.9	1.30	5.05	9.40	1.20	4.45	6.95
4	1.50	5.45	10.5	1.35	5.00	9.90	1.20	4.35	7.00
5	1.25	5.50	9.85	1.20	4.90	9.20	1.10	4.20	7.30
6	1.30	5.00	10.1	1.05	4.70	9.40	1.10	4.25	7.40
7	1.20	5.25	9.75	1.05	4.40	9.15	1.20	4.00	7.50
8	1.10	5.25	9.75	1.15	4.55	8.70	1.05	3.70	7.25
9	1.25	5.10	9.80	1.30	4.30	8.60	1.05	3.75	7.50
10	1.35	5.10	10.1	1.20	4.55	8.70	0.90	3.65	7.15

Table 2.2: Empirical sizes (%) of the portmanteau statistics (2.19) implemented by the Monte Carlo test procedure. The empirical size was calculated based on 2000 simulations of the α-stable MAR(1,1) process (X_t) solution of $(1 - \psi_0 F)(1 - \phi_0 B)X_t = \varepsilon_t$, with $\psi_0 = 0.7$ and $\phi_0 = 0.9$. Each Monte Carlo test was performed with 1000 simulations.

2.5.2 Selection based on extreme residuals clustering

We now gauge the usefulness of the results of Section 2.4.5 by simulating paths of the α-stable MAR(1,1) process $(1-\psi_0 F)(1-\phi_0 B)X_t = \varepsilon_t$ with different parameterisations and analysing the residuals of the competing

[15] It can indeed be noticed that the asymptotic distributions of the LS estimator and of the residuals autocorrelations remain unchanged whether X_t is defined as the solution of $\psi_0(F)\phi_0(B)X_t = \varepsilon_t$ or $\psi_0(B)\phi_0(B)X_t = \varepsilon_t$ in (2.7).

representations. There are four competing models yielding the same all-pass causal AR(2) representation:

$$\text{Pure causal AR(2):} \qquad (1-\psi_0 B)(1-\phi_0 B)X_t = \zeta_t, \tag{2.26}$$

$$\text{MAR(1,1):} \qquad (1-\psi_0 F)(1-\phi_0 B)X_t = \varepsilon_t, \tag{2.27}$$

$$\text{MAR(1,1):} \qquad (1-\psi_0 B)(1-\phi_0 F)X_t = \nu_t, \tag{2.28}$$

$$\text{Pure noncausal AR(2):} \qquad (1-\psi_0 F)(1-\phi_0 F)X_t = \omega_t, \tag{2.29}$$

where (ζ_t), (ν_t) and (ω_t) denote the sequences of errors of each all-pass representation. More specifically for each estimated model, we compute the errors at several horizons h (as in (2.24)-(2.25) for the AR(1). For each MAR(1,1) alternative, we disentangle the causal and noncausal components and compute their respective error series). For each errors series and a given threshold $x > 0$, we identify the clusters of consecutive extreme values, i.e., errors larger than x in modulus. As explained in Section 2.4.5, for any horizon h, we expect all-pass representations to display larger clusters of extreme errors than the strong model, for which clusters larger than h have zero probability. Letting $\hat{\xi}_{k,h}(x)$ denote the number of consecutive exceedances for the k-th cluster, we therefore propose an Excess Clustering (EC)[16]indicator defined as:

$$\mathrm{EC}_h = \frac{\sum\limits_{k/\hat{\xi}_{k,h}(x)>h} \left(\hat{\xi}_{k,h}(x) - h\right)}{\mathrm{Card}\{k : \hat{\xi}_{k,h}(x) > h\}}, \quad \text{if} \quad \mathrm{Card}\{k : \hat{\xi}_{k,h}(x) > h\} > 0, \quad \text{else} \quad \mathrm{EC}_h = 0. \tag{2.30}$$

We start by generating 10,000 sample paths of MAR(1,1) processes. For each path, we fit a backward AR(2), estimate the set $\{\phi_0, \psi_0\}$, and for each of the four competing models we compute the indicator (2.30) for $h = 1, \ldots, 20$. We refer to vectors $(EC_1, \ldots, EC_{20})$ as estimators of the *term structure* of residuals excess clustering. Averaging model-wise across the 10,000 simulations yields the typical excess clustering behaviors of the residuals of each competing model. We perform this experiment for several MAR(1,1) processes and display the results of two parameterisations in Figure 2.3 (see Appendix for additional results and details regarding the methodology).

It can be noticed that the all-pass models feature excessively clustering residuals at any horizon whereas the residuals of the strong model barely deviate from no excess clustering. As we could expect from (2.22), the heavier the tails the easier it is to identify dependent residuals. This is in line with the findings of Hecq, Lieb and Telg (2016) [67] who are concerned with identification of causal/noncausal models using the LAD estimator. Noticeably, even with very heavy tails ($\alpha = 0.5$), the residuals at any horizons of the strong representation still barely deviate from no excess clustering. These experiments highlight in addition the usefulness of considering residuals at various horizons, instead of focusing only on basic residuals. Indeed, all the term structures of excess clustering show that the contrast between the competing models does not

[16] For a given h, EC_h defined at (2.30) corresponds to the average size of clusters larger than h, from which we subtract h, and is 0 if all the clusters are smaller than h. It is related to the Extremal Index, more common in the literature, which is the reciprocal of the average size of clusters. Also, the choice of clustering scheme, i.e. how the sequence $(\hat{\xi}_{k,h}(x))_k$ is constructed, can have an impact on the estimated excess clustering : more elaborate clustering schemes could be considered (see for instance [48, 109]).

arise for $h = 1$ but rather tends to peak for intermediate values of h.

Last, we assess how well we can discriminate between the all-pass models and the strong representation by exploiting the excess clustering feature. For each of the 10,000 simulations, we rank the four competing models according to the area under the term structure curve of excess clustering (AUC) and select the candidate with least AUC. Table 2.3 reports the true positive rates of this procedure. For $\alpha = 1.5$ and $n = 500$, the strong representation was correctly identified in above 88% of the 10,000 simulated paths and this proportion increases with n.

$n = 500$	$n = 2000$	$n = 5000$
88.4%	95.8%	97.5%

Table 2.3: Correct model selection rates based on least excess clustering across 10,000 simulated paths of the MAR(1,1) process (X_t) solution of $(1 - 0.7F)(1 - 0.9B)X_t = \varepsilon_t$ with i.i.d. 1.5-stable noise.

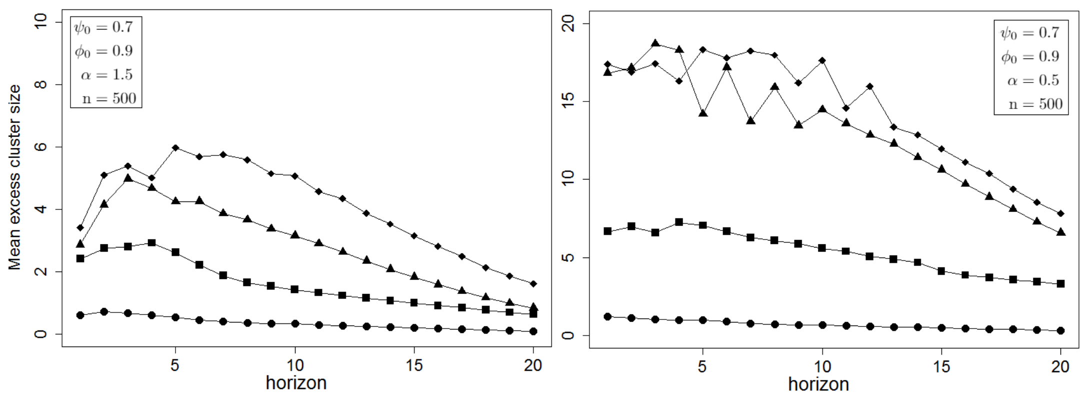


Figure 2.3: Across 10,000 simulations of the α-stable MAR(1,1) process (X_t) solution of $(1 - \psi_0 F)(1 - \phi_0 B)X_t = \varepsilon_t$, the plots show the average of the term structure of excess clustering of the linear residuals of the four competing models (2.26) (squares), the strong representation (2.27) (points), (2.28) (triangles) and (2.29) (diamonds). The parameterisations and path lengths are indicated on each panel.

2.6 An application to financial series

In this section, we illustrate the adequacy of MAR models for real economic series. We fitted MAR models on six financial and economic series: cotton price (monthly, USD/pound, August 1972 to July 2017)[17], soybean price (monthly, USD/bushel, January 1973 to May 2006)[18], sugar price (monthly, USD/pound, November 1962 to August 2018)[19], coffee price (monthly, USD/pound, April 1976 to May 2018)[20], Hang Seng Index

[17] https://www.macrotrends.net/2533/cotton-prices-historical-chart-data

[18] https://www.macrotrends.net/2531/soybean-prices-historical-chart-data

[19] https://www.macrotrends.net/2537/sugar-prices-historical-chart-data

[20] https://www.macrotrends.net/2535/coffee-prices-historical-chart-data

(HSI) (monthly, HKD, November 1986 to Mars 2017)[21], and the quarterly Shiller Price/Earning ratio (Q1 1881 to Q2 2017)[22]. In the following study, a linear trend is fitted and subtracted from the HSI series and the other series are centered. These series are depicted on Figure 2.4.

2.6.1 AR estimation and validation using the Monte Carlo Portmanteau test

We start by investigating the appropriate total AR order ($r = p + q$) for each series using the Monte Carlo portmanteau test of Section 2.5.1. For each series, starting from total AR order 1, we estimate all-pass causal representations of increasing order by LS and perform the portmanteau test with $H = 50$ lags using the Monte Carlo portmanteau procedure of Lin and McLeod (1000 paths were simulated for each test). The results of the portmanteau test, reported in Table 2.4, allow to discard non-admissible low order models at the level 5%. We retain for the following the lowest orders (indicated in bold) which pass the portmanteau procedure: cotton: 2; soybean: 5; sugar: 4; coffee: 4; HSI: 3; Shiller P/E: 6.[23]

Total AR order r	Cotton	Soybean	Sugar	Coffee	HSI	Shiller P/E
1	0.30	1.50	0.30	0.70	4.50	0.10
2	**6.60**	0.90	3.30	0.80	4.50	0.20
3		1.60	4.50	4.00	**7.10**	0.20
4		2.60	**8.20**	**7.99**		0.40
5		**15.4**				1.00
6						**5.90**

Table 2.4: P-values (%) of the Monte Carlo portmanteau tests with $H = 50$ lags for increasing AR order r. Rejection if P-value $< 5\%$.

2.6.2 MAR selection based on extreme clustering

For each of the mentioned series, we apply the methodology of Section 2.5.2: we fit all possible MAR models of total order $r = 1, \ldots, 6$, compute the term structure of excess clustering of the residuals of each competing model and the associated term structure of EC and we then rank the competing models according to the area under the term structure curve. In Table 2.5, we report for each total AR order r the MAR(p, q) specification which displays the lowest AUC of excess clustering and the median AUC of its competitors. We can notice that overall, excess clustering decreases as r increases. Combining the results from the portmanteau tests of the previous section and of the extreme clustering analysis, we select a final specification for each series as

[21] https://finance.yahoo.com/quote/%5EHSI/history?p=%5EHSI

[22] https://www.quandl.com/data/multpl/shiller_pe_ratio_month-shiller-pe-ratio-by-month

[23] This procedure yields as a by-product the McCulloch quantile estimates of the tail index α (see McCulloch (1986)) for the six series. Values of $\hat{\alpha}$: cotton: 1.53 ; soybean: 1.41 ; sugar: 1.35 ; coffee: 1.39 ; HSI: 1.34 ; Shiller P/E: 1.49. In all the cases, the infinite variance hypothesis is plausible.

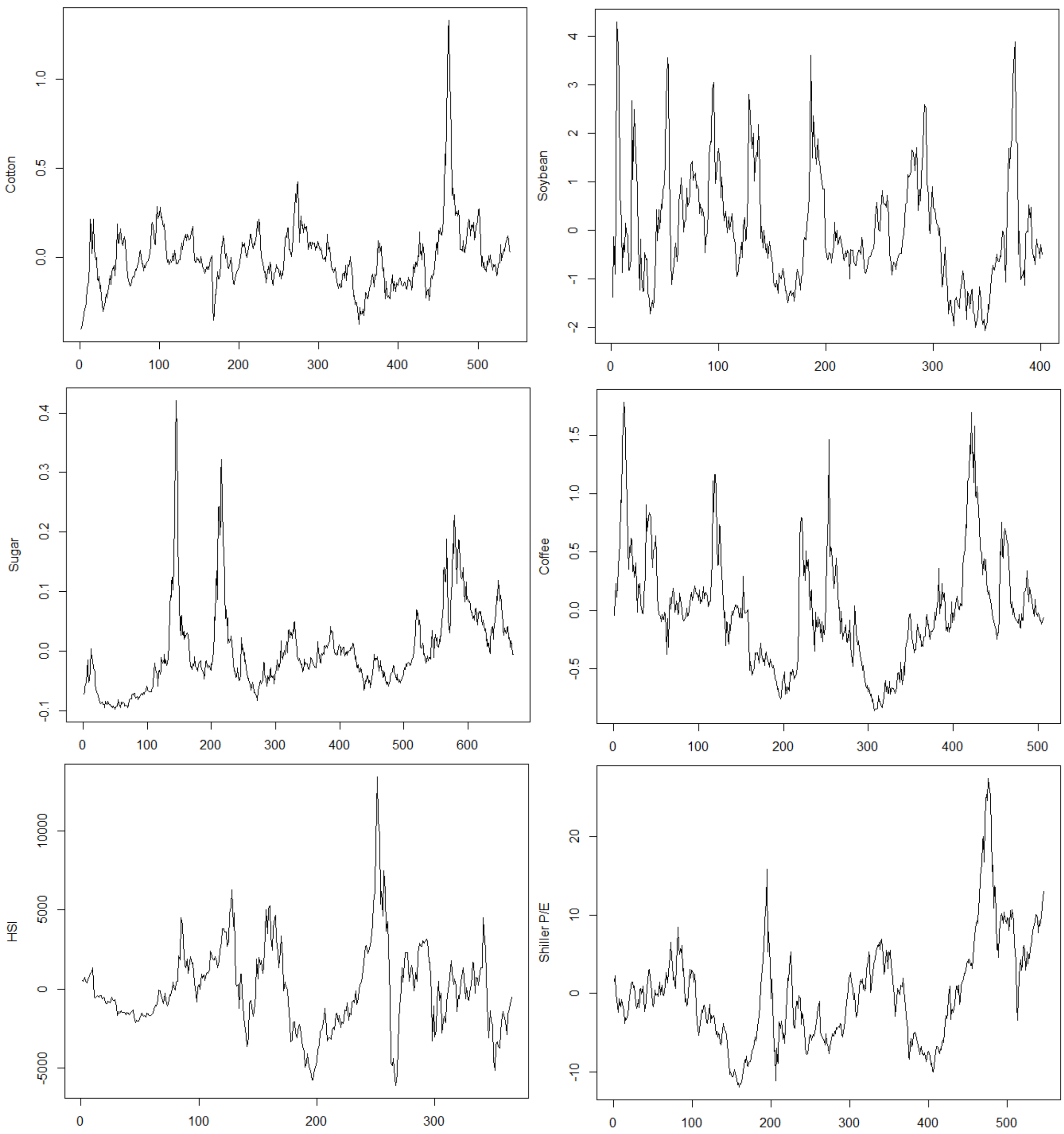


Figure 2.4: Financial series paths: cotton (8/1972 to 7/2017), soybean (1/1973 to 5/2006), sugar (11/1962 to 8/2018), coffee (4/1976 to 5/2018), Hang Seng index (HSI, 11/1986 to 3/2017) and Shiller's P/E ratio (Q1/1881 to Q2/2017). All series are monthly except the latter which is quarterly. A linear trend is fitted and subtracted from the HSI series and the other series are centered.

follows. For a given series,

- Assign the total AR order r validated by the portmanteau test (see Table 2.4).
- For the assigned order r, select then the least excess clustering competing representation (see the

favoured specifications in Table 2.5).

The selection is reported in Table 2.6 which shows the causal and noncausal orders as well as the (inverted) roots of the corresponding polynomials. For all series, mixed models with non-trivial causal and non-causal components are favoured, which is compatible with the upward/downward trends displayed by most extreme events featured in the trajectories.

Total AR order		Cotton	Soybean	Sugar	Coffee	HSI	Shiller P/E
	Favoured specification	(0,1)	(0,1)	(1,0)	(0,1)	(0,1)	(1,0)
1	AUC Excess Clustering	23.3	6.50	28.2	9.53	7.75	15.7
	Median of competitors	29.7	7.03	41.7	13.6	16.5	51.9
	Favoured specification	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)
2	AUC Excess Clustering	8.25	3.9	22.3	5.35	5.75	8.54
	Median of competitors	16.1	6.10	29.1	8.97	9.33	26.4
	Favoured specification	(1,2)	(2,1)	(1,2)	(1,2)	(1,2)	(1,2)
3	AUC Excess Clustering	6.20	2.30	21.4	2.83	0.50	2.13
	Median of competitors	10.5	3.45	83.1	9.60	6.17	12.4
	Favoured specification	(1,3)	(3,1)	(3,1)	(1,3)	(1,3)	(1,3)
4	AUC Excess Clustering	5.83	2.42	6.67	3.00	0.50	1.63
	Median of competitors	10.1	3.38	14.8	5.06	5.66	6.25
	Favoured specification	(1,4)	(2,3)	(2,3)	(1,4)	(3,2)	(3,2)
5	AUC Excess Clustering	3.00	1.75	3.17	3.00	2.00	2.08
	Median of competitors	8.30	3.00	13.3	5.38	5.67	5.17
	Favoured specification	(1,5)	(2,4)	(2,4)	(3,3)	(1,5)	(2,4)
6	AUC Excess Clustering	3.00	1.40	4.82	2.50	0.50	1.33
	Median of competitors	8.30	3.61	17.3	6.30	4.20	4.67

Table 2.5: Selection based on extreme clustering.

Series	Final specification	Noncausal (inverted) roots	Causal (inverted) roots
Cotton	MAR(1,1)	0.92	0.17
Soybean	MAR(2,3)	$0.37 \pm 0.40i$	$0.87,\ -0.60\ ,\ -0.07$
Sugar	MAR(3,1)	$0.94,\ -0.08 \pm 0.49$	0.52
Coffee	MAR(1,3)	0.55	$0.92,\ -0.29 \pm 0.21i$
HSI	MAR(1,2)	0.37	$0.89,\ -0.26$
Shiller P/E	MAR(2,4)	$0.58 \pm 0.29i$	$0.97,\ -0.70,\ -0.21 \pm 0.60i$

Table 2.6: Selection of the MAR specification for each financial series among the favoured ones of Table 2.5 based on the total AR order determined in Table 2.4. The MAR(p, q) specifications indicate the noncausal p and causal q orders as well as the (inverted) roots of the corresponding polynomials.

2.7 Concluding remarks

Noncausal models may provide better understanding of the dynamic features of a time series that are not perceived via causal models. Even the addition of a very simple noncausal component to an arbitrarily complex classical causal AR is sufficient to profoundly alter its motion and this in turn impacts the way we infer its future. Building on Gouriéroux and Zakoïan (2017) [63], we showed in this paper that several important properties of the pure noncausal AR(1) with stable errors extend to mixed AR models: the Markov property, the existence of a conditional mean whatever the size of the tails of the errors distribution, and the presence of ARCH effects in the Cauchy case. On the other hand, if the unit root property continues to hold when $\alpha = 1$, the martingale property is lost when a causal part is present in the model. A more complete description of the conditional distribution would require deriving higher-order moments, in particular to study the conditional skewness and kurtosis. We leave this issue for further research.

In the statistical part of the paper, we showed that LS estimation of a causal representation of the process allows to consistently identify the roots of the MAR polynomials, though not to distinguish causal and noncausal roots. Such identification issues were addressed by Hecq at al. (2016) [67] for MAR processes, and by Cavaliere et al. (2018) [24] using bootstrap inference for pure noncausal processes. We proposed an alternative strategy based on extreme clustering and leave its asymptotic properties for further investigations.

2.8 Postponed proofs

2.8.1 Proof of Proposition 2.2.1

Using the MA(∞) representation (2.3) of X_t and the assumption that $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha,\beta,\sigma,\mu)$, it follows that

$$\forall s \in \mathbb{R}, \quad \psi_{X_t}(s) := \mathbb{E}\left[e^{isX_t}\right] = \mathbb{E}\left[e^{is\sum\limits_{k=-\infty}^{+\infty} d_k\varepsilon_{t+k}}\right] = \prod_{k=-\infty}^{+\infty} \mathbb{E}\left[e^{isd_k\varepsilon_{t+k}}\right]$$

$$= \prod_{k=-\infty}^{+\infty} \exp\left\{-\sigma^\alpha |d_k s|^\alpha \left(1 - i\beta \mathrm{sign}(d_k s) w(\alpha, d_k s)\right) + i d_k s\mu\right\}.$$

If $\alpha \neq 1$, then,

$$\forall s \in \mathbb{R}, \quad \ln\psi_{X_t}(s) = \sum_{k=-\infty}^{+\infty} -\sigma^\alpha |d_k s|^\alpha \left(1 - i\beta\mathrm{sign}(d_k s)\mathrm{tg}\left(\frac{\pi\alpha}{2}\right)\right) + i d_k s\mu$$

$$= -\tilde{\sigma}^\alpha |s|^\alpha \left(1 - i\beta \frac{\sum\limits_{k=-\infty}^{+\infty} |d_k|^\alpha \mathrm{sign}(d_k)}{\sum\limits_{k=-\infty}^{+\infty} |d_k|^\alpha} \mathrm{sign}(s)\mathrm{tg}\left(\frac{\pi\alpha}{2}\right)\right) + is\mu \sum_{k=-\infty}^{+\infty} d_k.$$

Whereas if $\alpha = 1$, then, for any $s \in \mathbb{R}$

$$
\begin{aligned}
\ln \psi_{X_t}(s) &= \sum_{k=-\infty}^{+\infty} -\sigma|d_k s| \left(1 + i\frac{2}{\pi}\beta \operatorname{sign}(d_k s) \ln |d_k s|\right) + i d_k s \mu \\
&= -|s|\sigma \left(\sum_{k=-\infty}^{+\infty} |d_k|\right) \left(1 + i\frac{2}{\pi}\beta \frac{\sum_{k=-\infty}^{+\infty} d_k}{\sum_{k=-\infty}^{+\infty} |d_k|} \operatorname{sign}(s) \ln |s|\right) + is \left(\sum_{k=-\infty}^{+\infty} d_k \mu - \frac{2}{\pi}\sigma\beta \sum_{k=-\infty}^{+\infty} d_k \ln |d_k|\right).
\end{aligned}
$$

The conclusion follows from the characterization of stable laws in (2.2).

2.8.2 Proof of Proposition 2.3.1

We decompose (X_t) into its pure causal AR(q) and noncausal AR(p) components (see [59, 87]) respectively (v_t) and (u_t), defined by

$$
u_t = \phi(B) X_t \iff \psi(F) u_t = \varepsilon_t, \tag{2.31}
$$

$$
v_t = \psi(F) X_t \iff \phi(B) v_t = \varepsilon_t. \tag{2.32}
$$

We first show that (u_t) is a Markov process of order p. When there is no risk of ambiguity we denote by f a generic density, whose definition can change along the proof. Since by (2.31), $u_t = \psi_1 u_{t+1} + \ldots + \psi_p u_{t+p} + \varepsilon_t$, for $k > p$, the conditional density of u_t given its k past values is

$$
\begin{aligned}
f\left(u_t \middle| u_{t-1}, \ldots, u_{t-k}\right) &= \frac{f\left(u_t, \ldots, u_{t-k}\right)}{f\left(u_{t-1}, \ldots, u_{t-k}\right)} \\
&= \frac{f\left(u_{t-k} \middle| u_{t-k+1}, \ldots, u_{t-k+p}\right) f\left(u_t, \ldots, u_{t+1-k}\right)}{f\left(u_{t-k} \middle| u_{t-k+1}, \ldots, u_{t-k+p}\right) f\left(u_{t-1}, \ldots, u_{t+1-k}\right)} \\
&= \frac{f\left(u_t, \ldots, u_{t-p}\right)}{f\left(u_{t-1}, \ldots, u_{t-p}\right)} = f\left(u_t \middle| u_{t-1}, \ldots, u_{t-p}\right),
\end{aligned}
$$

where the second equality follows from the Bayes formula and (2.31), that is $u_t = \phi(B) X_t$, and the third equality is obtained by decreasing induction on k. We now turn to the MAR process (X_t). From $u_t = \phi(B) X_t$, we have $X_t = \sum_{i=1}^q \phi_i X_{t-i} + u_t$. Thus, with obvious notation, for any $x, x_1, \ldots, x_{p+q} \in \mathbb{R}$,

$$
\begin{aligned}
&f_{X_t}\left(x \middle| X_{t-1} = x_1, X_{t-2} = x_2, \ldots\right) \\
&= f_{u_t + \sum_{i=1}^q \phi_i x_i}\left(x \middle| X_{t-1} = x_1, \ldots\right) \\
&= f_{u_t}\left(x - \sum_{i=1}^q \phi_i x_i \middle| u_{t-1} = x_1 - \sum_{i=1}^q \phi_i x_{1+i}, u_{t-2} = x_2 - \sum_{i=1}^q \phi_i x_{2+i}, \ldots\right) \\
&= f_{u_t}\left(x - \sum_{i=1}^q \phi_i x_i \middle| u_{t-1} = x_1 - \sum_{i=1}^q \phi_i x_{1+i}, \ldots, u_{t-p} = x_p - \sum_{i=1}^q \phi_i x_{p+i}\right)
\end{aligned}
$$

using the Markov property of (u_t). The latter quantity is a function of $(x, x_1, \ldots, x_{p+q})$, showing that process (X_t) is Markov of order $p + q$.

2.8.3 Proof of Theorem 2.3.1

We first show that the theorem holds for $q = 0$ and we then extend it to general MAR(p, q) processes.

Lemma 2.8.1 *Let (X_t) be an α-stable pure noncausal AR(p) process solution of $X_t = \psi_1 X_{t+1} + \ldots + \psi_p X_{t+p} + \varepsilon_t$, where the roots of $\psi(z)$ are outside the unit circle. Then,*

$$\mathbb{E}\Big[|X_t|^\gamma \Big| X_{t-1}, \ldots, X_{t-p}\Big] < +\infty, \quad \textit{for any} \quad \gamma \in (0, 2\alpha + 1).$$

Proof. Suppose $p > 1$ (the result is already known from GZ for $p = 1$). For any $(x_0, \ldots, x_p) \in \mathbb{R}^{p+1}$,

$$f_{X_t|(X_{t+1},\ldots,X_{t+p})=(x_1,\ldots,x_p)}(x_0) = f_\varepsilon(x_0 - \psi_1 x_1 - \ldots - \psi_p x_p),$$

because ε_t is independent from $X_{t+1}, \ldots, X_{t+p}$. By the Bayes formula,

$$f_{X_t|(X_{t+1},\ldots,X_{t+p})=(x_1,\ldots,x_p)}(x_0) = \frac{f_{X_{t+p}|(X_t,\ldots,X_{t+p-1})=(x_0,\ldots,x_{p-1})}(x_p)}{f_{X_{t+1},\ldots,X_{t+p}}(x_1,\ldots,x_p)} f_{X_t,\ldots,X_{t+p-1}}(x_0,\ldots,x_{p-1}).$$

Thus,

$$f_{X_{t+p}|(X_t,\ldots,X_{t+p-1})=(x_0,\ldots,x_{p-1})}(x_p) = \frac{f_\varepsilon(x_0 - \psi_1 x_1 - \ldots - \psi_p x_p) f_X(x_p) f_{X_{t+1},\ldots,X_{t+p-1}|X_{t+p}=x_p}(x_1,\ldots,x_{p-1})}{f_{X_t,\ldots,X_{t+p-1}}(x_0,\ldots,x_{p-1})}. \tag{2.33}$$

On the one hand, when $x_p \to \pm\infty$,

$$f_X(x_p) \sim C(x_p)|x_p|^{-\alpha-1}, \tag{2.34}$$

$$f_\varepsilon(x_0 - \psi_1 x_1 - \ldots - \psi_p x_p) \sim C^*(x_p)|x_p|^{-\alpha-1}, \tag{2.35}$$

where $C(x_p)$ and $C^*(x_p)$ are constants depending on x_p, which may change according to whether $x_p \to +\infty$ or $x \to -\infty$. On the other hand, we show that

$$f_{X_{t+1},\ldots,X_{t+p-1}|X_{t+p}=x_p}(x_1,\ldots,x_{p-1}) \underset{|x_p|\to+\infty}{\longrightarrow} 0. \tag{2.36}$$

Let $Z_t = X_t - \psi_1 X_{t+1} - \cdots - \psi_{p-1} X_{t+p-1}$. Conditionally on $X_{t+p} = x_p$, we have $Z_t = \psi_p x_p + \varepsilon_t$. Since X_{t+p} and ε_t are independent and $\psi_p \neq 0$, we have $|Z_t| \to +\infty$ a.s. as $|x_p| \to +\infty$. Therefore, for any $z_0 \in \mathbb{R}$ and any neighbourhood V_{z_0} of z_0, when $|x_p| \to +\infty$,

$$\mathbb{P}\Big(Z_t \in V_{z_0} \Big| X_{t+p} = x_p\Big) \longrightarrow 0, \quad \text{which implies,} \quad \mathbb{P}\Big((X_t, \ldots, X_{t+p-1}) \in V_{\boldsymbol{x}} \Big| X_{t+p} = x_p\Big) \longrightarrow 0,$$

for any point $\boldsymbol{x} \in \mathbb{R}^p$ and neighbourhood $V_{\boldsymbol{x}}$ around this point. Hence the convergence in Equation (2.36). Combining Equations (2.33), (2.34), (2.35) and (2.36), we obtain, for $|x_p|$ large enough,

$$f_{X_{t+p}|(X_t,\ldots,X_{t+p-1})=(x_0,\ldots,x_{p-1})}(x_p) = o\Big(|x_p|^{-2(\alpha+1)}\Big).$$

Thus Lemma 2.8.1 is established.

Let us now prove Theorem 2.3.1. Let $\gamma \in (0, 2\alpha+1)$. Decomposing (X_t) into its pure causal and noncausal components (v_t) and (u_t), defined in (2.32) and (2.31), we have the equivalence between the information sets

$$(X_{t-1}, \dots, X_{t-p-q}) \quad \text{and} \quad (u_{t-1}, \dots, u_{t-p}, v_{t-p-1}, \dots, v_{t-p-q}),$$

and the independence between $(u_{t-1}, \dots, u_{t-p})$ and $(v_{t-p-1}, \dots, v_{t-p-q})$ (see Lanne and Saikkonen (2011), Gouriéroux and Jasiak (2016)). From Equation (2.31), we have for $\gamma \geq 1$ by the triangle inequality,

$$\begin{aligned}
&\Big(\mathbb{E}\Big[|X_t|^\gamma \Big| X_{t-1}, \dots, X_{t-p-q}\Big]\Big)^{1/\gamma} \qquad (2.37)\\
&= \Big(\mathbb{E}\Big[|u_t - \phi_1 X_{t-1} - \dots - \phi_q X_{t-q}|^\gamma \Big| X_{t-1}, \dots, X_{t-p-q}\Big]\Big)^{1/\gamma}\\
&\leq |\phi_1 X_{t-1} + \dots + \phi_q X_{t-q}| + \Big(\mathbb{E}\Big[|u_t|^\gamma \Big| X_{t-1}, \dots, X_{t-p-q}\Big]\Big)^{1/\gamma}\\
&= |\phi_1 X_{t-1} + \dots + \phi_q X_{t-q}| + \Big(\mathbb{E}\Big[|u_t|^\gamma \Big| u_{t-1}, \dots, u_{t-p}, v_{t-p-1}, \dots, v_{t-p-q}\Big]\Big)^{1/\gamma}\\
&= |\phi_1 X_{t-1} + \dots + \phi_q X_{t-q}| + \Big(\mathbb{E}\Big[|u_t|^\gamma \Big| u_{t-1}, \dots, u_{t-p}\Big]\Big)^{1/\gamma}, \qquad (2.38)
\end{aligned}$$

which is finite almost surely by Lemma 2.8.1 since (u_t) is an α-stable pure noncausal AR(p) process. If $\gamma \in (0,1)$, by the inequality $(a+b)^\gamma \leq a^\gamma + b^\gamma$ for any $a, b \geq 0$, we have that $|a+b|^\gamma \leq \Big(|a|+|b|\Big)^\gamma \leq |a|^\gamma + |b|^\gamma$, for any $(a,b) \in \mathbb{R}$. Thus, similarly to (2.38), we show that

$$\mathbb{E}\Big[|X_t|^\gamma \Big| X_{t-1}, \dots, X_{t-p-q}\Big] \leq |\phi_1 X_{t-1} + \dots + \phi_q X_{t-q}|^\gamma + \mathbb{E}\Big[|u_t|^\gamma \Big| u_{t-1}, \dots, u_{t-p}\Big],$$

which completes the proof.

2.8.4 Proof of Proposition 2.3.2

The following Lemma will be useful for the proofs of Proposition 2.3.2 and Corollary 2.3.1.

Lemma 2.8.2 *Let (X_t) be a MAR(p,q) process with $q > 0$. For any $h \geq 0$, there exist polynomials P_h and Q_h with $d^\circ(P_h) = q-1$ and $d^\circ(Q_h) = h$, such that for any $t \in \mathbb{Z}$,*

$$X_{t+h} = P_h(B) X_{t-1} + Q_h(F) u_t, \qquad (2.39)$$

where (u_t) is defined in (2.31).

Proof. We prove (2.39) by induction on h. In view of (2.31) we have $X_t = \sum_{i=1}^q \phi_i X_{t-i} + u_t$ from. Thus (2.39) holds for $h = 0$, with $P_0(B) = \sum_{i=0}^{q-1} \phi_{i+1} B^i$ and $Q_0(F) = I$. Assume that the property holds up to the order $h-1$, for $h \geq 1$. For $r = \min(h,q)$, $X_{t+h} = \sum_{i=1}^r \phi_{i+1} X_{t+h-i} + \sum_{i=r+1}^q \phi_{i+1} X_{t+h-i} + u_{t+h}$ where, by convention, the second sum vanishes if $r = q$. Thus

$$X_{t+h} = \sum_{i=1}^r \phi_{i+1} P_{h-i}(B) X_{t-1} + \sum_{i=r+1}^q \phi_{i+1} X_{t+h-i} + u_{t+h} + \sum_{i=1}^r Q_{h-i}(F) u_t,$$

which is of the form (2.39) with

$$P_h(B) = \sum_{i=1}^r \phi_{i+1} P_{h-i}(B) + \sum_{i=r+1}^q \phi_{i+1} B^{i-h-1}, \quad Q_h(B) = F^h + \sum_{i=1}^r Q_{h-i}(F). \qquad (2.40)$$

Therefore, (2.39) is established.

We now extend Theorem 2.3.1 in the case $p = 1$, by showing that for any $h \geq 0$, $\mathbb{E}\Big[|X_{t+h}|^\gamma \Big| X_{t-1}, \dots, X_{t-q-1}\Big] < +\infty$ whenever $0 < \gamma < 2\alpha + 1$. By Lemma 2.8.2 we have, proceeding as for Equation (2.38) and letting $Q_h(z) = \sum_{i=0}^h q_{i,h} z^i$,

$$\Big(\mathbb{E}\Big[|X_{t+h}|^\gamma \Big| X_{t-1}, \dots, X_{t-q-1}\Big]\Big)^{1/\gamma} \leq |P_h(B)X_{t-1}| + \sum_{i=0}^{h} |q_{i,h}| \Big(\mathbb{E}\Big[|u_{t+h}|^\gamma \Big| u_{t-1}\Big]\Big)^{1/\gamma},$$

which is finite almost surely for any $h \geq 0$ whenever $1 \leq \gamma < 2\alpha + 1$ by GZ (Proposition 3.2) since (u_t) is a noncausal AR(1). For $\gamma \in (0, 1)$, we proceed similarly using the inequality $|a + b|^\gamma \leq |a|^\gamma + |b|^\gamma$, for any $(a, b) \in \mathbb{R}$. We now turn to the conditional expectation of X_{t+h}. We have by the independence between u_{t-1} and $(v_{t-2}, \dots, v_{t-q-1})$

$$\begin{aligned}
\mathbb{E}\Big[X_{t+h} \Big| X_{t-1}, \dots, X_{t-q-1}\Big] &= P_h(B)X_{t-1} + \sum_{i=0}^{h} q_{i,h} \mathbb{E}\Big[u_{t+i} \Big| X_{t-1}, \dots, X_{t-q-1}\Big] \\
&= P_h(B)X_{t-1} + \sum_{i=0}^{h} q_{i,h} \mathbb{E}\Big[u_{t+i} \Big| u_{t-1}, v_{t-2}, \dots, v_{t-q-1}\Big] \\
&= P_h(B)X_{t-1} + \sum_{i=0}^{h} q_{i,h} \mathbb{E}\Big[u_{t+i} \Big| u_{t-1}\Big]. \qquad (2.41)
\end{aligned}$$

By GZ (Proposition 3.3), we have for any $i \geq 0$,

$$\mathbb{E}\Big[u_{t+i} \Big| u_{t-1}\Big] = \Big(\psi^{<\alpha-1>}\Big)^{i+1} u_{t-1},$$

and therefore,

$$\begin{aligned}
\mathbb{E}\Big[X_{t+h} \Big| X_{t-1}, \dots, X_{t-q-1}\Big] &= P_h(B)X_{t-1} + \psi^{<\alpha-1>} u_{t-1} \sum_{i=0}^{h} q_{i,h} \Big(\psi^{<\alpha-1>}\Big)^i \\
&= \Big(P_h(B) + \psi^{<\alpha-1>} Q_h(\psi^{<\alpha-1>})\phi(B)\Big) X_{t-1} \\
&:= \mathscr{P}_h(B) X_{t-1}.
\end{aligned}$$

To conclude, we invoke the fact that (X_t) is a Markov chain of order $q + 1$, which gives the equality $\mathbb{E}\Big[X_{t+h} \Big| X_{t-1}, \dots, X_{t-q-1}\Big] = \mathbb{E}\Big[X_{t+h} \Big| \mathscr{F}_{t-1}\Big]$. The formula for $h = 0$ is obtained by noting that $P_0(B) = \sum_{i=0}^{q-1} \phi_{i+1} B^i$ and $Q_0(F) = I$.

2.8.5 Proof of Corollary 2.3.1

We will derive the asymptotic behaviour of $\mathscr{P}_h(B)X_{t-1} = \Big(P_h(B) + \psi^{<\alpha-1>} Q_h(\psi^{<\alpha-1>})\phi(B)\Big) X_{t-1}$ when $1 \leq \alpha < 2$. We start by a result giving details about the behaviours of the coefficients of the polynomials P_h and Q_h defined in Lemma 2.8.2. Denote $P_h(z) := \sum_{i=0}^{q-1} a_{i,h} z^i$ and $Q_h(z) := \sum_{i=0}^{h} b_{i,h} z^i$.

Lemma 2.8.3 *For $h \geq q$, the coefficients of polynomial P_h and Q_h verify:*

$$a_{0,h} = C_1(h)\lambda_1^h + \ldots + C_s(h)\lambda_s^h, \qquad a_{i,h} = \sum_{j=0}^{q-i-1} a_{0,h-j-1}\phi_{i+1+j}, \quad \textit{for} \quad 0 \leq i \leq q-1,$$

$$b_{i,h} = a_{0,h-i-1}, \quad \textit{for} \quad 0 \leq i \leq h, \qquad a_{0,-1} := 1,$$

where the $\lambda_1, \ldots, \lambda_s$ are the distinct (inverse of the) roots with multiplicities $m_1, \ldots, m_s$ of ϕ and $C_1, \ldots, C_s$ are polynomials with degrees $m_1 - 1, \ldots, m_s - 1$.

The proof is relegated to Appendix.

The proof of Corollary 2.3.1 involves several steps.

i) Equivalent of $a_{0,h}$

Without loss of generality we can assume that the (inverses of the) roots of $\phi(z)$ are ordered: $0 < |\lambda_s| < \cdots < |\lambda_1| < 1$. For ease of notation, we drop the indexes of the largest root (in modulus) λ_1 and m_1 and we will denote also by C the coefficient associated to the monomial of highest degree of C_1. We thus have

$$a_{0,h} \underset{h \to +\infty}{\sim} Ch^{m-1}\lambda^h, \quad \text{and} \quad |a_{0,h}| \underset{h \to +\infty}{\longrightarrow} 0. \tag{2.42}$$

ii) Limit of $P_h(B)X_{t-1}$

From Lemma 2.8.3, it appears that $P_h(B)X_{t-1} = \sum_{i=0}^{q-1} a_{i,h}X_{t-i-1} \underset{h \to +\infty}{\overset{a.s.}{\longrightarrow}} 0.$

iii) Limit of $Q_h(\psi^{<\alpha-1>})$

$$Q_h(\psi^{<\alpha-1>}) = \sum_{i=0}^{h} a_{0,h-i-1}(\psi^{<\alpha-1>})^i = \left(\psi^{<\alpha-1>}\right)^{h-1}\left[\psi^{<\alpha-1>} + \sum_{i=0}^{h-1} a_{0,i}(\psi^{<1-\alpha>})^i\right].$$

Let us study the general term of the above series. We have

$$a_{0,i}(\psi^{<1-\alpha>})^i \underset{i \to +\infty}{\sim} Ci^{m-1}\lambda^i(\psi^{<1-\alpha>})^i = C\text{sign}(\lambda\psi)^i i^{m-1}(|\lambda||\psi|^{1-\alpha})^i. \tag{2.43}$$

Different cases arise.

ι) Assume $\alpha = 1$. According to Equation (2.43) for $\alpha = 1$, $|a_{0,i}(\psi^{<1-\alpha>})^i| \sim |C|i^{m-1}|\lambda|^i$ which is the general term of an absolutely convergent series. Thus, $|Q_h(\psi^{<\alpha-1>})| = |Q_h(\text{sign}(\psi))| = \left|\text{sign}(\psi) + \sum_{i=0}^{h-1} a_{0,i}\text{sign}(\psi)^i\right| \underset{i \to +\infty}{\longrightarrow} D$, for some $D \geq 0$.

$\iota\iota$) Assume $1 < \alpha < 1 + \dfrac{\ln|\lambda|}{\ln|\psi|}$.

Then $|Q_h(\psi^{<\alpha-1>})| = |\psi|^{(\alpha-1)(h-1)}\left|\psi^{<\alpha-1>} + \sum_{i=0}^{h-1} a_{0,i}(\psi^{<1-\alpha>})^i\right| \underset{i \to +\infty}{\longrightarrow} 0 \cdot D = 0.$

$\iota\iota\iota$) Assume $\alpha = 1 + \dfrac{\ln|\lambda|}{\ln|\psi|}$.

For $i \geq q$, there exists a positive constant A such that

$$|a_{0,i}| = \left|\sum_{j=1}^{q} C_j(i)\lambda_j^i\right| \leq Ai^m|\lambda|^i. \tag{2.44}$$

Thus, since $|\lambda||\psi|^{1-\alpha} = 1$,

$$
\begin{aligned}
|\psi|^{(\alpha-1)(h-1)} \left| \sum_{i=0}^{h-1} a_{0,i} (\psi^{<1-\alpha>})^i \right| &\leq A|\psi|^{(\alpha-1)(h-1)} \sum_{i=0}^{h-1} i^m |\lambda|^i |\psi|^{(1-\alpha)i} \\
&\leq A|\psi|^{(\alpha-1)(h-1)} h^{m+1} \underset{h\to+\infty}{\longrightarrow} 0.
\end{aligned}
$$

ν) Assume $\alpha > 1 + \dfrac{\ln|\lambda|}{\ln|\psi|}$. From Equation (2.44),

$$
\begin{aligned}
|\psi|^{(\alpha-1)(h-1)} \left| \sum_{i=0}^{h-1} a_{0,i} (\psi^{<1-\alpha>})^i \right| &\leq A|\psi|^{(\alpha-1)(h-1)} \sum_{i=0}^{h-1} i^m |\lambda|^i |\psi|^{(1-\alpha)i} \\
&\leq A|\psi|^{(\alpha-1)(h-1)} h^m \frac{1 - |\lambda|^h |\psi|^{(1-\alpha)h}}{1 - |\lambda||\psi|^{1-\alpha}} \\
&\leq \frac{A h^m |\psi|^{1-\alpha}}{1 - |\lambda||\psi|^{1-\alpha}} \left(|\psi|^{(\alpha-1)h} - |\lambda|^h \right) \qquad \underset{h\to+\infty}{\longrightarrow} 0.
\end{aligned}
$$

The proof of the diverging conditional expectation in the MAR$(1, q)$ case with $\alpha \in (0, 1)$ is provided in Appendix.

2.8.6 Proof of Proposition 2.4.1

The $\rho(h)$'s are only function of the AR coefficients and coincide with the theoretical autocorrelations of the process $\sum_{k=-\infty}^{\infty} d_k Z_{t-k}$, where (Z_t) is an i.i.d. noise (with finite variance). Thus, the $\rho(h)$'s are the theoretical autocorrelations of the stationary solution (Y_t) of the AR model $\psi_0(F)\phi_0(B)Y_t = Z_t$. We know from Brockwell and Davis (1991, Proposition 3.5.1) [19] that (Y_t) satisfies the causal AR model $\psi_0(B)\phi_0(B)Y_t = Z_t^*$, for some white noise sequence (Z_t^*), from which the recursion on the coefficients $\rho(h)$ is deduced. The conclusion follows.

2.8.7 Proof of Proposition 2.4.2

The consistency of $\hat{\boldsymbol{\eta}}$ follows from Davis and Resnick (1986, Section 5.4) [41].

2.8.8 Proof of Proposition 2.4.3

Let $\hat{\boldsymbol{\rho}} = [\hat{\rho}(i)]_{i=1,\dots,p+q}, \hat{\boldsymbol{R}} = [\hat{\rho}(i-j)]_{i,j=1,\dots,p+q}$. In view of (2.11) and (2.14), we have $\hat{\boldsymbol{\eta}} = \hat{\boldsymbol{R}}^{-1} \hat{\boldsymbol{\rho}}$ and $\boldsymbol{\eta}_0 = \boldsymbol{R}^{-1}\boldsymbol{\rho}$. We have

$$
\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0 \quad = \quad \hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}) + (\hat{\boldsymbol{R}}^{-1} - \boldsymbol{R}^{-1})\boldsymbol{\rho} = \hat{\boldsymbol{R}}^{-1} \left\{ (\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}) + (\boldsymbol{R} - \hat{\boldsymbol{R}})\boldsymbol{R}^{-1}\boldsymbol{\rho} \right\}. \tag{2.45}
$$

We have $\boldsymbol{R} - \hat{\boldsymbol{R}} = \sum_{i=1}^{p+q} \{\rho(i) - \hat{\rho}(i)\} \boldsymbol{K}^{(i)}$. It follows that

$$
(\boldsymbol{R} - \hat{\boldsymbol{R}})\boldsymbol{R}^{-1}\boldsymbol{\rho} = -\boldsymbol{L}(\boldsymbol{I}_{p+q} \otimes \boldsymbol{R}^{-1}\boldsymbol{\rho})(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}). \tag{2.46}
$$

Thus, since $\hat{\boldsymbol{R}}^{-1} \to \boldsymbol{R}^{-1}$ in probability as $n \to \infty$, $\frac{a_n^2}{\tilde{a}_n}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$ has the same asymptotic distribution as $\boldsymbol{R}^{-1}\{\boldsymbol{I}_{p+q} - \boldsymbol{L}(\boldsymbol{I}_{p+q} \otimes \boldsymbol{R}^{-1}\boldsymbol{\rho})\} \frac{a_n^2}{\tilde{a}_n}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho})$. The convergence in distribution in (2.16) is a direct consequence of Davis and Resnick (1986) [41] who showed that $\frac{a_n^2}{\tilde{a}_n}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}) \xrightarrow{d} \mathbf{Z}$.

2.8.9 Proof of Proposition 2.4.4

The case $0 < \alpha < 1$ is a direct consequence of Theorem 4.4 (i) by Davis and Resnick (1986) [41].
Consider the case $1 < \alpha < 2$. From the proof of Corollary 1 p. 553 by Davis and Resnick (1986) [41], we know that $\tilde{\rho}(h) - \hat{\rho}(h) = o_p(\tilde{a}_n a_n^{-2})$ for $h \geq 1$. Given $\hat{\rho}(h) \xrightarrow{p} \rho(h)$, it holds that $\tilde{\rho}(h) \xrightarrow{p} \rho(h)$ for $h \geq 1$. Following the proof of Proposition 2.4.3 with obvious notations, it can then be shown that $\tilde{a}_n^{-1} a_n^2 (\tilde{\eta} - \eta_0)$ has the same asymptotic distribution as $\boldsymbol{R}^{-1}\{\boldsymbol{I}_{p+q} - \boldsymbol{L}(\boldsymbol{I}_{p+q} \otimes \boldsymbol{R}^{-1}\boldsymbol{\rho})\}\frac{a_n^2}{\tilde{a}_n}(\tilde{\boldsymbol{\rho}} - \boldsymbol{\rho})$. The conclusion follows from Corollary 1 by Davis and Resnick (1986) [41].

2.8.10 Proof of Proposition 2.4.5

Write, for $t = p + q + 1, \dots, n$,

$$\hat{\zeta}_t = -\sum_{i=0}^{p+q} \eta_{0i} X_{t-i} - \sum_{i=1}^{p+q} (\hat{\eta}_i - \eta_{0i}) X_{t-i} = -\sum_{i=0}^{p+q} \eta_{0i} X_{t-i} - (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)' \boldsymbol{X}_{t-1},$$

with $\eta_{00} = -1$ and $\boldsymbol{X}_{t-1} = (X_{t-1}, \dots, X_{t-p-q})'$. Hence

$$\begin{aligned} \tilde{a}_n^{-1} a_n^2 \hat{\rho}_{\hat{\zeta}}(h) \quad &= \quad \frac{\tilde{a}_n^{-1} a_n^2}{\hat{\gamma}_{\hat{\zeta}}(0)} \sum_{t=p+q+1}^{n} \left\{ \sum_{i,j=0}^{p+q} \eta_{0i}\eta_{0j} X_{t-i} X_{t-h-j} \right. \\ &\qquad \left. + (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)' \sum_{i=0}^{p+q} \eta_{0i} (X_{t-i} \boldsymbol{X}_{t-h-1} + X_{t-h-i} \boldsymbol{X}_{t-1}) \right\} + o_P(1), \end{aligned}$$

with by convention $X_s = 0$ for $s \leq 0$. Let the $(p+q+1) \times (p+q+1)$ matrices $\hat{\boldsymbol{R}}_{\boldsymbol{h}} = [\hat{\rho}(h+i-j)]_{i,j=0,\dots,p+q}$, $\boldsymbol{R}_{\boldsymbol{h}} = [\rho(h+i-j)]_{i,j=0,\dots,p+q}$, and for any strictly positive integers, m, m' such that $m \leq m'$, let $\hat{\boldsymbol{\rho}}_{m:m'} = [\hat{\rho}(i)]_{i=m,\dots,m'}$ and $\boldsymbol{\rho}_{m:m'} = [\rho(i)]_{i=m,\dots,m'}$. Then,

$$\begin{aligned} \sum_{t=p+q+1}^{n} \sum_{i,j=0}^{p+q} \eta_{0i}\eta_{0j} X_{t-i} X_{t-h-j} \quad &= \quad \hat{\gamma}(0) \sum_{i,j=0}^{p+q} \eta_{0i}\eta_{0j} \hat{\rho}(h+j-i) + o_P(1) \\ &= \quad \hat{\gamma}(0) \sum_{i,j=0}^{p+q} \eta_{0i}\eta_{0j} \{\hat{\rho}(h+j-i) - \rho(h+j-i)\} + o_P(1) \\ &= \quad \hat{\gamma}(0) \boldsymbol{\eta}_0' \Big(\hat{\boldsymbol{R}}_{\boldsymbol{h}} - \boldsymbol{R}_{\boldsymbol{h}} \Big) \boldsymbol{\eta}_0 + o_P(1) \\ &= \quad \hat{\gamma}(0) \Big(\boldsymbol{\eta}_0' \otimes \boldsymbol{\eta}_0' \Big) \mathrm{vec} \Big(\hat{\boldsymbol{R}}_{\boldsymbol{h}} - \boldsymbol{R}_{\boldsymbol{h}} \Big) + o_P(1), \end{aligned}$$

where the second equality follows from (2.11). Moreover,

$$\begin{aligned} &\sum_{t=p+q+1}^{n} (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)' \sum_{i=0}^{p+q} \eta_{0i} (X_{t-i} \boldsymbol{X}_{t-h-1} + X_{t-h-i} \boldsymbol{X}_{t-1}) \\ &\qquad = \hat{\gamma}(0) \sum_{i=0}^{p+q} \sum_{j=1}^{p+q} (\hat{\eta}_{nj} - \eta_{0j}) \eta_{0i} \Big(\hat{\rho}(h+j-i) + \hat{\rho}(h+i-j) \Big) + o_P(1) \\ &\qquad = \hat{\gamma}(0) \boldsymbol{\eta}_0' \Big(\hat{\boldsymbol{R}}_{\boldsymbol{h}} + \hat{\boldsymbol{R}}_{\boldsymbol{h}}' \Big) (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) + o_P(1). \end{aligned}$$

Let the $(p+q+1)\times(p+q+1)$ matrices $\boldsymbol{D}_i = J^i$ and $\boldsymbol{D}_{-i} = {}^{\mathbf{t}}\boldsymbol{J}^{\boldsymbol{i}}$ for $i \geq 0$. We have:

$$\begin{aligned}
\hat{\boldsymbol{R}}_{\boldsymbol{h}} - \boldsymbol{R}_{\boldsymbol{h}} = \sum_{i=1}^{p+q-h} \Big(\hat{\rho}(i) - \rho(i)\Big)\Big(\boldsymbol{D}_{h-i} + \boldsymbol{D}_{h+i}\Big) & \\
+ \sum_{i=p+q-h+1}^{h+p+q} \Big(\hat{\rho}(i) - \rho(i)\Big)\boldsymbol{D}_{h-i}, & \qquad \text{if} \quad 1 \leq h \leq p+q-1, \\
\hat{\boldsymbol{R}}_{\boldsymbol{h}} - \boldsymbol{R}_{\boldsymbol{h}} = \sum_{i=h-p-q}^{h+p+q} \Big(\hat{\rho}(i) - \rho(i)\Big)\boldsymbol{D}_{h-i}, & \qquad \text{if} \quad h \geq p+q.
\end{aligned}$$

Thus, with

$$\begin{aligned}
\boldsymbol{L}_{\boldsymbol{h}} &= \Big[\mathrm{vec}(\boldsymbol{D}_{h-1} + \boldsymbol{D}_{h+1}) \ldots \mathrm{vec}(\boldsymbol{D}_{2h-p-q} + \boldsymbol{D}_{p+q}) \;\; \mathrm{vec}(\boldsymbol{D}_{2h-p-q-1}) \ldots \mathrm{vec}(\boldsymbol{D}_{-p-q})\Big], \quad \text{if} \quad 1 \leq h \leq p+q, \\
\boldsymbol{L}_{\boldsymbol{h}} &= \Big[\mathrm{vec}(\boldsymbol{D}_{p+q}) \ldots \mathrm{vec}(\boldsymbol{D}_{-p-q})\Big], \qquad \text{if} \quad h \geq p+q,
\end{aligned}$$

we can write

$$\begin{aligned}
\mathrm{vec}\Big(\hat{\boldsymbol{R}}_{\boldsymbol{h}} - \boldsymbol{R}_{\boldsymbol{h}}\Big) &= \boldsymbol{L}_{\boldsymbol{h}}\Big(\hat{\boldsymbol{\rho}}_{1:h+p+q} - \boldsymbol{\rho}_{1:h+p+q}\Big), && \text{if} \quad 1 \leq h \leq p+q, \\
\mathrm{vec}\Big(\hat{\boldsymbol{R}}_{\boldsymbol{h}} - \boldsymbol{R}_{\boldsymbol{h}}\Big) &= \boldsymbol{L}_{\boldsymbol{h}}\Big(\hat{\boldsymbol{\rho}}_{h-p-q:h+p+q} - \boldsymbol{\rho}_{h-p-q:h+p+q}\Big), && \text{if} \quad h \geq p+q+1.
\end{aligned}$$

The two last expressions point to the fact that $\Big(\hat{\rho}_{\hat{\zeta}}(h)\Big)_{h=1,\ldots,H}$ will depend on $\Big(\hat{\boldsymbol{\rho}}(i) - \boldsymbol{\rho}(i)\Big)_{i=1,\ldots,H+p+q}$. We therefore rewrite $\mathrm{vec}\Big(\hat{\boldsymbol{R}}_{\boldsymbol{h}} - \boldsymbol{R}_{\boldsymbol{h}}\Big)$ as

$$\mathrm{vec}\Big(\hat{\boldsymbol{R}}_{\boldsymbol{h}} - \boldsymbol{R}_{\boldsymbol{h}}\Big) = \boldsymbol{L}_{\boldsymbol{h}}\boldsymbol{M}_{\boldsymbol{h}}\Big(\hat{\boldsymbol{\rho}}_{1:H+p+q} - \boldsymbol{\rho}_{1:H+p+q}\Big),$$

with $\boldsymbol{M}_{\boldsymbol{h}}$ being the matrix of size $(h+p+q)\times(H+p+q)$ if $0 \leq h \leq p+q$ and $(2(p+q))\times(H+p+q)$ if $h \geq p+q+1$ picking the appropriate components of $\Big(\hat{\boldsymbol{\rho}}_{1:H+p+q} - \boldsymbol{\rho}_{1:H+p+q}\Big)$. More explicitly,

$$\begin{aligned}
\boldsymbol{M}_{\boldsymbol{h}} &= \Big(\boldsymbol{I}_{h+p+q} \;\; \mathbf{0}_{h+p+q\times H-h}\Big), && \text{if} \quad 0 \leq h \leq p+q, \\
\boldsymbol{M}_{\boldsymbol{h}} &= \Big(\mathbf{0}_{2(p+q)+1\times h-p-q-1} \;\; \boldsymbol{I}_{2(p+q)+1} \;\; \mathbf{0}_{2(p+q)+1\times H-h}\Big), && \text{if} \quad h \geq p+q+1.
\end{aligned}$$

Thus, using equations (2.45) and (2.46),

$$\tilde{a}_n^{-1}a_n^2\hat{\rho}_{\hat{\zeta}}(h) = \tilde{a}_n^{-1}a_n^2\frac{\hat{\gamma}(0)}{\hat{\gamma}_{\hat{\zeta}}(0)}\left[\Big(\boldsymbol{\eta}_0' \otimes \boldsymbol{\eta}_0'\Big)\boldsymbol{L}_{\boldsymbol{h}}\boldsymbol{M}_{\boldsymbol{h}} + \boldsymbol{\eta}_0'\Big(\hat{\boldsymbol{R}}_{\boldsymbol{h}} + \hat{\boldsymbol{R}}_{\boldsymbol{h}}'\Big)\hat{\boldsymbol{P}}\right]\Big(\hat{\boldsymbol{\rho}}_{1:H+p+q} - \boldsymbol{\rho}_{1:H+p+q}\Big) + o_P(1),$$

with $\hat{\boldsymbol{P}} := \begin{pmatrix} \mathbf{0}_{1\times p+q} \\ \boldsymbol{I}_{p+q} \end{pmatrix} \hat{\boldsymbol{R}}^{-1}\{\boldsymbol{I}_{p+q} - \boldsymbol{L}(\boldsymbol{I}_{p+q} \otimes \boldsymbol{R}^{-1}\boldsymbol{\rho})\}\boldsymbol{M}_{\mathbf{0}}$.

Finally, letting $\hat{\boldsymbol{A}}_{\boldsymbol{H}} = \left[\Big(\boldsymbol{\eta}_0' \otimes \boldsymbol{\eta}_0'\Big)\boldsymbol{L}_{\boldsymbol{h}}\boldsymbol{M}_{\boldsymbol{h}} + \boldsymbol{\eta}_0'\Big(\hat{\boldsymbol{R}}_{\boldsymbol{h}} + \hat{\boldsymbol{R}}_{\boldsymbol{h}}'\Big)\hat{\boldsymbol{P}}\right]_{h=1,\ldots,H}$ denote the matrix resulting from the vertical piling of vectors, we have

$$\frac{a_n^2}{\tilde{a}_n}\hat{\boldsymbol{\rho}}_{\hat{\zeta}} = \hat{\boldsymbol{A}}_{\boldsymbol{H}}\frac{a_n^{-2}\hat{\gamma}(0)}{a_n^{-2}\hat{\gamma}_{\hat{\zeta}}(0)}\tilde{a}_n^{-1}a_n^2\Big(\hat{\boldsymbol{\rho}}_{1:H+p+q} - \boldsymbol{\rho}_{1:H+p+q}\Big) + o_P(1).$$

By Theorem 4.2 by Davis and Resnick (1985) [40], Theorem 4.4 by Davis and Resnick (1986) [41] and Lemma 2.8.5 below, $\hat{\boldsymbol{P}} \xrightarrow{p} \boldsymbol{P} := \begin{pmatrix} \boldsymbol{0}_{1\times p+q} \\ \boldsymbol{I}_{p+q} \end{pmatrix} \boldsymbol{R}^{-1}\{\boldsymbol{I}_{p+q} - \boldsymbol{L}(\boldsymbol{I}_{p+q} \otimes \boldsymbol{R}^{-1}\boldsymbol{\rho})\}\boldsymbol{M_0}$,
$\hat{\boldsymbol{A}}_{\boldsymbol{H}} \xrightarrow{p} \left[\left(\boldsymbol{\eta}_0' \otimes \boldsymbol{\eta}_0'\right)\boldsymbol{L_h}\boldsymbol{M_h} + \boldsymbol{\eta}_0'\boldsymbol{R}_{\boldsymbol{h}}'\boldsymbol{P}\right]_{h=1,\dots,H} := \boldsymbol{A_H}$ and $\hat{\boldsymbol{\rho}}_{\hat{\zeta}} \xrightarrow{d} \gamma(0)\boldsymbol{A_H}\boldsymbol{Z}$ where $\boldsymbol{Z} = (Z_1, \dots, Z_{H+p+q})$, and where the (Z_i) are defined at Proposition 2.4.3.

Lemma 2.8.4 *Under the assumptions of Proposition 2.4.5,* $a_n^{-2}\left(\hat{\gamma}(h) - \gamma(h)\hat{\gamma}_\zeta(0)\right) \xrightarrow{p} 0.$

Lemma 2.8.5 *Under the assumptions of Proposition 2.4.5,* $a_n^{-2}\hat{\gamma}_{\hat{\zeta}}(0) = a_n^{-2}\dfrac{\hat{\gamma}(0)}{\gamma(0)} + o_P(1).$

2.8.11 Proof of Lemma 2.8.4

We have

$$\hat{\gamma}(h) = \sum_{t=1}^n X_t X_{t-h} = \sum_{t=1}^n \sum_{i\in\mathbb{Z}} \sum_{j\in\mathbb{Z}} d_i d_j \varepsilon_{t+i}\varepsilon_{t+j-h} = \sum_{t=1}^n \sum_{i\in\mathbb{Z}} \sum_{j\neq i} d_i d_{j+h}\varepsilon_{t+i}\varepsilon_{t+j} + \sum_{t=1}^n \sum_{i\in\mathbb{Z}} d_i d_{i+h}\varepsilon_{t+i}^2.$$

From Proposition 4.2 by Davis and Resnick (1986) [41], we have

$$a_n^{-2}\sum_{t=1}^n \sum_{i\in\mathbb{Z}} \sum_{j\neq i} d_i d_{j+h}\varepsilon_{t+i}\varepsilon_{t+j} \xrightarrow{p} 0. \tag{2.47}$$

A direct extension of Proposition 4.3.ii by Davis and Resnick (1986) [41] (see also the proof of Proposition 4.3 by GZ in the AR(1) case) yields

$$a_n^{-2}\left(\sum_{t=1}^n \sum_{i\in\mathbb{Z}} d_i d_{i+h}\varepsilon_{t+i}^2 - \gamma(h)\sum_{t=1}^n \varepsilon_t^2\right) \xrightarrow{p} 0. \tag{2.48}$$

Combining equations (2.47) and (2.48), we get $a_n^{-2}\left(\hat{\gamma}(h) - \gamma(h)\hat{\gamma}_\zeta(0)\right) \xrightarrow{p} 0.$

2.8.12 Proof of Lemma 2.8.5

$$\begin{aligned}
a_n^{-2}\sum_{t=1}^n \hat{\zeta}_t^2 &= a_n^{-2}\sum_{t=1}^n \left(X_t - \sum_{i=1}^{p+q}\boldsymbol{\eta}_{0i}X_{t-i} + \sum_{i=1}^{p+q}\left(\hat{\boldsymbol{\eta}}_i - \boldsymbol{\eta}_{0i}\right)X_{t-i}\right)^2 \\
&= a_n^{-2}\sum_{t=1}^n \Bigg[\left(X_t - \sum_{i=1}^{p+q}\boldsymbol{\eta}_{0i}X_{t-i}\right)^2 + 2\sum_{i=1}^{p+q}\left(\hat{\boldsymbol{\eta}}_i - \boldsymbol{\eta}_{0i}\right)\left(X_tX_{t-i} - \sum_{j=1}^{p+q}\boldsymbol{\eta}_{0j}X_{t-i}X_{t-j}\right) \\
&\qquad + \sum_{i=1}^{p+q}\sum_{i=1}^{p+q}\left(\hat{\boldsymbol{\eta}}_i - \boldsymbol{\eta}_{0i}\right)\left(\hat{\boldsymbol{\eta}}_j - \boldsymbol{\eta}_{0j}\right)X_{t-i}X_{t-j}\Bigg] \\
&= a_n^{-2}\Bigg[\hat{\gamma}(0) - \sum_{i=1}^{p+q}\boldsymbol{\eta}_{0i}\hat{\gamma}(-j) - \sum_{i=1}^{p+q}\boldsymbol{\eta}_{0i}\left(\hat{\gamma}(i) - \sum_{j=1}^{p+q}\boldsymbol{\eta}_{0j}\hat{\gamma}(i-j)\right) \\
&\qquad + \sum_{i=1}^{p+q}\sum_{i=1}^{p+q}\left(\hat{\boldsymbol{\eta}}_i - \boldsymbol{\eta}_{0i}\right)\left(\hat{\boldsymbol{\eta}}_j - \boldsymbol{\eta}_{0j}\right)\hat{\gamma}(i-j)\Bigg].
\end{aligned}$$

Using Lemma 2.8.4, the fact that $\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0 \longrightarrow 0$ in probability and the convergence in distribution of the vector $a_n^{-2}\Big(\hat{\gamma}(i), \quad 0 \leq i \leq L\Big)$ for any integer L, we get:

$$a_n^{-2}\hat{\gamma}_{\hat{\zeta}}(0) = a_n^{-2}\hat{\gamma}_{\zeta}(0)\left[\gamma(0) - \sum_{i=1}^{p+q} \boldsymbol{\eta}_{0i}\gamma(-j) - \sum_{i=1}^{p+q} \boldsymbol{\eta}_{0i}\left(\gamma(i) - \sum_{j=1}^{p+q} \boldsymbol{\eta}_{0j}\gamma(i-j)\right)\right] + o_P(1).$$

From Proposition 2.4.1, we have that $\boldsymbol{\eta}_0(B)\gamma(i) = 0$ for any $i \geq 1$ and $\boldsymbol{\eta}_0(B)\gamma(0) = 1$. Thus

$$a_n^{-2}\hat{\gamma}_{\hat{\zeta}}(0) = a_n^{-2}\hat{\gamma}_{\zeta}(0) + o_P(1) = a_n^{-2}\frac{\hat{\gamma}(0)}{\gamma(0)} + o_P(1).$$

2.9 Appendix

This Appendix consists of six sections of additional results: 2.9.1) asymptotic prediction of the MAR$(1, q)$ when $\alpha \in (0,1)$ and an explicit example in the MAR$(1,1)$ case; 2.9.2) expectation of MAR(p,q) processes conditionally on a linear combination of past values and proof of the unit root property; 2.9.3) conditional correlation structure of noncausal AR(1) processes, proofs of Proposition 2.3.3 and of the conditional variance of the MAR(1,1); 2.9.4) proof of Lemma 2.8.3; 2.9.5) recursion over polynomials P_h and Q_h; 2.9.6) Cluster size distribution, an illustration with the noncausal AR(1); 2.9.7) complementary results on the empirical study and details about the estimation of excess clustering term structures; 2.9.9) Complementary estimation of the financial series using the R package 'MARX'.

2.9.1 A complement to Corollary 2.3.1 in the case $\boldsymbol{\alpha \in (0,1)}$ and $q > 1$

Under the conditions of Proposition 2.3.2, when $\alpha \in (0,1)$, we have almost surely

$$\left|\mathbb{E}\left[\, X_{t+h} \middle| \mathscr{F}_{t-1}\right]\right| \underset{h\to+\infty}{\longrightarrow} \begin{cases} 0 & if \quad \psi^{<\alpha-1>} + \sum_{i=0}^{+\infty} a_{0,i}(\psi^{<1-\alpha>})^i = 0, \\ +\infty & else, \end{cases}$$

where the $a_{0,i}$'s are defined in Lemma 2.8.3.

Proof.

To complete the proof of Corollary 2.3.1 in this case, we will derive the limit of $Q_h(\psi^{<\alpha-1>}) = \left(\psi^{<\alpha-1>}\right)^{h-1}\left[\psi^{<\alpha-1>} + \sum_{i=0}^{h-1} a_{0,i}(\psi^{<1-\alpha>})^i\right]$ when $\alpha < 1$. Recall that we have shown $a_{0,h} \underset{h\to+\infty}{\sim} Ch^{m-1}\lambda^h$.

In this case, we have $|\lambda||\psi|^{1-\alpha} < 1$, thus $\left|\psi^{<\alpha-1>} + \sum_{i=0}^{h-1} a_{0,i}(\psi^{<1-\alpha>})^i\right| \underset{i\to+\infty}{\longrightarrow} D$, where D is a nonnegative constant.

- Assume $D > 0$. Then $|Q_h(\psi^{<\alpha-1>})| \to +\infty$ as h tends to infinity, since $|\psi|^{(\alpha-1)(h-1)} \to +\infty$.
- Assume $D = 0$. We will show that $|Q_h(\psi^{<\alpha-1>})| \longrightarrow 0$.

  Indeed, we have

$$\psi^{<\alpha-1>} + \sum_{i=0}^{+\infty} a_{0,i}(\psi^{<1-\alpha>})^i = 0$$
$$\psi^{<\alpha-1>} + \sum_{i=0}^{h-1} a_{0,i}(\psi^{<1-\alpha>})^i = -\sum_{i=h}^{+\infty} a_{0,i}(\psi^{<1-\alpha>})^i.$$

  Thus,

$$\begin{aligned} |Q_h(\psi^{<\alpha-1>})| &= |\psi|^{(\alpha-1)(h-1)}\left|\psi^{<\alpha-1>} + \sum_{i=0}^{h-1} a_{0,i}(\psi^{<1-\alpha>})^i\right| \\ &= |\psi|^{(\alpha-1)(h-1)}\left|\sum_{i=h}^{+\infty} a_{0,i}(\psi^{<1-\alpha>})^i\right| \\ &\le |\psi|^{(\alpha-1)(h-1)} \sum_{i=h}^{+\infty} |a_{0,i}||\psi|^{i(1-\alpha)}, \end{aligned}$$

and

$$\sum_{i=h}^{+\infty} |a_{0,i}||\psi|^{i(1-\alpha)} \underset{h\to+\infty}{\sim} |C| \sum_{i=h}^{+\infty} i^{m-1}(|\lambda||\psi|^{1-\alpha})^i.$$

We will show that for any $x \in (0,1)$, and any integer $r \geq 0$,

$$\sum_{i=h}^{+\infty} i^r x^i \underset{h\to+\infty}{\sim} h^r x^h (1-x)^{-1}, \tag{2.49}$$

which will imply

$$|\psi|^{(\alpha-1)(h-1)} \sum_{i=h}^{+\infty} |a_{0,i}||\psi|^{i(1-\alpha)} \underset{h\to+\infty}{=} O(h^{m-1}|\lambda|^h),$$

and thus $|Q_h(\psi^{<\alpha-1>})| \longrightarrow 0$, yielding the conclusion.

Let us now prove Equation (2.49). Notice that for $x \in (0,1)$, the sequences $\Big(i^r x^i\Big)_i$ and $\Big(i(i-1)\dots(i-r+1)x^i\Big)_i$ are equivalent as i tends to infinity and are both general terms of absolutely convergent series. Thus,

$$\sum_{i=h}^{+\infty} i^r x^i \underset{h\to+\infty}{\sim} \sum_{i=h}^{+\infty} i(i-1)\dots(i-r+1)x^i = x^r g^{(r)}(x),$$

where $g(x) := \sum_{i=h}^{+\infty} x^i = x^h(1-x)^{-1}$.

By Leibniz formula, we obtain

$$g^{(r)}(x) = \sum_{j=0}^{r} \frac{h!(r-j)!}{(h-j)!} \frac{x^{h-j}}{(1-x)^{r-j+1}} \underset{h\to+\infty}{\sim} \frac{h^r x^{h-r}}{1-x},$$

and thus,

$$\sum_{i=h}^{+\infty} i^r x^i \underset{h\to+\infty}{\sim} x^r \frac{h^r x^{h-r}}{1-x} = \frac{h^r x^h}{1-x}.$$

Substituting x by $|\lambda||\psi|^{1-\alpha}$ concludes the proof.

In the case $\alpha \in (0,1)$, i.e. for the heavier tails within the stable family, the absolute conditional expectation tends to $+\infty$ in modulus whenever the quantity $\psi^{<\alpha-1>} + \sum_{i=0}^{+\infty} a_{0,i}(\psi^{<1-\alpha>})^i$ does not vanish. This divergence is coherent with the fact that the unconditional expectation of (X_t) does not exist when $\alpha < 1$. It would be striking to have a case for which the above quantity is exactly zero, which would imply that the conditional expectation vanishes even for this class of particularly extreme processes. However, as the following example shows, all MAR(1,1) feature diverging conditional expectation when $\alpha < 1$.

Example 2.9.1 (Asymptotic predictions of the MAR(1,1) process) Let (X_t) be defined by Equation (2.6). From the explicit predictions formulated in Section 2.3.4, we deduce the asymptotic equivalents as the horizon h tends to infinity:

$$\mathbb{E}\Big[X_{t+h}\Big|\mathscr{F}_{t-1}\Big] \underset{h\to+\infty}{\overset{a.s.}{\sim}} \begin{cases} \dfrac{(\psi^{<\alpha-1>})^{h+1}}{1-\psi^{<1-\alpha>}\phi}\left(X_{t-1}-\phi X_{t-2}\right), & \text{if} \quad |\phi| \quad < |\psi|^{\alpha-1}, \\ \dfrac{\phi^{h+2}}{\phi-\psi^{<\alpha-1>}}\Big(X_{t-1}-\psi^{<\alpha-1>}X_{t-2}\Big), & \text{if} \quad |\phi| \quad > |\psi|^{\alpha-1}, \\ \phi^{h+1}\left(X_{t-1}-\dfrac{1+(-1)^h}{2}(X_{t-1}-\phi X_{t-2})\right), & \text{if} \quad \phi \quad = -\psi^{<\alpha-1>}, \\ (h+1)\phi^{h+1}(X_{t-1}-\phi X_{t-2}), & \text{if} \quad \phi \quad = \psi^{<\alpha-1>}. \end{cases}$$

Noticing that the condition $|\phi| < |\psi|^{\alpha-1}$ is equivalent to $\alpha < 1+\frac{\ln|\phi|}{\ln|\psi|}$, with $\frac{\ln|\phi|}{\ln|\psi|} > 0$, it can be seen that the three asymptotic limits of Corollary 2.3.1 are consistent with these equivalents. In particular, when $\alpha = 1$, we always have $|\phi| < 1 = |\psi|^{\alpha-1}$ and we get that, almost surely,

$$\left|\mathbb{E}\left[X_{t+h}\middle| X_{t-1}, X_{t-2}\right]\right| \underset{h\to+\infty}{\longrightarrow} \ell_{t-1} = \left|\frac{X_{t-1} - \phi X_{t-2}}{1-\operatorname{sign}(\psi)\phi}\right|.$$

2.9.2 Unit root property and extension

The equality $\mathbb{E}\big[X_t\big|X_{t-1}\big] = X_{t-1}$ for the noncausal Cauchy AR(1) with positive AR coefficient shows the existence of a unit root. Indeed, we have $X_t = X_{t-1} + \eta_t$ where $\mathbb{E}\big[\eta_t\big|X_{t-1}\big] = 0$. We show in this section that this property actually extends to more general MAR processes. The next result provides the conditional expectation of X_t given X_{t-1}.

Proposition 2.9.1 *Let X_t be the MAR(p,q) process solution of* (2.1) *with symmetric α-stable errors,* $1 < \alpha < 2$. *Denoting (d_k) the coefficients sequence of its MA(∞) representation, we have*

$$\mathbb{E}\left[X_t\middle| X_{t-1}\right] = \frac{\sum\limits_{k\in\mathbb{Z}} d_k \left(d_{k+1}\right)^{<\alpha-1>}}{\sum\limits_{k\in\mathbb{Z}} \left|d_{k+1}\right|^{\alpha}} X_{t-1}.$$

The condition for the existence of a unit root is now straightforward.

Corollary 2.9.1 *Under the assumptions of Proposition 2.9.1,*

$$\mathbb{E}\left[X_t\middle| X_{t-1}\right] = X_{t-1} \qquad \Longleftrightarrow \qquad \sum_{k\in\mathbb{Z}} d_k \left(d_{k+1}\right)^{<\alpha-1>} = \sum_{k\in\mathbb{Z}} \left|d_{k+1}\right|^{\alpha}.$$

The case $\alpha \leq 1$ is more intricate because the expectation on the left-hand side of (2.50) might not exist. However, the conditions for existence can be established using Theorem 2.13 of Samorodnistky, Taqqu (1994) [117]. This is left for further research. Proposition 2.9.1 is a consequence of the more general conditional expectation of X_t given any linear combination of the past that we provide in the next result.

Proposition 2.9.2 *Let X_t be the MAR(p,q) process solution of* (2.1) *with symmetric α-stable errors,* $1 < \alpha < 2$. *Denote (d_k) the coefficients sequence of its MA(∞) representation. Then for any $h \geq 0$, $k \geq 1$, and $a_1, \dots, a_k$ such that there exists $\ell \in \mathbb{Z}$, $a_1 d_{\ell+1} + \dots + a_k d_{\ell+k} \neq 0$, we have*

$$\mathbb{E}\Big[X_{t+h}\Big|\sum_{j=1}^{k} a_j X_{t-j}\Big] = \frac{\sum_{\ell\in\mathbb{Z}} d_{\ell-h}\left(\sum_{j=1}^{k} a_j d_{\ell+j}\right)^{<\alpha-1>}}{\sum_{\ell\in\mathbb{Z}}\left|\sum_{j=1}^{k} a_j d_{\ell+j}\right|^{\alpha}} (a_1 X_{t-1} + \dots + a_k X_{t-k}). \tag{2.50}$$

Proposition 2.9.1 is obtained for $k = 1$, $a_1 = 1$.

Proof.

Let us introduce $Y_{t-1,k} = a_1 X_{t-1} + \dots + a_k X_{t-k}$. Let $\varphi(u,v) = \mathbb{E}\left[e^{iuY_{t-1,k}+ivX_{t+h}}\right]$. For any $(u,v) \in \mathbb{R}^2$

we have,

$$
\begin{aligned}
\varphi(u,v) &= \mathbb{E}\left[\exp\left\{iu\sum_{j=1}^{k}a_j\sum_{\ell\in\mathbb{Z}}d_\ell\varepsilon_{t+\ell-j}+v\sum_{\ell\in\mathbb{Z}}d_\ell\varepsilon_{t+\ell+h}\right\}\right]\\
&= \mathbb{E}\left[\exp\left\{i\sum_{\ell\in\mathbb{Z}}\left(u\sum_{j=1}^{k}a_jd_{\ell+j}+vd_{\ell-h}\right)\varepsilon_{t+\ell}\right\}\right]\\
&= \exp\left\{-\sigma^\alpha\sum_{\ell\in\mathbb{Z}}\left|u\sum_{j=1}^{k}a_jd_{\ell+j}+vd_{\ell-h}\right|^\alpha\right\}.
\end{aligned}
$$

Thus,

$$
\frac{\partial\varphi}{\partial u}(u,v) = -\alpha\sigma^\alpha\varphi(u,v)\sum_{\ell\in\mathbb{Z}}\left(\sum_{j=1}^{k}a_jd_{\ell+j}\right)\left(u\sum_{j=1}^{k}a_jd_{\ell+j}+vd_{\ell-h}\right)^{<\alpha-1>},
$$

and

$$
\left.\frac{\partial\varphi}{\partial u}\right|_{v=0} = -\alpha\sigma^\alpha u^{<\alpha-1>}\varphi(u,0)\sum_{\ell\in\mathbb{Z}}\left|\sum_{j=1}^{k}a_jd_{\ell+j}\right|^\alpha.
$$

We also have

$$
\begin{aligned}
\frac{\partial\varphi}{\partial v}(u,v) &= -\alpha\sigma^\alpha\varphi(u,v)\sum_{\ell\in\mathbb{Z}}d_{\ell-h}\left(u\sum_{j=1}^{k}a_jd_{\ell+j}+vd_{\ell-h}\right)^{<\alpha-1>},\\
\left.\frac{\partial\varphi}{\partial v}\right|_{v=0} &= -\alpha\sigma^\alpha u^{<\alpha-1>}\varphi(u,0)\sum_{\ell\in\mathbb{Z}}d_{\ell-h}\left(\sum_{j=1}^{k}a_jd_{\ell+j}\right)^{<\alpha-1>}.
\end{aligned}
$$

Therefore,

$$
\left.\frac{\partial\varphi}{\partial v}\right|_{v=0} = \frac{\sum_{\ell\in\mathbb{Z}}d_{\ell-h}\left(\sum_{j=1}^{k}a_jd_{\ell+j}\right)^{<\alpha-1>}}{\sum_{\ell\in\mathbb{Z}}\left|\sum_{j=1}^{k}a_jd_{\ell+j}\right|^\alpha}\left.\frac{\partial\varphi}{\partial u}\right|_{v=0} \tag{2.51}
$$

On the other hand, for $u\neq 0$:

$$
\left.\frac{\partial\varphi}{\partial u}\right|_{v=0} = i\mathbb{E}\left[Y_{t-1,k}e^{iuY_{t-1,k}}\right],\quad \left.\frac{\partial\varphi}{\partial v}\right|_{v=0} = i\mathbb{E}\left[X_{t+h}e^{iuY_{t-1,k}}\right].
$$

Therefore, for $u\in\mathbb{R}^*$:

$$
\mathbb{E}\left[\left(X_{t+h}-\frac{\sum_{\ell\in\mathbb{Z}}d_{\ell-h}\left(\sum_{j=1}^{k}a_jd_{\ell+j}\right)^{<\alpha-1>}}{\sum_{\ell\in\mathbb{Z}}\left|\sum_{j=1}^{k}a_jd_{\ell+j}\right|^\alpha}Y_{t-1,k}\right)e^{iuY_{t-1,k}}\right] = 0. \tag{2.52}
$$

Hence, from Bierens (Theorem 1, 1982) [12]: Thus

$$
\mathbb{E}\left[X_{t+h}\middle|Y_{t-1,k}\right] = \frac{\sum_{\ell\in\mathbb{Z}}d_{\ell-h}\left(\sum_{j=1}^{k}a_jd_{\ell+j}\right)^{<\alpha-1>}}{\sum_{\ell\in\mathbb{Z}}\left|\sum_{j=1}^{k}a_jd_{\ell+j}\right|^\alpha}Y_{t-1,k}.
$$

2.9.3 Conditional heteroscedasticity of the MAR$(1, q)$ process

In order to prove Proposition 2.3.3, we need to show some preliminary results about the conditional covariance of noncausal AR(1) processes. We will then turn to the conditional covariance of a MAR$(1, q)$ process from which the conditional variance will be obtained.

Conditional correlation structure of the MAR$(1, q)$

Lemma 2.9.1 *Let X_t be a noncausal AR(1) process satisfying $X_t = \psi X_{t+1} + \varepsilon_t$, with $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(1, 0, \sigma, 0)$. Then, for any nonnegative integers h and τ:*

$$\mathbb{E}\Big[X_{t+h}X_{t+h+\tau}\Big|X_{t-1}\Big] = (sign\ \psi)^{\tau}\left[|\psi|^{-h-1}\left(X_{t-1}^2 + \frac{\sigma^2}{(1-|\psi|)^2}\right) - \frac{\sigma^2}{(1-|\psi|)^2}\right].$$

Remark 2.9.1 From the previous result, it is possible to derive the whole conditional correlation structure of (X_t). It can be shown that for any $t \in \mathbb{Z}$, and any positive integers h and τ:

$$\frac{\mathrm{Cov}\Big(X_{t+h}, X_{t+h+\tau}\Big|X_{t-1}\Big)}{\sqrt{\mathbb{V}\Big(X_{t+h}\Big|X_{t-1}\Big)}\sqrt{\mathbb{V}\Big(X_{t+h+\tau}\Big|X_{t-1}\Big)}} = (\mathrm{sign}\ \psi)^{\tau}\sqrt{\frac{|\psi|^{-h-1}-1}{|\psi|^{-h-\tau-1}-1}},$$

which, when $\tau \to +\infty$, is asymptotically equivalent to $(\mathrm{sign}\ \psi)^{\tau}|\psi|^{\tau/2}\sqrt{1-|\psi|^{h+1}}$ for any $h \geq 0$, and to $(\mathrm{sign}\ \psi)^{\tau}|\psi|^{\tau/2}$ when h becomes large. Although in our infinite variance framework, the unconditional correlation is not defined, empirical correlations can always be computed. We know from Davis and Resnick (1985,1986) [40, 41] that they converge in probability towards the theoretical autocorrelations that would prevail in the L^2 framework. Given n observations of process (X_t), we have for any $\tau \geq 0$,

$$\frac{\sum_{t=1}^{n-\tau+1} X_t X_{t+\tau}}{\sum_{t=1}^{n} X_t^2} \xrightarrow[n\to+\infty]{p} \psi^{\tau}.$$

Surprisingly, the "unconditional" autocorrelations of (X_t) do not converge to the conditional ones when $n \to +\infty$, and vanish at a much slower rate ($|\psi|^{\tau/2}$ instead of $|\psi|^{\tau}$).

We now turn to the MAR$(1, q)$ process.

Proposition 2.9.3 *Let X_t be a MAR$(1, q)$ process, $q \geq 0$, solution of Equation* (2.1) *with $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(1, 0, \sigma, 0)$. Then, for any positive integers h and τ, there exist polynomials P_h, $P_{h+\tau}$, both of degrees $q-1$, and Q_h, $Q_{h+\tau}$ of respective degrees h and $h+\tau$ such that*

$$\begin{aligned}
\mathbb{E}\left[X_{t+h}X_{t+h+\tau}\,|\,\mathscr{F}_{t-1}\right] &= (P_h(B)X_{t-1})(P_{h+\tau}(B)X_{t-1}) \\
&\quad + sign(\psi)(\phi(B)X_{t-1})\Big[(P_h(B)X_{t-1})Q_{h+\tau}(sign\ \psi) + (P_{h+\tau}(B)X_{t-1})Q_h(sign\ \psi)\Big] \\
&\quad + c_{h,\tau}\left((\phi(B)X_{t-1})^2 + \frac{\sigma^2}{(1-|\psi|)^2}\right) - \frac{\sigma^2}{(1-|\psi|)^2}Q_h(sign\ \psi)Q_{h+\tau}(sign\ \psi),
\end{aligned}$$

with $c_{h,\tau} = \sum_{i=0}^{h+\tau}\sum_{j=0}^{h} q_{i,h+\tau}q_{j,h}(sign\ \psi)^{i+j}|\psi|^{-\min(i,j)-1}$ and $Q_k(z) = \sum_{i=0}^{k} q_{i,k}z^i$, for any $k \geq 0$.

This result yields Proposition 2.3.2 by taking $h = \tau = 0$, with $P_0(B) = \phi_1 + \phi_2 B + \ldots + \phi_q B^q$ and $Q_0(B) = 1$.

Proof of Lemma 2.9.1

Consider $\varphi(x,y,z) := \mathbb{E}\Big(e^{ixX_{t+k}+iyX_{t+\ell}+izX_{t-1}}\Big)$, with $0 \le \ell \le k,\ X_t = \psi X_{t+1} + \varepsilon_t$ and $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha,0,\sigma,0)$. We have

$$\varphi(x,y,z) = \mathbb{E}\left(e^{i\sum_{n\in\mathbb{Z}}(xd_{n-k}+yd_{n-\ell}+zd_{n+1})\varepsilon_{t+n}}\right) = \exp\left\{-\sigma^\alpha \sum_{n\in\mathbb{Z}} |xd_{n-k}+yd_{n-\ell}+zd_{n+1}|^\alpha\right\}.$$

Thus, on the one hand,

$$\begin{aligned}
\frac{\partial\varphi}{\partial z} &= -\alpha\sigma^\alpha \sum_{n\in\mathbb{Z}} d_{n+1}(xd_{n-k}+yd_{n-\ell}+zd_{n+1})^{<\alpha-1>}\varphi(x,y,z),\\
\frac{\partial^2\varphi}{\partial z^2} &= (\alpha\sigma^\alpha)^2\left(\sum_{n\in\mathbb{Z}} d_{n+1}(xd_{n-k}+yd_{n-\ell}+zd_{n+1})^{<\alpha-1>}\right)^2\varphi(x,y,z)\\
&\qquad - \alpha(\alpha-1)\sum_{n\in\mathbb{Z}} d_{n+1}^2|xd_{n-k}+yd_{n-\ell}+zd_{n+1}|^{\alpha-2}\varphi(x,y,z),\\
\left.\frac{\partial^2\varphi}{\partial z^2}\right|_{\substack{x=0\\y=0}} &= (\alpha\sigma^\alpha)^2|z|^{2(\alpha-1)}\left(\sum_{n\in\mathbb{Z}}|d_{n+1}|^\alpha\right)^2\varphi(0,0,z) - \alpha(\alpha-1)|z|^{\alpha-2}\sum_{n\in\mathbb{Z}}|d_{n+1}|^\alpha\varphi(0,0,z).
\end{aligned}$$

And on the other hand,

$$\begin{aligned}
\frac{\partial\varphi}{\partial y} &= -\alpha\sigma^\alpha \sum_{n\in\mathbb{Z}} d_{n-\ell}(xd_{n-k}+yd_{n-\ell}+zd_{n+1})^{<\alpha-1>}\varphi(x,y,z),\\
\frac{\partial^2\varphi}{\partial x\partial y} &= (\alpha\sigma^\alpha)^2\left(\sum_{n\in\mathbb{Z}} d_{n-\ell}(xd_{n-k}+yd_{n-\ell}+zd_{n+1})^{<\alpha-1>}\right)\\
&\qquad\times\left(\sum_{n\in\mathbb{Z}} d_{n-k}(xd_{n-k}+yd_{n-\ell}+zd_{n+1})^{<\alpha-1>}\right)\varphi(x,y,z)\\
&\qquad - \alpha(\alpha-1)\sum_{n\in\mathbb{Z}} d_{n-\ell}d_{n-k}|xd_{n-k}+yd_{n-\ell}+zd_{n+1}|^{\alpha-2}\varphi(x,y,z),\\
\left.\frac{\partial^2\varphi}{\partial x\partial y}\right|_{\substack{x=0\\y=0}} &= (\alpha\sigma^\alpha)^2|z|^{2(\alpha-1)}\left(\sum_{n\in\mathbb{Z}} d_{n-\ell}(d_{n+1})^{<\alpha-1>}\right)\left(\sum_{n\in\mathbb{Z}} d_{n-k}(d_{n+1})^{<\alpha-1>}\right)\varphi(0,0,z)\\
&\qquad - \alpha(\alpha-1)|z|^{\alpha-2}\sum_{n\in\mathbb{Z}} d_{n-\ell}d_{n-k}|d_{n+1}|^{\alpha-2}\varphi(0,0,z).
\end{aligned}$$

Hence,

$$\begin{aligned}
\frac{1}{A_2}\left[\left.\frac{\partial^2\varphi}{\partial x\partial y}\right|_{\substack{x=0\\y=0}} - (\alpha\sigma^\alpha)^2A_1|z|^{2(\alpha-1)}\varphi(0,0,z)\right] &= -\alpha(\alpha-1)|z|^{\alpha-2}\varphi(0,0,z),\\
\frac{1}{A_3}\left[\frac{\partial^2\varphi}{\partial z^2} - (\alpha\sigma^\alpha)^2A_3^2|z|^{2(\alpha-1)}\varphi(0,0,z)\right] &= -\alpha(\alpha-1)|z|^{\alpha-2}\varphi(0,0,z),
\end{aligned}$$

with

$$A_1 = \left(\sum_{n\in\mathbb{Z}} d_{n-\ell}(d_{n+1})^{<\alpha-1>}\right)\left(\sum_{n\in\mathbb{Z}} d_{n-k}(d_{n+1})^{<\alpha-1>}\right),$$

$$A_2 = \sum_{n\in\mathbb{Z}} d_{n-\ell}d_{n-k}|d_{n+1}|^{\alpha-2},$$

$$A_3 = \sum_{n\in\mathbb{Z}} |d_{n+1}|^{\alpha}.$$

Therefore,

$$\frac{1}{A_2}\left[\left.\frac{\partial^2\varphi}{\partial x\partial y}\right|_{\substack{x=0\\y=0}} - (\alpha\sigma^\alpha)^2 A_1|z|^{2(\alpha-1)}\varphi(0,0,z)\right] = \frac{1}{A_3}\left[\frac{\partial^2\varphi}{\partial z^2} - (\alpha\sigma^\alpha)^2 A_3^2|z|^{2(\alpha-1)}\varphi(0,0,z)\right],$$

This yields for $\alpha = 1$,

$$\frac{1}{A_2}\left[\left.\frac{\partial^2\varphi}{\partial x\partial y}\right|_{\substack{x=0\\y=0}} - \sigma^2 A_1\varphi(0,0,z)\right] = \frac{1}{A_3}\left[\frac{\partial^2\varphi}{\partial z^2} - \sigma^2 A_3^2\varphi(0,0,z)\right].$$

Taking into account that $d_n = \psi^n \mathbb{1}_{\{n\geq 0\}}$ for the noncausal AR(1) and noticing that

$$\frac{\partial^2\varphi}{\partial x\partial y} = -\mathbb{E}\left[X_{t+k}X_{t+\ell}e^{izX_{t-1}}\right],$$

$$\frac{\partial^2\varphi}{\partial z^2} = -\mathbb{E}\left[X_{t-1}^2 e^{izX_{t-1}}\right],$$

we get for any $z \in \mathbb{R}^*$:

$$\mathbb{E}\left[\left\{X_{t+kX_{t+\ell}} - (\text{sign } \psi)^{k+\ell}\left(|\psi|^{-\ell-1}(X_{t-1}^2+\tilde{\sigma}^2) - \tilde{\sigma}^2\right)\right\}e^{izX_{t-1}}\right] = 0,$$

with $\tilde{\sigma} = \dfrac{\sigma}{1-|\psi|}$. From Bierens (Theorem 1, 1982) [12]:

$$\mathbb{E}\left[X_{t+k}X_{t+\ell}\middle|X_{t-1}\right] = (\text{sign } \psi)^{k+\ell}\left(|\psi|^{-\ell-1}(X_{t-1}^2+\tilde{\sigma}^2) - \tilde{\sigma}^2\right),$$

which concludes the proof.

Proof of Proposition 2.9.3

Let k and ℓ be two positive integers such that $\ell \leq k$. From Lemma 2.8.2, we know that for any $h \geq 0$, there exist two polynomials P_h and Q_h of respective degrees $q-1$ and h such that:

$$X_{t+h} = P_h(B)X_{t-1} + Q_h(F)u_t.$$

Thus, using the same device as in the Proof of Proposition 2.3.2,

$$\begin{aligned}\mathbb{E}\Big[X_{t+k}X_{t+\ell}\Big|X_{t-1},\dots,X_{t-q-1}\Big] &= \mathbb{E}\Big[\Big(P_k(B)X_{t-1}+Q_k(F)u_t\Big)\Big(P_\ell(B)X_{t-1}+Q_\ell(F)u_t\Big)\Big|X_{t-1},\dots,X_{t-q-1}\Big],\\ &= \Big(P_k(B)X_{t-1}\Big)\Big(P_\ell(B)X_{t-1}\Big)\\ &\quad + \Big(P_k(B)X_{t-1}\Big)\mathbb{E}\Big[Q_\ell(F)u_t\Big|u_{t-1}\Big] + \Big(P_\ell(B)X_{t-1}\Big)\mathbb{E}\Big[Q_k(F)u_t\Big|u_{t-1}\Big]\\ &\quad + \sum_{i=0}^{k}\sum_{j=0}^{\ell} q_i q_j \mathbb{E}\Big[u_{t+i}u_{t+j}\Big|u_{t-1}\Big].\end{aligned}$$

The second and third terms can be expressed as:

$$\Big(P_k(B)X_{t-1}\Big)\mathbb{E}\Big[Q_\ell(F)u_t\Big|u_{t-1}\Big] + \Big(P_\ell(B)X_{t-1}\Big)\mathbb{E}\Big[Q_k(F)u_t\Big|u_{t-1}\Big] =$$
$$\operatorname{sign}(\psi)\Big(\phi(B)X_{t-1}\Big)\Big[Q_\ell(\operatorname{sign}\,\psi)\Big(P_k(B)X_{t-1}\Big) + Q_k(\operatorname{sign}\,\psi)\Big(P_\ell(B)X_{t-1}\Big)\Big],$$

whereas the fourth term can be rewritten using Lemma 2.9.1:

$$\begin{aligned}\sum_{i=0}^{k}\sum_{j=0}^{\ell} q_i q_j \mathbb{E}\Big[u_{t+i}u_{t+j}\Big|u_{t-1}\Big] &= \sum_{i=0}^{k}\sum_{j=0}^{\ell} q_i q_j (\operatorname{sign}\,\psi)^{i+j}\left[|\psi|^{-\min(i,j)-1}\Big((\phi(B)X_{t-1})^2+\tilde{\sigma}^2\Big)-\tilde{\sigma}^2\right],\\ &= -\tilde{\sigma}^2 Q_k(\operatorname{sign}\,\psi)Q_\ell(\operatorname{sign}\,\psi)\\ &\quad+\Big((\phi(B)X_{t-1})^2+\tilde{\sigma}^2\Big)\sum_{i=0}^{k}\sum_{j=0}^{\ell} q_i q_j(\operatorname{sign}\,\psi)^{i+j}|\psi|^{-\min(i,j)-1}.\end{aligned}$$

Proof of Proposition 2.3.3

The result of Proposition 2.3.3 is obtained by substituting $\mathbb{E}\Big[X_{t+h}\Big|\mathscr{F}_{t-1}\Big]$ and $\mathbb{E}\Big[X_{t+h}^2\Big|\mathscr{F}_{t-1}\Big]$ in

$$\mathbb{V}\Big(X_{t+h}\Big|\mathscr{F}_{t-1}\Big) = \mathbb{E}\Big[X_{t+h}^2\Big|\mathscr{F}_{t-1}\Big] - \Big(\mathbb{E}\Big[X_{t+h}\Big|\mathscr{F}_{t-1}\Big]\Big)^2,$$

using the formulas of Propositions 2.3.2 and 2.9.3.

Details on the conditional variance of the MAR(1,1) of Section 2.3.4

By Lemma 2.8.3, the polynomial Q_h intervening in Proposition 2.3.3 reads in the case of the MAR(1,1)

$$Q_h(z) = \sum_{i=0}^{h}\phi^{h-i}z^i.$$

Applying Proposition 2.3.3, we know that

$$\mathbb{V}\Big(X_{t+h}\Big|\mathscr{F}_{t-1}\Big) = \Big((X_{t-1}-\phi X_{t-2})^2 + \frac{\sigma^2}{(1-|\psi|)^2}\Big)\Big(c_h - \Big(Q_h(\operatorname{sign}\,\psi)\Big)^2\Big),$$

with $c_h = \sum_{i=0}^{h}\sum_{j=0}^{h} q_{i,h}q_{j,h}(\operatorname{sign}\,\psi)^{i+j}|\psi|^{-\min(i,j)-1}$. Using the explicit form of the $q_{i,h}$'s, the coefficients of polynomial Q_h, we can deduce that for $\psi > 0$

$$\begin{aligned}Q_h(\operatorname{sign}\psi) &= \frac{1-\phi^{h+1}}{1-\phi},\\ c_h &= \psi^{-h-1}\sum_{i=0}^{h}\sum_{j=0}^{h}\phi^i\phi^j\psi^{\max(i,j)},\end{aligned}$$

which can be simplified by elementary calculations after splitting the sums according to whether $i \geq j$ or $j > i$.

2.9.4 Proof of Lemma 2.8.3

For $h = 0$, Equation (2.39) holds with $P_0(B) = \phi_1 + \phi_2 B^2 \ldots + \phi_q B^{q-1}$ and $Q_0(B) = 1$. We have

$$
\begin{aligned}
X_{t+h} &= a_{0,h} X_{t-1} + \sum_{i=1}^{q-1} a_{i,h} X_{t-i-1} + \sum_{i=0}^{h} b_{i,h} u_{t+i} \\
&= a_{0,h} \left(\sum_{i=0}^{q-1} \phi_{i+1} X_{t-i-2} + u_{t-1} \right) + \sum_{i=1}^{q-1} a_{i,h} X_{t-i-1} + \sum_{i=0}^{h} b_{i,h} u_{t+i} \\
&= \sum_{i=0}^{q-2} \Big(a_{i+1,h} + a_{0,h} \phi_{i+1} \Big) X_{t-i-2} + a_{0,h} \phi_q X_{t-q-1} + a_{0,h} u_{t-1} + \sum_{i=0}^{h} b_{i,h} u_{t+i}.
\end{aligned}
$$

Since this last formula holds at any $t \in \mathbb{Z}$, this last equation yields

$$
X_{t+h+1} = \sum_{i=0}^{q-2} \Big(a_{i+1,h} + a_{0,h} \phi_{i+1} \Big) X_{t-i-1} + a_{0,h} \phi_q X_{t-q} + a_{0,h} u_t + \sum_{i=1}^{h+1} b_{i-1,h} u_{t+i}.
$$

However, we also have by definition

$$
X_{t+h+1} = P_{h+1}(B) X_{t-1} + Q_{h+1}(F) u_t = \sum_{i=0}^{q-1} a_{i,h+1} X_{t-i-1} + \sum_{i=0}^{h+1} b_{i,h+1} u_{t+i}.
$$

Thus, by identification,

$$
\begin{aligned}
a_{q-1,h+1} &= a_{0,h} \phi_q, \\
a_{i,h+1} &= a_{i+1,h} + a_{0,h} \phi_{i+1}, \quad \text{for} \quad 0 \le i \le q-2, \\
a_{0,h} &= b_{0,h+1}, \\
b_{i,h+1} &= b_{i-1,h}, \quad \text{for} \quad 1 \le i \le h+1.
\end{aligned}
$$

We deduce from these equations that for any $h \ge 0$,

$$
\begin{aligned}
b_{i,h+1} &= a_{0,h-i}, \quad \text{for} \quad 0 \le i \le h+1, \\
a_{i,h+1} &= \sum_{j=0}^{\min(q-i-1,h)} a_{0,h-j} \phi_{i+1+j}, \quad \text{for} \quad 0 \le i \le q-1,
\end{aligned}
$$

with the convention $a_{0,-1} = 1$. We obtain that $(a_{0,h})$ is the solution of the linear recurrent equation of order q

$$
a_{0,h+q} = \phi_1 a_{0,h+q-1} + \ldots + \phi_q a_{0,h}, \quad \text{for} \quad h \ge 0, \tag{2.53}
$$

with initial values $(a_{0,0}, \ldots, a_{0,q-1})$ that could be expressed as functions of $\phi_1, \ldots, \phi_q$. Denote $\lambda_1, \ldots, \lambda_s$ the distinct roots of the polynomial $F^q \phi(B)$ with respective multiplicities $m_1, \ldots, m_s$, with $s \le q$, $m_1 + \ldots + m_s = q$. Since ϕ has all its roots outside the unit circle, we know that $|\lambda_i| < 1$ for all i. Therefore, there exist polynomials $C_1, \ldots, C_q$ of respective degrees $m_1, \ldots, m_s$ such that for any $h \ge q$,

$$
a_{0,h} = C_1(h) \lambda_1^h + \ldots + C_s(h) \lambda_s^h.
$$

2.9.5 A recursive scheme for computing polynomials P_h and Q_h of Lemma 2.8.2

Lemma 2.9.2 *Polynomials P_h and Q_h of Lemma 2.8.2 satisfy the following recursive equations:*

$$BP_{h+1}(B) = P_h(B) - P_h(0)\phi(B), \qquad\qquad Q_{h+1}(F) = FQ_h(F) + P_h(0), \tag{2.54}$$

with initial conditions $Q_0(B) = 1$, $P_0(B) = \phi_1 + \phi_2 B + \ldots + \phi_q B^{q-1}$.

Proof. By applying polynomial $\phi(B)$ to (2.39), we get by (2.31)

$$\begin{aligned}\phi(B)X_{t+h} &= P_h(B)\phi(B)X_{t-1} + Q_h(F)\phi(B)u_t,\\ B^{-h}u_t &= BP_h(B)u_t + Q_h(F)\phi(B)u_t,\end{aligned}$$

which implies $B^{h+1}P_h(B) + B^h Q_h(F)\phi(B) = 1$. The same holds at rank $h+1$. Thus, denoting $Q_h(F) = \sum_{i=0}^{h} q_{i,h}F^i$ and $Q_h^*(B) := B^h Q_h(F) = \sum_{i=0}^{h} q_{h-i,h}B^i$, we also have: $B^{h+2}P_{h+1}(B)+Q_{h+1}(B)\psi^*(B)\phi(B) = 1$. Subtracting the expressions at ranks h and $h+1$ yields:

$$B^{h+1}\Big(BP_{h+1}(B) - P_h(B)\Big) + \phi(B)\Big(Q^*_{h+1}(B) - Q^*_h(B)\Big) = 0. \tag{2.55}$$

We can notice that the term of degree zero in this expression is: $\phi(0)\Big(Q^*_{h+1}(0) - Q^*_h(0)\Big) = 0$, hence $q_{h+1,h+1} = q_{h,h}$. Focusing on the next terms of degrees $i = 1, \ldots, h$, we can iteratively show that $q_{h+1-i,h+1} = q_{h-i,h}$. Finally, focusing on the term of degree $h+1$, we now deduce that $-P_h(0) + q_{1,h+1} - q_{0,h} = 0$. This leads us to the equality

$$Q^*_{h+1}(B) = Q^*_h(B) + B^{h+1}P_h(0), \tag{2.56}$$

or equivalently $Q_{h+1}(F) = FQ_h(F)+P_h(0)$, which establishes the right-hand side equation of (2.54). Finally, replacing (2.56) in (2.55) concludes the proof of Lemma 2.9.2.

2.9.6 Cluster size distribution: the noncausal AR(1) case

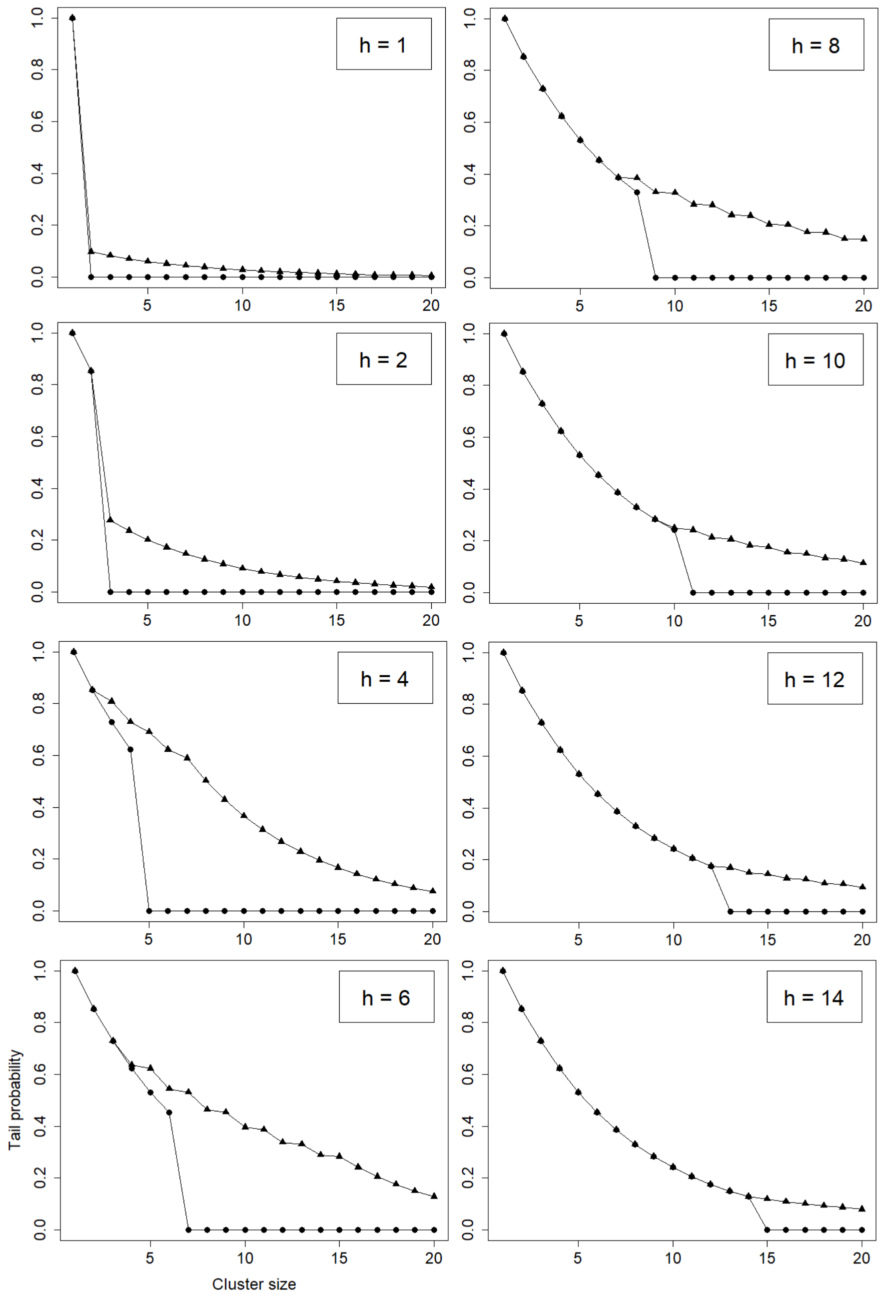


Figure 2.5: Theoretical tail probability given by Equation (2.22) of cluster sizes of extreme errors (2.24) (strong representation, points) and (2.25) (all-pass representation, triangles) for $\alpha = 1.5$, $\psi_0 = 0.9$ at different horizons h.

We illustrate the extreme clustering behaviours of the two error sequences (2.24) and (2.25) for various horizons and parameter values $\alpha = 1.5$, $\psi_0 = 0.9$. From equations (2.24) and (2.25), we deduce the sequence $(c_{(k)})$ and compute the tail probability distributions of the cluster size using (2.22). As depicted in Figure 2.5, the contrast between the errors of the all-pass representations and those of the strong representations is the highest for intermediate values of h.

2.9.7 Monte Carlo study: complementary results and methodology

Asymptotic distribution of the LS estimator

			$\alpha=1.5$	$\psi=0.7$	$\phi=0.9$			$\alpha=1$	$\psi=0.7$	$\phi=0.9$	
n		$q_{0.1}$	$q_{0.25}$	Median	$q_{0.75}$	$q_{0.9}$	$q_{0.1}$	$q_{0.25}$	Median	$q_{0.75}$	$q_{0.9}$
500	$\hat{\delta}_1$	-2.759	-1.338	-0.527	-0.061	0.231	-12.69	-3.569	-0.731	0.012	0.691
	$\hat{\delta}_2$	-0.265	0.038	0.495	1.284	2.653	-0.873	-0.049	0.694	3.430	12.13
2000	$\hat{\delta}_1$	-1.558	-0.746	-0.226	0.086	0.417	-6.321	-1.732	-0.221	0.247	1.382
	$\hat{\delta}_2$	-0.448	-0.105	0.214	0.730	1.521	-0.662	-0.320	0.001	0.322	0.655
5000	$\hat{\delta}_1$	-1.188	-0.565	-0.132	0.156	0.513	-4.564	-1.269	-0.097	0.387	1.824
	$\hat{\delta}_2$	-0.536	-0.172	0.125	0.561	1.177	-2.098	-0.469	0.096	1.357	4.749
∞	$\hat{\delta}_1$	-0.726	-0.252	0.000	0.246	0.719	-5.470	-0.856	0.000	0.954	5.686
	$\hat{\delta}_2$	-0.762	-0.264	0.000	0.268	0.768	-6.687	-1.110	0.000	1.006	6.503
			$\alpha=0.5$	$\psi=0.7$	$\phi=0.9$			$\alpha=1.7$	$\psi=0.3$	$\phi=0.4$	
500	$\hat{\delta}_1$	-1307	-114.6	-5.247	0.157	14.06	-1.003	-0.513	-0.042	0.408	0.870
	$\hat{\delta}_2$	-21.31	-0.412	5.176	114.8	1239	-0.958	-0.484	-0.008	0.466	0.956
2000	$\hat{\delta}_1$	-524.3	-40.97	-0.493	2.804	54.63	-0.662	-0.328	-0.016	0.290	0.618
	$\hat{\delta}_2$	-74.37	-4.171	0.506	46.28	563.9	-0.662	-0.320	0.001	0.322	0.655
5000	$\hat{\delta}_1$	-385.3	-28.11	-0.109	5.402	96.34	-0.641	-0.313	-0.008	0.292	0.608
	$\hat{\delta}_2$	-127.1	-7.493	0.111	33.07	445.0	-0.647	-0.318	-0.001	0.316	0.648
∞	$\hat{\delta}_1$	-1546	-31.43	0.000	32.34	1614	-0.555	-0.235	0.000	0.231	0.554
	$\hat{\delta}_2$	-2129	-42.88	0.000	41.63	2068	-0.614	-0.257	0.001	0.261	0.621

Table 2.7: Characteristics of the empirical distribution of $\hat{\delta}_i = \left(\frac{n}{\ln n}\right)^{1/\alpha}(\hat{\boldsymbol{\eta}}_i - \boldsymbol{\eta}_{0i})$, for $i = 1, 2$ over 100,000 simulated paths of α-stable MAR(1,1) processes (X_t) solution of $(1-\psi F)(1-\phi B)X_t = \varepsilon_t$ with four different parametrisations $(\alpha, \psi_0, \phi_0) \in \{(1.5, 0.7, 0.9), (1, 0.7, 0.9), (0.5, 0.7, 0.9), (1.7, 0.3, 0.4)\}$. The empirical a-quantile is denoted q_a. The results for $n = \infty$ are obtained by simulations of the asymptotic distribution in (2.16). [See Example 2.4.1]

Direct implementation of the Portmanteau test

We conducted an experiment to assess the direct implementation of the portmanteau test (without Monte Carlo) and focused on $\alpha = 1.5$. We computed the residuals of the 100,000 simulated paths based on the all-pass causal AR(2) fits, evaluate the statistic (2.19) for $h = 1, \dots, 10$ and simulate its asymptotic distribution. For each path, we performed the test at three different nominal sizes 1%, 5% and 10% by comparing the statistics to the appropriate quantile of the asymptotic distribution. The empirical sizes are reported in Table 2.8. The test suffers heavy distortions, especially in smaller samples, which was expected from the results by Lin and McLeod (2008) [94] in the pure causal AR framework. It is generally oversized for small lags and progressively becomes undersized as more lags are included. The empirical sizes slowly approach the nominal sizes as the number of observations increases and the discrepancy between few and more lags also gets smaller.

	$n=500$			$n=2000$			$n=5000$		
H	1%	5%	10%	1%	5%	10%	1%	5%	10%
1	6.69	21.2	31.7	3.08	9.42	17.0	1.92	6.28	12.5
2	4.54	16.4	27.1	2.40	7.80	14.7	1.60	5.77	11.6
3	3.40	13.4	22.8	1.96	6.41	12.4	1.36	4.84	10.1
4	2.65	10.7	19.0	1.64	5.38	10.3	1.17	4.17	8.74
5	2.11	8.96	16.2	1.37	4.58	8.96	1.04	3.59	7.61
6	1.61	7.58	13.8	1.16	3.93	7.94	0.91	3.20	6.84
7	1.24	6.49	12.1	1.01	3.51	7.17	0.80	2.86	6.22
8	0.96	5.66	10.6	0.89	3.19	6.58	0.70	2.62	5.73
9	0.74	5.08	9.62	0.81	2.94	5.99	0.64	2.42	5.30
10	0.57	4.55	8.74	0.75	2.70	5.50	0.60	2.26	5.00

Table 2.8: Empirical sizes of portmanteau tests with nominal sizes 1%, 5% and 10% using the first H lags, $H = 1, \ldots, 10$ of the residuals' autocorrelations of 100,000 simulated paths of process (X_t) solution of $(1-0.7F)(1-0.9B)X_t = \varepsilon_t$, with 1.5-stable noise.

2.9.8 Extreme residuals clustering

Estimating the term structure of excess clustering

In practice, for one simulated path of the MAR(1,1) process (X_t) and one horizon h, we have six series of residuals $(\hat{\zeta}^i_{t+h|t})_t$, $i = 1, \ldots, 6$, one each for the pure causal and noncausal AR(2) competitors, and two each for the two MAR(1,1) competitors (one for the causal component, one for the noncausal component). To compute the cluster sequences $(\hat{\xi}^i_{k,h}(x))_k$ as defined in Section 2.5.2 for each residuals series, we need to choose a threshold $x > 0$. It would be desirable to use thresholds such that we can harmoniously compare the clustering behaviours of the six series of residuals. For the experiment detailed below, we worked with the autostandardised series of residuals

$$\hat{v}^i_{t+h|t} := \left(\frac{\hat{\zeta}^i_{t+h|t}}{\max_s |\hat{\zeta}^i_{s+h|s}|} \right)_t, \tag{2.57}$$

which lie between 0 and 1, and for each horizon h, we used the threshold

$$x_h := \max_{i=1,\ldots,6} q_a\left(|\hat{v}^i_{t+h|t}|\right), \tag{2.58}$$

where $q_a(\cdot)$ the a-percent quantile. In our experiments, $a = 0.9$ was used.

Outline of the experiment

For a given parameterisation (α, ψ_0, ϕ_0) and path length n, we simulate 10000 paths of process (X_t) solution of $(1-\psi_0 F)(1-\phi_0 B)X_t = \varepsilon_t$ and conducted the experiment as follows. For each simulated path of (X_t) and a given horizon $h \geq 1$:

ι) Estimate the regression $X_t = \hat{\eta}_1 X_{t-1} + \hat{\eta}_2 X_{t-2} + \hat{\zeta}_t$.

$\iota\iota$) Obtain the set of (inverted) roots $\{\hat{\psi}, \hat{\phi}\}$ by solving for the zeros of $\hat{\eta}(z) = 1 - \hat{\eta}_1 z - \hat{\eta}_2 z^2$.

$\iota\iota\iota$) For each of the four competing models (2.26)-(2.29), decompose the process into pure causal and noncausal components and compute $(\hat{v}^i_{t+h|t})$, the series of autostandardised errors at horizons h as in (2.57).

$\iota\nu$) Compute x_h as in (2.58) and obtain the cluster sizes sequences $(\hat{\xi}^i_{k,h}(x_h))_k$ for each series $(\hat{v}^i_{t+h|t})$, $i = 1, \ldots, 6$.

ν) Compute the Excess Clustering at horizon h of each residuals series as in (2.30).

$\nu\iota$) For the two MAR(1,1) competitors, average the Excess Clustering indicators obtained from the residuals of the causal and noncausal components.

For a given simulated path (X_t), we repeat the above steps for horizons $h = 1, \ldots, H$ and obtain four term structures of Excess Clustering, one for each of the competing models (2.26)-(2.29). Across the 10000 simulated paths of (X_t), one can then either:
(i) average model-wise across the obtained term structures to gauge the typical excess clustering behaviour of each competing model (as in Figures 2.3 and 2.6), or
(ii) for each of the simulated paths (X_t), compute the area under the four term structures, select the least clustering model and evaluate the rate of correct selections (as in Table 2.3).

Excess clustering for additional parameterisations

We evaluated the residuals excess clustering behaviours of the four alternatives (2.26)-(2.29) for additional parameterisations and sample sizes of the MAR(1,1) data generating process. Excess clustering in all-pass residuals is apparent even for small sample sizes. The contrast between the residuals of the strong representation and those of the all-pass increases as the sample size grows (see the left panel of Figure 2.3 and the two upper panels of Figure 2.6). Also, even with a much smaller noncausal parameter $\psi = 0.2$ (lower right panel of Figure 2.6), the strong representation still clearly displays the least excess clustering compared to the three other competitors. We can nevertheless notice in this case that the pure causal AR(2) alternative is not far from the strong representation (points). This is coherent with the fact that the noncausal parameter ψ is relatively small, especially compared to the causal parameter ϕ, yielding much weaker dependence across the residuals of the misspecified pure causal AR(2).

2.9.9 Real data: complementary results using the R package 'MARX'

Total AR orders selection by Information Criterion

The portmanteau procedure of Section 2.6.1 allowed to discard non-admissible low order models for the six financial and economic time series considered. Portmanteau tests are however not designed to select an «optimal» model. To go further, we report in Table 2.9 the orders that minimise Akaike's information criterion (AIC) using the R package 'MARX' available on CRAN (see Hecq, Telg and Lieb (2017b)). The validity of such AIC's for innovations in the domain of attraction of a stable law has been studied by Knight

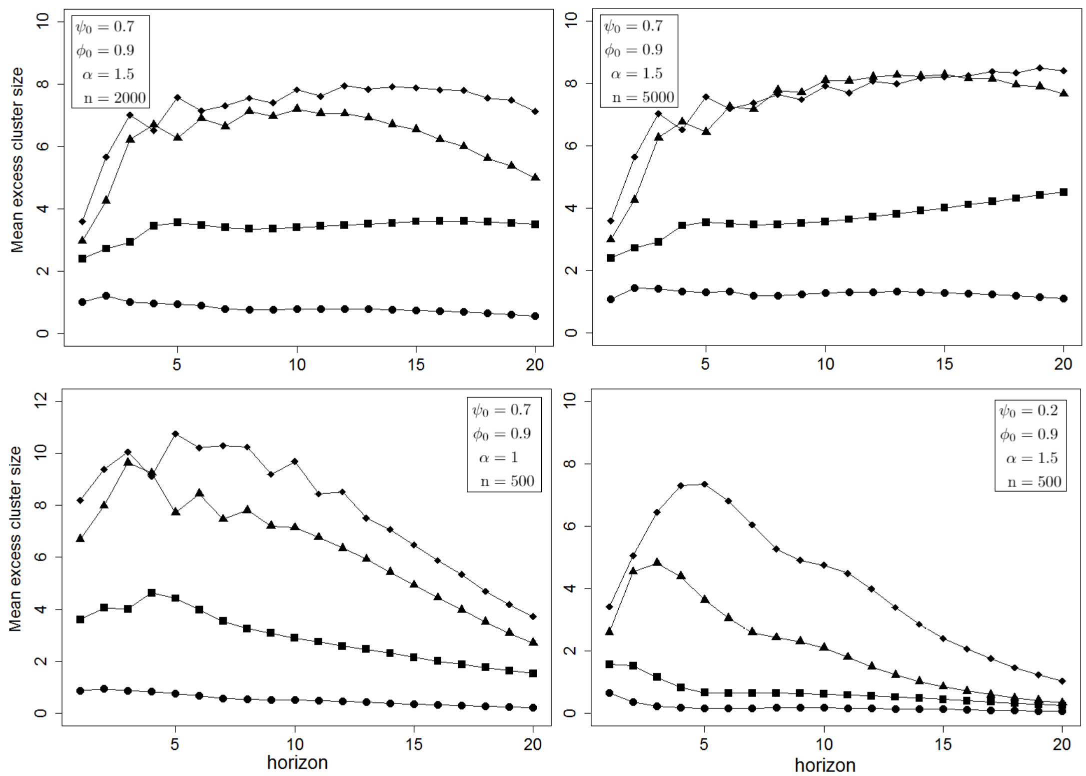


Figure 2.6: Across 10,000 simulations of the α-stable MAR(1,1) process (X_t) solution of $(1 - \psi_0 F)(1 - \phi_0 B)X_t = \varepsilon_t$, the plots show the average of the term structure of excess clustering of the linear residuals of the four competing models (2.26) (squares), the strong representation (2.27) (points), (2.28) (triangles) and (2.29) (diamond). The parameterisations and path lengths are indicated on each panel.

(1989). Except for the HSI, the results of the two procedures are compatible, the AIC criterion tending to select higher optimal orders.

	Cotton	Soybean	Sugar	Coffee	HSI	Shiller P/E
Selected total AR order	3	8	7	9	1	8

Table 2.9: Optimal order minimising the AIC criterion.

2.9.10 Identification of causal and noncausal roots

Given the lowest total AR orders validated by the portmanteau procedure (see Table 2.6.1), we used the routine marx.t of the 'MARX' package to fit MAR models on the six series by t-Student ML. The results are presented in Table 2.10. Except for the HSI and the sugar series, the causal/non-causal orders obtained are equal to those of the final specifications in Table 2.6. The estimated roots are also similar, but we note some discrepancies in their causal/non-causal allocations.

Series	Final specification	Noncausal (inverted) roots	Causal (inverted) roots
Cotton	MAR(2,0)	0.93, 0.11	–
Soybean	MAR(2,3)	$0.16 \pm 0.42i$	0.94, −0.55, 0.30
Sugar	MAR(2,2)	$0.29 \pm 0.41i$	0.96, −0.43
Coffee	MAR(1,3)	0.41	0.95, -0.23 ± 0.20
HSI	MAR(3,0)	0.92, 0.28, −0.21	–
Shiller P/E	MAR(2,4)	0.95, −0.48, $0.50 \pm 0.23i$	$-0.21 \pm 0.43i$

Table 2.10: Estimation of the MAR(p,q) specification for each financial series by t-Student ML using the routine marx.t of the 'MARX' package. This routine requires as input the total AR order $p+q$, for which we used the validated orders given by Table 4.

Chapter 3

Conditional Moments of Anticipative α-Stable Markov Processes

Sébastien Fries

Abstract The anticipative α-stable autoregression of order 1 (AR(1)) is a stationary Markov process undergoing explosive episodes akin to bubbles in financial time series data. Although featuring infinite variance, integer conditional moments up to order four may exist. The conditional expectation, variance, skewness and kurtosis are provided at any forecast horizon under any admissible parameterisation. During bubble episodes, these moments become equivalent to that of a two-point distribution charging complementary probabilities to two polarly-opposite outcomes: pursued explosion or collapse. Parallel results are obtained for the continuous time anticipative α-stable Ornstein-Uhlenbeck process. The proofs build on and extend properties of arbitrary, not necessarily symmetric α-stable bivariate random vectors. Other processes are considered such as the anticipative AR(2) and the aggregation of anticipative AR(1).

Keywords: Anticipative processes, Stable processes, Stable random vectors, Conditional moments, Explosive bubbles

MSC classes: 60G52, 60E07, 60G25

3.1 Introduction

Dynamic models often admit solution processes for which the current value of the variable is a function of future values of an independent error process. Such solutions, called *anticipative*, have attracted increasing attention in the financial and econometric literatures. In particular, anticipative processes have been shown to be convenient for modelling speculative bubbles [24, 51, 61, 63, 67, 68, 69, 70] (see also [3, 26, 87, 88]). However, lack of knowledge about the predictive distribution of anticipative processes is impeding the ability to forecast them, thus limiting their use in practical applications. Partial results have been obtained in [63] for the anticipative stable AR(1), defined as the stationary solution of

$$X_t = \rho X_{t+1} + \varepsilon_t, \qquad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, \beta, \sigma, 0), \tag{3.1}$$

where $0 < |\rho| < 1$ and $\mathcal{S}(\alpha, \beta, \sigma, 0)$ denotes the univariate α-stable distribution with tail parameter $\alpha \in (0, 2)$, asymmetry $\beta \in [-1, 1]$ and scale $\sigma > 0$. Figure 3.1 depicts a typical simulated path of an anticipative stable AR(1) featuring multiple bubbles.

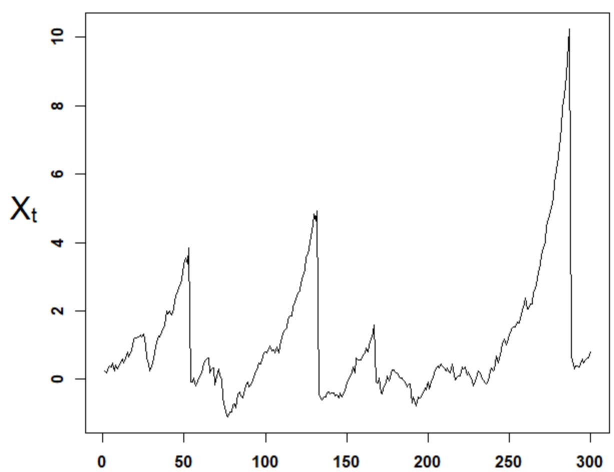


Figure 3.1: Sample path of the solution of (3.1) with $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(1.7, 0.8, 0.1, 0)$ and $\rho = 0.95$.

This paper proposes a complete characterisation of the conditional moments at any horizon, when existing, for the stable AR(1) process and two related models: the anticipative Ornstein-Uhlenbeck (OU) process and the aggregated stable process, defined as a linear combination of α-stable anticipative processes of the form (3.1). Explicit expressions of the conditional moments generally have a complex form. However, we will show that the conditional distributions of X_{t+h}, say, given $X_t = x$ displays dramatic simplifications when $x \to \pm\infty$, which provides illuminating interpretations on the behaviour of the process during bubble episodes.

Section 3.2 starts by recalling characterisations and properties of multivariate stable distributions and then provides our results on the anticipative stable AR(1) and OU processes. Section 3.3 analyses the aggregation of AR(1). Section 3.4 finds a new upper bound for the existence of conditional moments of anticipative AR(2) processes. Complementary results on bivariate stable vectors are stated in Section 3.6. Postponed proofs are collected in Section 3.7.

3.2 Anticipative α-stable Markov processes

Before analysing the anticipative α-stable AR(1) and OU processes, we begin by recalling some characterisations of multivariate stable distributions which will be the cornerstone of our proofs.

3.2.1 Characterisation of α-stable random vectors

Stable random vectors are defined in a similar way as when considering stable variables on the real line. Denote by «$\stackrel{d}{=}$» the equality in distribution between two random variables.

Definition 3.2.1 *A random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ is said to be a stable random vector in $\mathbb{R}^d$ if for any positive numbers A and B there is a positive number C and a non-random vector $\boldsymbol{D} \in \mathbb{R}^d$ such that*

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

where $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$ are independent copies of $\boldsymbol{X}$. Moreover, if $\boldsymbol{X}$ is stable, then there exists a constant $\alpha \in (0, 2]$ such that the above holds with $C = (A^\alpha + B^\alpha)^{1/\alpha}$, and $\boldsymbol{X}$ is then called α-stable.

We exclude the intensively-studied Gaussian case ($\alpha = 2$) from our analysis. Let S_d be the unit sphere of $\mathbb{R}^d$ equipped with the Euclidean norm $\|\cdot\|$ induced by the canonical scalar product, denoted $\langle \cdot, \cdot \rangle$. The distribution of stable random vectors are characterised (see Theorem 2.3.1 in [117]) by a unique pair $(\Gamma, \boldsymbol{\mu}^0)$, where Γ is a finite measure on the unit sphere S_d and a vector $\boldsymbol{\mu}^0 \in \mathbb{R}^d$. Let $0 < \alpha < 2$, then $\boldsymbol{X} = (X_1, \ldots, X_d)$ is an α-stable random vector if and only if there exists a unique pair $(\Gamma, \boldsymbol{\mu}^0)$ such that, for any $\boldsymbol{u} \in \mathbb{R}^d$, the characteristic function of $\boldsymbol{X}$ writes

$$\varphi_{\boldsymbol{X}}(\boldsymbol{u}) := \mathbb{E}\Big[e^{i\langle \boldsymbol{u}, \boldsymbol{X}\rangle}\Big] = \exp\Bigg\{ - \int_{S_d} |\langle \boldsymbol{u}, \boldsymbol{s}\rangle|^\alpha \Big(1 - i \operatorname{sign}(\langle \boldsymbol{u}, \boldsymbol{s}\rangle) w(\alpha, \langle \boldsymbol{u}, \boldsymbol{s}\rangle)\Big)\Gamma(d\boldsymbol{s}) + i\,\langle \boldsymbol{u}, \boldsymbol{\mu}^0\rangle \Bigg\}, \tag{3.2}$$

where $w(\alpha, s) = \operatorname{tg}\left(\frac{\pi\alpha}{2}\right)$, if $\alpha \neq 1$, and $w(1, s) = -\frac{2}{\pi} \ln |s|$ otherwise, for $s \in \mathbb{R}$. The finite measure Γ is called the *spectral measure* of $\boldsymbol{X}$ and captures the information about the scale, asymmetry and dependence between its components. The non-random vector $\boldsymbol{\mu}^0$ is a location parameter and is called the *shift vector*. The pair $(\Gamma, \boldsymbol{\mu}^0)$ is said to be the *spectral representation* of the random vector $\boldsymbol{X}$. In the univariate case, (3.2) boils down to

$$\varphi_X(u) = \exp\Bigg\{ - \sigma^\alpha |u|^\alpha \Big(1 - i\beta \operatorname{sign}(u) w(\alpha, u)\Big) + iu\mu \Bigg\},$$

for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\mu \in \mathbb{R}$.

Stable distributions are known to have very little moments. However, the distribution of one component conditionally on the others can have more moments according to the degree of dependence between them. A sufficient condition for the existence of conditional moments of bivariate α-stable vectors (X_1, X_2) is given in the following Proposition.

Proposition 3.2.1 (Samorodnistky and Taqqu (Theorem 5.1.3, 1994)) *Let $\boldsymbol{X} = (X_1, X_2)$ be an α-stable random vector with spectral measure Γ, satisfying*

$$\int_{S_2} |s_1|^{-\nu}\Gamma(ds) < +\infty, \quad \text{for some} \quad \nu \geq 0. \tag{3.3}$$

Then, $\mathbb{E}\Big[|X_2|^{\gamma}\Big|X_1 = x\Big] < +\infty$ for almost every x if

$$0 \leq \gamma < \min(\alpha + \nu, 2\alpha + 1).$$

The less concentrated around the points $(0, 1)$ and $(0, -1)$ of the unit circle is the spectral measure, the higher the moments of $X_2|X_1$. In the next section, condition (3.3) will be shown to hold for any $\nu \geq 0$ for the vector (X_t, X_{t+h}) when (X_t) is the solution of (3.1).

3.2.2 Discrete time: the anticipative α-stable AR(1)

Operating the arsenal of properties of multivariate α-stable distributions we provide in the previous section and Section 3.6.1, we analyse in detail the predictive distribution of the anticipative α-stable AR(1) solution of (3.1), $X_t = \sum_{k\geq 0} \rho^k \varepsilon_{t+k}$. The following result shows that when the noise sequence (ε_t) in (3.1) is α-stable distributed, then (X_t, X_{t+h}) is itself an α-stable random vector with a very specific spectral representation.

Proposition 3.2.2 *Let (X_t) be the anticipative AR(1) solution of (3.1) with $0 < \alpha < 2$, $\beta \in [-1, 1]$ and $|\rho| < 1$. Then, for any $h \geq 1$, (X_t, X_{t+h}) is α-stable and its spectral representation, denoted $(\Gamma_h, \boldsymbol{\mu}^0)$ with $\boldsymbol{\mu}^0 = (\mu_1^0, \mu_2^0)$, is such that*

$$\Gamma_h = \frac{\bar{\sigma}^{\alpha}}{2} \sum_{\vartheta\in S_1} \left[\left(1 - |\rho|^{\alpha h} + \left(1 - \left(\rho^{<\alpha>}\right)^h\right)\vartheta\bar{\beta}\right)\delta_{\{(\vartheta,0)\}} + \left(1 + |\rho|^{2h}\right)^{\alpha/2} (1 + \vartheta\bar{\beta})\delta_{\{\vartheta \boldsymbol{s}_h\}} \right], \tag{3.4}$$

where $S_1 = \{-1, +1\}$, $\delta_{\{x\}}$ is the Dirac measure at point $x \in \mathbb{R}^2$, $\bar{\sigma}^{\alpha} = \dfrac{\sigma^{\alpha}}{1 - |\rho|^{\alpha}}$, $\bar{\beta} = \beta\dfrac{1 - |\rho|^{\alpha}}{1 - \rho^{<\alpha>}}$, $y^{<r>} = sign(y)|y|^r$ for any $y, r \in \mathbb{R}$ and $\boldsymbol{s}_h = \dfrac{(\rho^h, 1)}{\sqrt{1 + |\rho|^{2h}}} \in S_2$. Moreover, if $\alpha \neq 1$, then $\boldsymbol{\mu}^0 = (0, 0)$, and if $\alpha = 1$ then,

$$\mu_1^0 = \bar{\mu} - \frac{2}{\pi}\bar{\sigma}\bar{\beta}\frac{\rho \ln |\rho|}{(1 - \rho)}, \qquad \mu_2^0 = \rho^{-h}\bar{\mu} - \frac{2}{\pi}\bar{\sigma}\bar{\beta}\ln |\rho| \left(h + \frac{\rho}{1 - \rho}\right).$$

with $\bar{\mu} = -\dfrac{\bar{\sigma}\bar{\beta}}{\pi}\rho^h \ln\left(1 + \rho^{-2h}\right)$.

It can be noticed from the previous Proposition that the spectral measure of (X_t, X_{t+h}) is discrete and concentrated on at most four points of the unit circle: $(\pm 1, 0)$ and $\pm(\rho^h, 1)/\sqrt{1 + |\rho|^{2h}}$. It collapses on exactly two points when $\rho > 0$ and $\beta = 1$ (resp. $\beta = -1$), that is, when the marginal distribution of X_t is *totally skewed to the right* (resp. *to the left*).[1] In particular, for any fixed $h \geq 1$, Γ_h is always charging zero mass to sufficiently small neighbourhoods around the points $(0, \pm 1)$, which leads to the following result and the existence of conditional moments.[2]

Lemma 3.2.1 *Let (X_t) be the anticipative AR(1) solution of (3.1) with $0 < \alpha < 2$, $\beta \in [-1, 1]$ and $0 < |\rho| < 1$. Then, for any $h \geq 1$, the spectral measure of (X_t, X_{t+h}) is such that*

$$\int_{S_2} |s_1|^{-\nu} \Gamma_h(d\boldsymbol{s}) < +\infty, \quad \text{for any} \quad \nu \geq 0. \tag{3.5}$$

Proof.
Let $\nu \geq 0$ and $h \geq 1$. Decompose the integral of (3.5) into two parts

$$\begin{aligned} &\int_{S_2} |s_1|^{-\nu} \Gamma_h(d\boldsymbol{s}) \\ &\quad = \int_{S_2 \cap \{\boldsymbol{s} \in S_2 : |s_1| \leq |\rho|^h / 2\sqrt{1+|\rho|^{2h}}\}} |s_1|^{-\nu} \Gamma_h(d\boldsymbol{s}) + \int_{S_2 \cap \{\boldsymbol{s} \in S_2 : |s_1| > |\rho|^h / 2\sqrt{1+|\rho|^{2h}}\}} |s_1|^{-\nu} \Gamma_h(d\boldsymbol{s}). \end{aligned}$$

In view of (3.4), the second term on the right-hand side is finite while the first one is zero. □

Corollary 3.2.1 *Let (X_t) be the anticipative AR(1) solution of (3.1) with $0 < \alpha < 2$, $\beta \in [-1, 1]$ and $0 < |\rho| < 1$. Then, for any $h \geq 1$,*

$$\mathbb{E}\Big[|X_{t+h}|^\gamma \Big| X_t, X_{t-1}, \dots\Big] < +\infty, \quad \text{a.s. for any} \quad 0 < \gamma < 2\alpha + 1.$$

Proof.
As the anticipative AR(1) is a Markov process (see Proposition 2 in [63]), we have $\mathbb{E}\Big[|X_{t+h}|^\gamma \Big| X_t, X_{t-1}, \dots\Big] = \mathbb{E}\Big[|X_{t+h}|^\gamma \Big| X_t\Big]$ for any h and γ. The existence of conditional moments up to order $2\alpha + 1$ is now a direct consequence of Lemma 3.2.1 and Proposition 3.2.1. □

Analytical formulae were so far available only for the first and second conditional moments of the anticipative stable AR(1) processes, moreover only in the symmetric ($\beta = 0$) and Cauchy ($\alpha = 1$ and $\beta = 0$) cases [63]. Thus we extend the formulae to any admissible parameterisations $(\alpha, \beta) \in (0, 2) \times [-1, 1]$ and also provide the forms of the third and fourth conditional moments in the next Theorem. For expository purposes, the more intricate case $\alpha = 1$ has been singled out in Section 3.6.2. Recall that the anticipative AR(1) is a Markov process and that integer conditional moments may exist only up to order four under the most favourable dispositions of Corollary 3.2.1.[3]

[1] When $\rho > 0$ and $\beta = 1$ (resp. $\beta = -1$), we have from Gouriéroux and Zakoïan (2017) that the marginal distribution of X_t is univariate α-stable with asymmetry parameter $\beta_1 = \beta \dfrac{1 - |\rho|^\alpha}{1 - \rho^{<\alpha>}} = 1$ (resp. $\beta_1 = -1$). Zolotarev (1986) call such distributions *totally skewed to the right* (resp. *to the left*).

[2] While the existence of conditional moments has been established in Chapter 2 based on techniques from [63], the proof here differs and relies solely on stable vector properties.

[3] Higher conditional moments may however exist in some boundary cases, such as when X_t is totally skewed either to the right or left.

Theorem 3.2.1 *Let (X_t) be the anticipative α-stable AR(1) solution of (3.1) with $\beta \in [-1, 1]$ and $0 < |\rho| < 1$. Let $h > 0$.*

For $\alpha \in (0, 2), \quad \alpha \neq 1,$

$$\mathbb{E}\Big[X_{t+h}\Big|X_t = x\Big] = \kappa_1 x + \frac{a(\lambda_1 - \beta_1\kappa_1)}{1 + a^2\beta_1^2}\left[a\beta_1 x + \frac{1 - xH(x)}{\pi f_X(x)}\right]. \tag{3.6}$$

For $\alpha \in (1/2, 2), \quad \alpha \neq 1,$

$$\begin{aligned}\mathbb{E}\Big[X_{t+h}^2\Big|X_t = x\Big] = \kappa_2 x^2 &+ \frac{ax(\lambda_2 - \beta_1\kappa_2)}{1 + (a\beta_1)^2}\left[a\beta_1 x + \frac{1 - xH(x)}{\pi f_X(x)}\right] \\ &- \frac{\alpha^2\sigma_1^{2\alpha}}{\pi f_X(x)}\mathcal{H}\Big(2, \boldsymbol{\theta}_1; x\Big).\end{aligned} \tag{3.7}$$

For $\alpha \in (1, 2),$

$$\begin{aligned}\mathbb{E}\Big[X_{t+h}^3\Big|X_t = x\Big] = \kappa_3 x^3 &+ \frac{ax^2(\lambda_3 - \beta_1\kappa_3)}{1 + (a\beta_1)^2}\left[a\beta_1 x + \frac{1 - xH(x)}{\pi f_X(x)}\right] \\ &- \frac{\alpha^2\sigma_1^{2\alpha}}{2\pi f_X(x)}\left[x\mathcal{H}\Big(2, \boldsymbol{\theta}_2; x\Big) + \alpha\sigma_1^\alpha\mathcal{H}\Big(3, \boldsymbol{\theta}_3; x\Big)\right].\end{aligned} \tag{3.8}$$

For $\alpha \in (3/2, 2),$

$$\begin{aligned}\mathbb{E}\Big[X_{t+h}^4\Big|X_t = x\Big] = \kappa_4 x^4 &+ \frac{ax^3(\lambda_4 - \beta_1\kappa_4)}{1 + (a\beta_1)^2}\left[a\beta_1 x + \frac{1 - xH(x)}{\pi f_X(x)}\right] \\ &- \frac{\alpha^2\sigma_1^{2\alpha}}{\pi f_X(x)}\left[\frac{x^2}{2}\mathcal{H}\Big(2, \boldsymbol{\theta}_4; x\Big) + \frac{\alpha x\sigma^\alpha}{6}\mathcal{H}\Big(3, \boldsymbol{\theta}_5; x\Big) + \frac{\alpha^2\sigma^{2\alpha}}{3}\mathcal{H}\Big(4, \boldsymbol{\theta}_6; x\Big)\right].\end{aligned} \tag{3.9}$$

Here, $a = tg(\pi\alpha/2)$, and

$$\sigma_1^\alpha = \frac{\sigma^\alpha}{1 - |\rho|^\alpha}, \qquad \beta_1 = \beta\frac{1 - |\rho|^\alpha}{1 - \rho^{<\alpha>}}, \qquad \kappa_p = |\rho|^{\alpha h}\rho^{-hp}, \qquad \lambda_p = \beta_1\big(\rho^{<\alpha>}\big)^h\rho^{-hp},$$

for $p \in \{1, 2, 3, 4\}$. Furthermore, for any $n \in \mathbb{N}$, $\boldsymbol{\theta}_i = (\theta_{i1}, \theta_{i2}) \in \mathbb{R}^2$, $x \in \mathbb{R}$, $\mathcal{H}$ is defined by

$$\mathcal{H}(n, \boldsymbol{\theta}_i; x) = \int_0^{+\infty} e^{-\sigma_1^\alpha u^\alpha} u^{n(\alpha-1)}\Big(\theta_{i1}\cos(ux - a\beta_1\sigma_1^\alpha u^\alpha) + \theta_{i2}\sin(ux - a\beta_1\sigma_1^\alpha u^\alpha)\Big)du, \tag{3.10}$$

and we denote $H(\,\cdot\,) := \mathcal{H}\big(0, (0, 1); \,\cdot\,\big)$, and $f_X := \frac{1}{\pi}\mathcal{H}\big(0, (1, 0); \,\cdot\,\big)$.[4] *Finally, $\boldsymbol{\theta}_1 = (\theta_{11}, \theta_{12})$ in (3.7) is given by*

$$\theta_{11} = \kappa_1^2 - a^2\lambda_1^2 + a^2\beta_1\lambda_2 - \kappa_2, \qquad \theta_{12} = a(\lambda_2 + \beta_1\kappa_2) - 2a\lambda_1\kappa_1, \tag{3.11}$$

and the remaining $\boldsymbol{\theta}_i$'s in (3.8)-(3.9), which depend only on α, β_1, and the κ_p's and λ_p's above, are given in (3.66)-(3.75) in Section 3.7. If $\alpha < 1$ and $\beta_1 = 1$ (resp. $\beta_1 = -1$), Relations (3.6) and (3.7) are well defined only for $x \geq 0$ (resp. $x \leq 0$).

Remark 3.2.1 In the special cases considered in [63], the conditional expectation and variance are respectively linear and quadratic functions of the past. This does not appear to be the case in general. For

[4] Notice that f_X is the density of $X_t \sim \mathcal{S}(\alpha, \beta_1, \sigma_1, 0)$ when $\alpha \neq 1$.

instance, a necessary and sufficient condition for the linearity of the conditional expectation (3.26) requires that $\lambda_1 = \beta_1\kappa_1$. In the case of the anticipative AR(1), this holds if and only if at least one of the following is true: ι) $\beta = 0$, $\iota\iota$) $\rho > 0$, or $\iota\iota\iota$) h is even.

Remark 3.2.2 From a computational perspective, note that the above moments can be inexpensively calculated for various horizons h and conditioning values x. Notice indeed that the functions $\mathcal{H}\big(n, \boldsymbol{\theta}; x\big)$, $n = 2, 3, 4$, can be decomposed into $a_h u_n(x) + b_h v_n(x)$, where a_h and b_h are constants depending only on h -and fixed parameters of the process-, and $u_n(x) = \mathcal{H}(n, (0, 1); x)$ and $v_n = \mathcal{H}(n, (1, 0); x)$ are simple integrals which depend only on x. The constants a_h and b_h can be inexpensively computed for any horizons while the simple integrals $u_n(x)$ and $v_n(x)$ need only to be computed once for a given conditioning value.

Let us denote by $\mu(x, h)$, $\sigma^2(x, h)$, $\gamma_1(x, h)$ and $\gamma_2(x, h)$ the conditional expectation, variance, skewness and excess kurtosis respectively. When they are well defined, we denote for $x \in \mathbb{R}$ and $h > 0$,

$$\mu(x, h) := \mathbb{E}\Big[X_{t+h}\Big|X_t = x\Big], \tag{3.12}$$

$$\sigma^2(x, h) := \mathbb{E}\Big[\Big(X_{t+h} - \mu(x, h)\Big)^2\Big|X_t = x\Big], \tag{3.13}$$

$$\gamma_1(x, h) := \mathbb{E}\left[\left(\frac{X_{t+h} - \mu(x, h)}{\sigma(x, h)}\right)^3\middle|X_t = x\right], \tag{3.14}$$

$$\gamma_2(x, h) := \mathbb{E}\left[\left(\frac{X_{t+h} - \mu(x, h)}{\sigma(x, h)}\right)^4\middle|X_t = x\right] - 3. \tag{3.15}$$

To illustrate the results of Theorem 3.2.1, the conditional moments of the anticipative 1.7-stable AR(1) with $\rho = 0.95$, $\beta = 0.8$ and $\sigma = 0.1$ are depicted on Figure 3.2 as functions of the past observation $X_t = x$ and the horizon h. Notice in particular that the conditional volatility $\sigma(\,\cdot\,, h)$ appears naturally smile-shaped, which reproduces a well-known stylised fact of implied volatilities and news impact curves on financial markets.

Although X_t is marginally stable-distributed, the conditional distribution of X_{t+h} given X_t is typically non-stable. For $\rho > 0$, a clear interpretation of the distribution $X_{t+h}|X_t = x$ appears during explosive/bubble episodes, that is, as x becomes large relative to the central values of process (X_t).

Corollary 3.2.2 *Let (X_t) be the anticipative strictly stationary solution of* (3.1) *with $\rho > 0$ and $\beta \in [-1, 1]$. If $|\beta_1| = 1$, let $\beta_1 x \to +\infty$, and if $|\beta_1| \neq 1$, let $x \to \pm\infty$.*[5] *Also, let $s = 1$ if $x \to +\infty$ and $s = -1$ if $x \to -\infty$. Then, for any $h \geq 1$, as $x \to \pm\infty$,*[6]

$$\begin{aligned}
&\mu(x, h) \sim (\rho^{-h}x)\rho^{\alpha h}, && \textit{if} \quad \alpha \in (0, 2),\\
&\sigma^2(x, h) \sim (\rho^{-h}x)^2\rho^{\alpha h}(1 - \rho^{\alpha h}), && \textit{if} \quad \alpha \in (1/2, 2),\\
&\gamma_1(x, h) \longrightarrow s\frac{1 - 2\rho^{\alpha h}}{\sqrt{\rho^{\alpha h}(1 - \rho^{\alpha h})}}, && \textit{if} \quad \alpha \in (1, 2),\\
&\gamma_2(x, h) \longrightarrow \frac{1}{\rho^{\alpha h}} + \frac{1}{1 - \rho^{\alpha h}} - 6, && \textit{if} \quad \alpha \in (3/2, 2).
\end{aligned}$$

[5] See Remark 3.6.3 for details regarding the different behaviours when $|\beta_1| \neq 1$ and $|\beta_1| = 1$.

[6] For $f, g : \mathbb{R} \to \mathbb{R}$, g not vanishing in some neighbourhood of $a = \pm\infty$, we denote $f(x) \sim g(x)$ when $f(x) - g(x) \underset{x \to a}{=} o(g(x))$.

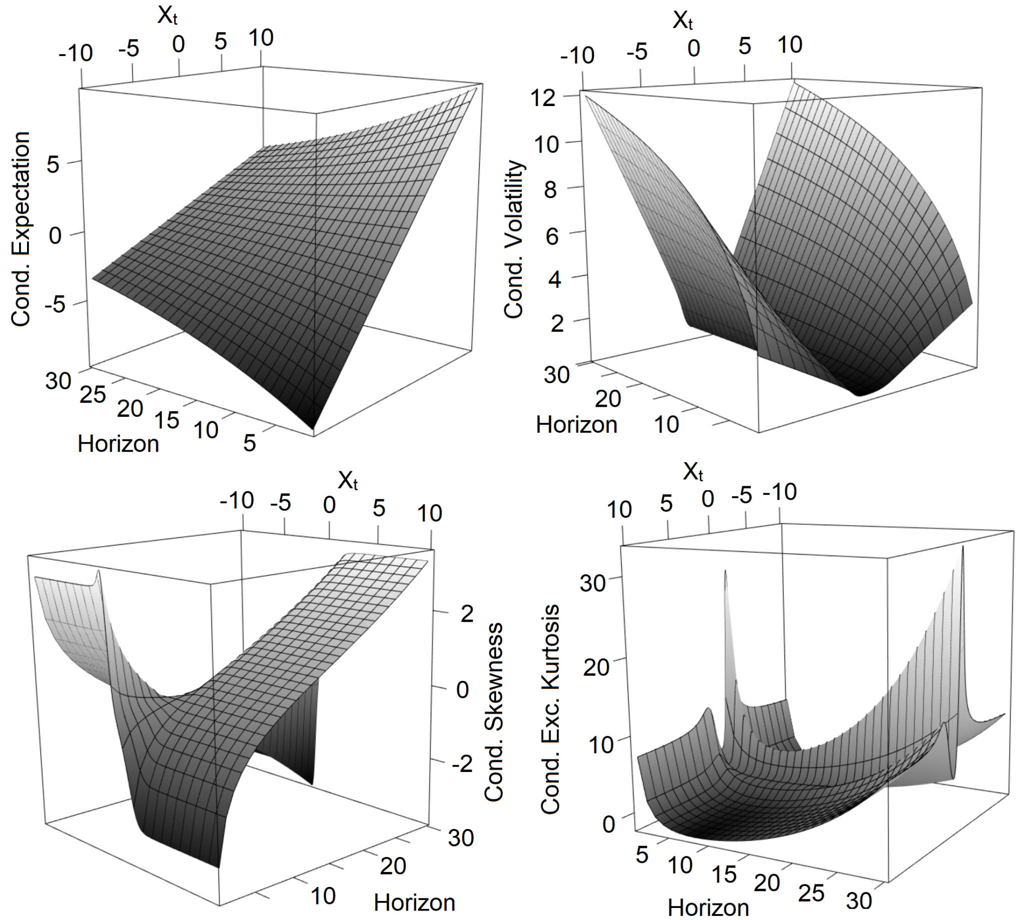


Figure 3.2: Conditional moments $\mu(\cdot)$, $\sigma(\cdot)$, $\gamma_1(\cdot)$, $\gamma_2(\cdot)$ given by (3.12)-(3.15) of the stable anticipative AR(1) solution of (3.1) with $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(1.7, 0.8, 0.1, 0)$ and $\rho = 0.95$, for horizons $h = 1, \dots, 30$ and conditioning values $X_t = x \in (-10, 10)$. Lower is darker, higher is whiter.

Remark 3.2.3 The strikingly simplistic forms of the conditional moments during explosive/bubble episodes yielded by Corollary 3.2.2 are characteristic of a weighted Bernoulli distribution charging probability $\rho^{\alpha h}$ to the value $\rho^{-h}x$ and probability $1 - \rho^{\alpha h}$ to 0. It is thus natural to interpret $\rho^{\alpha h}$ as the probability that the bubble survives at least h more time steps, conditionally on reaching the level $X_t = x$.[7] This interpretation surprisingly implies that the survival probability does not depend on the past longevity of the bubble neither on its current height. The bubbles generated by the stable anticipative AR(1) appear to display a memory-less property.

Remark 3.2.4 Corollary 3.2.2 also echoes the bubble model that was initially proposed in [13] and further

[7] The interpretation of $\rho^{\alpha h}$ as a survival probability of bubbles can also be reached using point processes under the more general assumption that the errors of (3.1) belong to the domain of attraction of an α-stable distribution (see Section 3.6.3).

studied recently in [77]. The approach therein consists in modelling X_t as

$$X_t = s_t \rho^* X_{t-1} + \eta_t, \quad \text{for} \quad t \geq 1, \tag{3.16}$$

given initial values $X_0 = \eta_0$, $s_0 = 0$, and where $\rho^* > 1$, (η_t) is a finite variance i.i.d. sequence and s_t is a $0-1$ Bernoulli taking value 1 with probability $p \in (0,1)$. The stable anticipative AR(1) (3.1) is reminiscent of (3.16) in two aspects. On the one hand, it is the unique solution of the linear recursive equation $X_t = \rho^* X_{t-1} + \varepsilon_t^*$ with explosive AR coefficient $\rho^* = 1/\rho > 1$. On the other hand, Corollary 3.2.2 shows that the anticipative AR(1) also behaves as a two-point conditional distribution during bubble episodes.

3.2.3 Continuous time: the anticipative α-stable Ornstein-Uhlenbeck

Financial applications are often inclined towards continuous-time representations and efforts are deployed to advance discrete- and continuous-time techniques side-by-side, including when it comes to bubble modelling [27]. A continuous time analogue of the AR(1) is the well-known Ornstein-Uhlenbeck process. When it is driven by a Brownian motion (α-stable Lévy process with $\alpha = 2$), it is the only continuous in probability stationary Markov Gaussian process. However, when driven by an α-stable Lévy process with $0 < \alpha < 2$, at least two distinct processes arise that are continuous in probability, stationary and Markov : the direct time OU and its reverse time counterpart (see Chapter 3 Section 6 in [117][8]).

Let us first introduce the objects upon which continuous time α-stable moving averages are defined. We borrow from the very concise introduction in [81]. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the underlying probability space and $L_0(\Omega)$ be the set of all real random variables defined on it. Let also $(E, \mathcal{E}, m)$ be an arbitrary measurable space, $\beta : E \longrightarrow [-1, 1]$ be a measurable function and define the set $\mathcal{E}_0 = \{A \in \mathcal{E} : m(A) < +\infty\}$.

Definition 3.2.2 *An independently scattered σ-additive set function $M : \mathcal{E}_0 \longrightarrow L_0(\Omega)$ such that for each $A \in \mathcal{E}_0$*

$$M(A) \sim \mathcal{S}\left(\alpha, \frac{\int_A \beta(x) m(dx)}{m(A)}, (m(A))^{1/\alpha}, 0\right)$$

is called α-stable random measure on $(E, \mathcal{E})$ with control measure m and skewness intensity β.

Independent scatteredness means that for any disjoints sets $A_1, A_2, \ldots, A_n \in \mathcal{E}_0$, $n \in \mathbb{N}$, the random variables $M(A_1), \ldots, M(A_n)$ are independent. One can consider random processes of the form

$$X_t = \int_E f(x-t) M(dx), \quad t \in \mathbb{R}, \tag{3.17}$$

where $f : E \longrightarrow \mathbb{R}$ is a measurable function such that $\int_E |f(x)|^\alpha m(dx) < +\infty$ and in the case $\alpha = 1$, additionally, $\int_E |f(x)\beta(x)| \ln |f(x)| m(dx) < +\infty$. As underlined in [81], the integral in (3.17) is constructed in the natural way by approximating the function f by simple functions in Chapter 3 Section 4 in [117].

We will focus on random processes for which $E = \mathbb{R}$ and m is the Lebesgue measure.

[8] Two-sided OU processes are also mentioned in [21, 128], where it is noticed they admit higher conditional than marginal moments. Anticipative stable OU are also alluded to in [23].

Definition 3.2.3 *Let $\lambda > 0$ and M be an α-stable random measure with Lebesgue control measure and constant skewness intensity $\beta \in [-1, 1]$. The non-anticipative and anticipative α-stable Ornstein-Uhlenbeck, denoted X_{na} and X_a respectively, are defined as*

$$X_{na}(t) = \int_{-\infty}^{t} e^{-\lambda(t-x)} M(dx), \quad t \in \mathbb{R}, \tag{3.18}$$

$$X_a(t) = \int_{t}^{+\infty} e^{-\lambda(x-t)} M(dx), \quad t \in \mathbb{R}. \tag{3.19}$$

Remark 3.2.5 The non-anticipative and anticipative α-stable OU are Markov processes. Indeed, for $s < t$, $X_{na}(t) - e^{-\lambda(t-s)} X_{na}(s) = \int_s^t e^{-\lambda(t-x)} M(dx)$ and $X_a(s) - e^{-\lambda(t-s)} X_a(t) = \int_s^t e^{-\lambda(x-s)} M(dx)$. By Theorem 3.5.3 in [117], we have the independence between $X_{na}(t) - e^{-\lambda(t-s)} X_{na}(s)$ and the σ-algebra generated by $\{X_{na}(u), u \leq s\}$ on the one hand, and between $X_a(s) - e^{-\lambda(t-s)} X_a(t)$ and the σ-algebra generated by $\{X_a(u), u \leq s\}$ on the other hand.

Close to this framework, generalised OU processes driven by Lévy processes (not necessarily stable) are also defined in integral forms and studied in [95]. In [97], these Lévy-driven OU are pointed out to be solutions to stochastic differential equations (SDE) of the form $dV_t = V_{t-} dU_t + dL_t$, where (U, L) is a bivariate Lévy process. It was moreover shown in [9] that the latter SDE may admit anticipative solutions.

The two definitions of the OU processes in (3.18) and (3.19) are very practical in our context as they can be readily embedded in the bivariate α-stable vector framework. Similarly to the discrete time case, we will consider for any $t \in \mathbb{R}$ and $h > 0$ the vectors $(X_i(t), X_i(t+h))$, for $i = a, na$. Just as for the α-stable non-anticipative AR(1), the non-anticipative OU does not feature more moments than the marginal distribution, namely $\mathbb{E}[|X_{na}(t+h)|^p | X_{na}(t)] = +\infty$ whenever $p \geq \alpha$. It displays infinite variance, and the expectation is also ill-defined when $0 < \alpha \leq 1$. On the contrary, the anticipative OU features conditional moments up to $2\alpha + 1$. From now on, we shall focus solely on the anticipative OU, hence we drop the subscript «a» and simply denote the process satisfying Equation (3.19) as X_t, for $t \in \mathbb{R}$. The next Lemma shows that, just as for the discrete time counterpart of the anticipative OU, the spectral measure of (X_t, X_{t+h}) is concentrated on either two or four points of the unit circle.

Proposition 3.2.3 *Let $\{X_t, t \in \mathbb{R}\}$ be the anticipative α-stable OU process defined by (3.19) with $\lambda > 0$ and M an α-stable random measure with Lebesgue control measure and constant skewness intensity $\beta \in [-1, 1]$. Then, for any $h \in \mathbb{R}_+^*$, (X_t, X_{t+h}) is α-stable and its spectral representation, denoted $(\Gamma_h, \boldsymbol{\mu}^0)$, with $\boldsymbol{\mu}^0 = (\mu_1^0, \mu_2^0)$, is such that*

$$\Gamma_h = \frac{1}{\alpha\lambda} \sum_{\vartheta \in S_1} \frac{1 + \vartheta\beta}{2} \left[\left(1 - e^{-\alpha\lambda h}\right) \delta_{\{(\vartheta, 0)\}} + \left(1 + e^{-2\lambda h}\right)^{\alpha/2} \delta_{\{\vartheta \boldsymbol{s}_h\}} \right],$$

with $\boldsymbol{s}_h = \dfrac{(e^{-\lambda h}, 1)}{\sqrt{1 + e^{-2\lambda h}}}$. Moreover, if $\alpha \neq 1$, then $\boldsymbol{\mu}^0 = (0, 0)$, and if $\alpha = 1$ then,

$$\mu_1^0 = \bar{\mu} + \frac{2}{\lambda\pi}\beta, \qquad \mu_2^0 = e^{\lambda h}\bar{\mu} + \frac{2}{\lambda\pi}\beta(1 + \lambda h),$$

where $\bar{\mu} = -\dfrac{\beta}{\lambda\pi} e^{-\lambda h} \ln(1 + e^{2\lambda h})$.

The following Theorem summarises the previous considerations and gives the expressions of the conditional moments in the case $\alpha \neq 1$. The case $\alpha = 1$ has been singled out in Proposition 3.6.8 for expository purposes.

Theorem 3.2.2 *Let $\{X_t, t \in \mathbb{R}\}$ be the anticipative α-stable OU process, $\alpha \neq 1$, defined by* (3.19) *with $\lambda > 0$ and M an α-stable random measure with Lebesgue control measure and constant skewness intensity $\beta \in [-1, 1]$. Then, for any $h \in \mathbb{R}_+^*$, the following hold*

ι) The anticipative α-stable Ornstein-Uhlenbeck is a Markov process.

$\iota\iota$) If $\quad 0 \leq \gamma < 2\alpha + 1, \quad$ then, $\quad \mathbb{E}\Big[|X_{t+h}|^\gamma \Big| X_t\Big] < +\infty.$

$\iota\iota\iota$) The first four moments of $X_{t+h}|X_t$, when they exist, are given by Theorem 3.2.1 with

$$\sigma_1^\alpha = \frac{1}{\alpha\lambda}, \qquad \beta_1 = \beta, \qquad \kappa_p = e^{-\lambda h(\alpha - p)}, \qquad \lambda_p = \beta\kappa_p, \qquad \textit{for} \quad p \in \{1, 2, 3, 4\}.$$

The expressions of the conditional moments simplify during explosive/bubble events.

Corollary 3.2.3 *Let $\{X_t, t \in \mathbb{R}\}$ be the anticipative α-stable OU as defined in Theorem 3.2.2. If $|\beta_1| = 1$, let $\beta_1 x \to +\infty$, and if $|\beta_1| \neq 1$, let $x \to \pm\infty$.*[9] *Also, let $s = 1$ if $x \to +\infty$ and $s = -1$ if $x \to -\infty$. Then, for any $h \in \mathbb{R}_+^*$,*

$$\begin{aligned}
\mu(x, h) &\sim (e^{\lambda h} x) e^{-\alpha\lambda h}, && \textit{if} \quad \alpha \in (0, 2),\\
\sigma^2(x, h) &\sim (e^{\lambda h} x)^2 e^{-\alpha\lambda h}(1 - e^{-\alpha\lambda h}), && \textit{if} \quad \alpha \in (1/2, 2),\\
\gamma_1(x, h) &\longrightarrow s\frac{1 - 2e^{-\alpha\lambda h}}{\sqrt{e^{-\alpha\lambda h}(1 - e^{-\alpha\lambda h})}}, && \textit{if} \quad \alpha \in (1, 2),\\
\gamma_2(x, h) &\longrightarrow \frac{1}{e^{-\alpha\lambda h}} + \frac{1}{1 - e^{-\alpha\lambda h}} - 6, && \textit{if} \quad \alpha \in (3/2, 2),
\end{aligned}$$

where the left-hand side quantities are defined in (3.12)-(3.15).

Remark 3.2.6 Echoing Remark 3.2.3, the anticipative OU behaves as its discrete time counterpart in that $X_{t+h}|X_t = x$, as x becomes large, can be interpreted as a distribution charging probability $e^{-\alpha\lambda h}$ to the value $e^{\lambda h}x$ and probability $1 - e^{-\alpha\lambda h}$ to 0. Focusing on the limiting behaviour of the conditional kurtosis, it can be easily seen that the function $h \longmapsto \dfrac{1}{e^{-\alpha\lambda h}} + \dfrac{1}{1 - e^{-\alpha\lambda h}} - 6$ is strictly convex and diverges to infinity as $h \to +\infty$, but also as $h \to 0$, illustrating that the paths of the anticipative OU are continuous only in probability. It reaches its global minimum at h_0 such that $e^{-\alpha\lambda h_0} = 1/2$, yielding $h_0 = \dfrac{\ln 2}{\alpha\lambda}$, and takes value -2 corresponding to the lowest achievable excess kurtosis amongst all probability distributions. Last, the horizon h_0 achieving the minimum is further away in the future for heavier-tailed and more persistent processes.

3.3 Aggregated anticipative AR(1)

Heavy-tailed anticipative AR processes generate trajectories that feature locally explosive phenomena such as financial bubbles. The higher the order of the AR process, the more complex patterns it is able to mimic

[9] See Remark 3.6.3 for details regarding the different behaviours when $|\beta_1| \neq 1$ and $|\beta_1| = 1$.

(see [51] for some examples). However, a given AR(p) process is constrained by the fact that it is specific to one particular explosive pattern which occurs recurrently through time. It is proposed in [63] to consider processes resulting from the aggregation of multiple AR(1) with different autoregressive coefficients. More formally, a process from this family can be defined by

$$X_t = c\sum_{j=1}^{J} \pi_j X_{j,t}, \quad X_{j,t} = \rho_j X_{j,t+1} + \varepsilon_{j,t}, \quad 0 < |\rho_j| < 1, \quad j = 1, \ldots, m \tag{3.20}$$

where $c > 0$, $\pi_j \in (0, 1)$ for any j, $\sum_{j=1}^{J} \pi_j = 1$ and $(\varepsilon_{j,t})_{t\in\mathbb{Z}} \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, \beta_j, \sigma_j, 0)$ are mutually independent sequences of i.i.d. noise. Process (X_t) will generate explosive/bubble episodes with rates of increase $1/\rho_j$. Unlike the latent $X_{j,t}$'s however, it is not a Markov process, and nothing is known about the predictive distribution of X_{t+h} given its past. We now give results regarding the conditional distribution of X_{t+h} given X_t, by first noticing that (X_t, X_{t+h}) can also be embedded in the multivariate α-stable framework. For $j = 1, \ldots, J$, denote $(\Gamma_{j,h}, \boldsymbol{\mu}_j^0)$, $\boldsymbol{\mu}_j^0 = (\mu_{1,j}^0, \mu_{2,j}^0)$ the spectral representation of $(X_{j,t}, X_{j,t+h})$ given by Proposition 3.2.2. For each $j = 1, \ldots, J$, denote also $\sigma_{1,j}$, $\beta_{1,j}$, $\kappa_{p,j}$ and $\lambda_{p,j}$ the quantities defined at Theorem 3.2.1 where ρ, σ and β are replaced by ρ_j, σ_j and β_j.

Lemma 3.3.1 *Let (X_t) be defined according to (3.20) with $0 < \alpha < 2$. Then, for any $h \geq 1$, (X_t, X_{t+h}) is a bivariate α-stable vector and its spectral representation, denoted $(\Gamma_h, \boldsymbol{\mu}^0)$ with $\boldsymbol{\mu}^0 = (\mu_1^0, \mu_2^0)$, is such that*

$$\Gamma_h = c^\alpha \sum_{j=1}^{J} \pi_j^\alpha \Gamma_{j,h},$$

and,

$$\mu_1^0 = c\sum_{j=1}^{J} \pi_j \Big(\mu_{1,j}^0 - \mathbb{1}_{\{\alpha=1\}} \frac{2}{\pi} \sigma_{1,j}\beta_{1,j} \ln|c\pi_j|\Big), \qquad \mu_2^0 = c\sum_{j=1}^{J} \pi_j \Big(\mu_{2,j}^0 - \mathbb{1}_{\{\alpha=1\}} \frac{2}{\pi} \sigma_{1,j}\lambda_{1,j} \ln|c\pi_j|\Big).$$

The techniques used in the previous sections are therefore available here as well and we are able to characterise the moments of X_{t+h} given X_t. As previously, we provide here the moments for $\alpha \neq 1$, the remaining case being given in Proposition 3.6.9.

Proposition 3.3.1 *Let (X_t) be defined according to (3.20) with $0 < \alpha < 2$. Let $h \geq 1$.*
ι) If $\gamma < 2\alpha + 1$, then $\mathbb{E}\Big[|X_{t+h}|^\gamma \Big| X_t = x\Big] < +\infty$.
$\iota\iota$) The first four moments of $X_{t+h}|X_t$, when they exist, are given by Theorem 3.2.1 with

$$\sigma_1^\alpha = c^\alpha \sum_{j=1}^{J} \pi_j^\alpha \sigma_{1,j}^\alpha, \qquad \beta_1 = \mathbb{E}(B), \qquad \kappa_p = \mathbb{E}(K_p), \qquad \lambda_p = \mathbb{E}(L_p), \qquad \textit{for} \quad p \in \{1, 2, 3, 4\}$$

where B, K_p and L_p are discrete random variables such that $\mathbb{P}\Big((B, K_p, L_p) = (\beta_{1,j}, \kappa_{p,j}, \lambda_{p,j})\Big) = w_j$ and $w_j = \dfrac{\pi_j^\alpha \sigma_{1,j}^\alpha}{\sum_{i=1}^{J} \pi_i^\alpha \sigma_{1,i}^\alpha}$ for $j = 1, \ldots, J$.

Proof. ι) From Lemma 3.3.1, we know that the spectral measure of (X_t, X_{t+h}) writes $\Gamma_h = c^\alpha \sum_{j=1}^{J} \pi_j^\alpha \Gamma_{j,h}$, for $0 < \alpha < 2$, where the $\Gamma_{j,h}$'s are the spectral measures of $(X_{j,t}, X_{j,t+h})$, with the $X_{j,t}$'s being simple

AR(1) processes. We know by Lemma 3.2.1 that for any j, any h and any $\nu \geq 0$,

$$\int_{S_2} |s_1|^{-\nu}\Gamma_{j,h}(ds) < +\infty.$$

Hence, for any $\nu \geq 0$,

$$\int_{S_2} |s_1|^{-\nu}\Gamma_h(ds) = c^\alpha \sum_{j=1}^{J} \pi_j^\alpha \int_{S_2} |s_1|^{-\nu}\Gamma_{j,h}(ds) < +\infty.$$

The existence of conditional moments follows from Proposition 3.2.1.

$\iota\iota$) The form of the conditional moments follow from Theorems 3.6.1, 3.6.3, 3.6.5 and 3.6.6. The parameters of the X_j's are obtained by first noticing that,

$$\sigma_1^\alpha = \int_{S_2} (s_2/s_1)^p |s_1|^\alpha \Gamma_h(d\boldsymbol{s}) = c^\alpha \sum_{j=1}^{J} \pi_j^\alpha \int_{S_2} (s_2/s_1)^p |s_1|^\alpha \Gamma_{j,h}(d\boldsymbol{s}) = c^\alpha \sum_{j=1}^{J} \pi_j^\alpha \sigma_{1,j}^\alpha.$$

And thus, for instance,

$$\kappa_p = \frac{1}{\sigma_1^\alpha} \int_{S_2} (s_2/s_1)^p |s_1|^\alpha \Gamma_h(d\boldsymbol{s}) = \frac{c^\alpha}{\sigma_1^\alpha} \sum_{j=1}^{J} \pi_j^\alpha \int_{S_2} (s_2/s_1)^p |s_1|^\alpha \Gamma_{j,h}(d\boldsymbol{s}) = \sum_{j=1}^{J} \frac{\pi_j^\alpha \sigma_{1,j}^\alpha}{\sum_{i=1}^{J} \pi_i^\alpha \sigma_{1,i}^\alpha} \kappa_{p,j}. \qquad \square$$

Remark 3.3.1 For the non-aggregated anticipative AR(1) considered at Section 3.2.2, linearity of the conditional expectation occurs when $\rho > 0$. However, assuming $\rho_j > 0$ for $j = 1, \ldots, J$ for the aggregated process X_t does not guarantee linearity in general. Indeed in Proposition 3.6.1, linearity is achieved if and only if $\lambda_1 - \beta_1\kappa_1 = 0$, which is equivalent to

$$\mathrm{Cov}\big(B, K_1\big) + \mathbb{E}\Big[B\big(|K_1| - K_1\big)\Big] = 0,$$

since $L_1 = B|K_1|$. Hence, if $\rho_j > 0$ for $j = 1, \ldots, J$, then $K_1 > 0$ a.s. and this condition becomes

$$\mathrm{Cov}\big(B, K_1\big) = 0.$$

Remark 3.3.2 It is easy to construct examples for which $\mathbb{E}\Big[X_{j,t+h}\Big|X_{j,t} = x\Big]$ are all linear in x for any j and h, and yet such that $y \longmapsto \mathbb{E}\Big[X_{t+h}\Big|X_t = y\Big]$ is a non-linear function of y. In view of the previous Remark, this can be achieved by taking for instance $J = 2$, $\rho_1 = \beta_1 = 0.1$ and $\rho_2 = \beta_2 = 0.9$ in (3.20).

3.4 A higher bound for the moments of $X_3|X_2, X_1$

To the best of our knowledge, Proposition 3.2.1 is the only result quantifying up to which order the conditional moments of a stable random vector may exist.[10] It is however restricted to the bivariate framework and whether this bound holds for higher dimension of the conditioning space is unknown. In this section, we take advantage both of the Markov property of anticipative AR(2) processes as shown in [51] and the result

[10] Sufficient conditions for the finiteness of the conditional variance are also known in higher dimensions (see [49] for instance) but do not tell anything about higher, possibly fractional, orders.

of Proposition 3.2.1 to show that a higher sufficient bound may hold when the dimension of conditioning is at least 2. Let (X_t) be the strictly stationary solution of

$$X_t = \psi_1 X_{t+1} + \psi_2 X_{t+2} + \varepsilon_t, \quad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, \beta, \sigma, 0), \tag{3.21}$$

where $\psi(z) := 1 - \psi_1 z - \psi_2 z^2 = (1 - a_1 z)(1 - a_2 z)$ for some real numbers a_1, a_2 such that $0 < |a_i| < 1$ for $i = 1, 2$. We exclude the uninteresting case $\psi_1 = 0$ since it implies that $\{X_{2t}, t \in \mathbb{Z}\}$ and $\{X_{2t+1}, t \in \mathbb{Z}\}$ are independent AR(1) processes. Under these conditions, X_t admits the moving average representation $X_t = \sum_{k\geq 0} d_k \varepsilon_{t+k}$, with $d_k = (a_1^{k+1} - a_2^{k+1})/(a_1 - a_2)$, if $a_1 \neq a_2$, and $d_k = (k+1)a^k$, if $a_1 = a_2 = a$.

Proposition 3.4.1 *Let X_t be the anticipative strictly stationary solution of* (3.21) *with* $0 < \alpha < 2$. *Then,*

$$\mathbb{E}\Big[|X_t|^\gamma \Big| X_{t-1}, X_{t-2}\Big] < +\infty, \quad \textit{a.s. for any} \quad 0 \leq \gamma < 3\alpha + 2. \tag{3.22}$$

Remark 3.4.1 Proposition 3.4.1 in particular demonstrates that for some α-stable random vectors (X_1, X_2, X_3), the moments of $X_3|X_2, X_1$ may exist up to order $3\alpha + 2 \in (2, 8)$. Obtaining bounds such as the latter and the one of Proposition 3.2.1 for general α-stable random vectors (X_1, X_2, X_3) is particularly delicate. Attempting a proof as in [30, 33] would require the sixth derivative of the characteristic function of $X_3|X_2, X_1$, knowing that in the bivariate case, the fourth derivative is already a sum of more than 20 terms requiring a two-page classification.

Proof. For any $(x_0, x_1, x_2) \in \mathbb{R}^3$,

$$f_{X_t|(X_{t+1},X_{t+2})=(x_1,x_2)}(x_0) = f_\varepsilon(x_0 - \psi_1 x_1 - \psi_2 x_2),$$

because ε_t is independent from X_{t+1}, X_{t+2}. By the Bayes formula,

$$f_{X_t|(X_{t+1},X_{t+2})=(x_1,x_2)}(x_0) = \frac{f_{X_{t+2}|(X_t,X_{t+1})=(x_0,x_1)}(x_2)}{f_{X_{t+1},X_{t+2}}(x_1,x_2)} f_{X_t,X_{t+1}}(x_0,x_1).$$

Thus,

$$f_{X_{t+2}|(X_t,X_{t+1})=(x_0,x_1)}(x_2) = \frac{f_\varepsilon(x_0 - \psi_1 x_1 - \psi_2 x_2) f_{X_{t+2}|X_{t+1}=x_1}(x_2) f_{X_{t+1}}(x_1)}{f_{X_t,X_{t+1}}(x_0,x_1)}.$$

On the one hand, when $|x_2| \to +\infty$,

$$f_\varepsilon(x_0 - \psi_1 x_1 - \psi_2 x_2) = O(|x_2|^{-\alpha-1}),$$

thus, for any $\gamma > 0$,

$$|x_2|^\gamma f_{X_{t+2}|(X_t,X_{t+1})=(x_0,x_1)}(x_2) \underset{|x_2|\to+\infty}{=} O\Big(|x_2|^{\gamma-\alpha-1} f_{X_{t+2}|X_{t+1}=x_1}(x_2)\Big). \tag{3.23}$$

On the other hand, we will show that $x \longmapsto |x|^r f_{X_{t+2}|X_{t+1}=x_1}(x)$ is integrable on $\mathbb{R}$ for any $r < 2\alpha + 1$, from which the conclusion will follow.

The integrability of the later function is equivalent to the finiteness of $\mathbb{E}\Big[|X_t|^r \Big| X_{t-1} = x_1\Big]$ which we will show using Proposition 3.2.1. From Lemma 3.7.1, (X_t, X_{t+1}) is α-stable and

$$\int_{S_2} |s_1|^{-\nu}\Gamma(d\boldsymbol{s}) = \sigma^\alpha \left(1 + \sum_{k=1}^{+\infty} |d_k|^{-\nu}(d_k^2 + d_{k-1}^2)^{(\alpha+\nu)/2}\right).$$

Given the form of the coefficients d_k's for X_t satisfying (3.21), we have for k large enough say $|d_k| \sim C(k)|a|^k$ where $|a| \in (0,1)$ and C is a polynomial with degree 0 or 1. It is easy to see that $|d_{k-1}/d_k| \longrightarrow \ell$, for some $\ell \geq 0$. Hence,

$$|d_k|^{-\nu}(d_k^2 + d_{k-1}^2)^{(\alpha+\nu)/2} = |d_k|^\alpha(1 + (d_{k-1}/d_k)^2)^{(\alpha+\nu)/2} \sim C(k)^\alpha |a|^{\alpha k}(1+\ell^2)^{(\alpha+\nu)/2},$$

which is the term of an absolutely convergent series for any $\nu \geq 0$. Thus, $\int_{S_2} |s_1|^{-\nu}\Gamma(d\boldsymbol{s}) < +\infty$ for all $\nu \geq 0$ and we conclude invoking Proposition 3.2.1. □

3.5 Concluding remarks

Our results constitute a first step towards a quantification of the odds of crashes of bubbles, which could be valuable for risk/portfolio managers and regulators. Specifically in a portfolio allocation context, where managers would decide both the composition of their portfolios and when to pull out from speculative assets for instance, the functional forms per se of the higher order moments could be valuable [65, 80]. Our results also open the possibility for alternative point predictors for the stable anticipative AR(1) and OU processes that exploit higher order conditional moments, as opposed to other predictors that were proposed to circumvent the infinite variance of α-stable processes, such as as minimum L^α*-dispersion* or maximum *covariation* ([81] and the references therein).

3.6 Complementary results

This section is composed of three subsections. The first provides the form of the conditional moments for arbitrary bivariate stable vectors (X_1, X_2). The second subsection completes Theorem 3.2.1 and Propositions 3.2.3, 3.4.1 in the case $\alpha = 1$. The third provides an interpretation of the quantity $\rho^{\alpha h}$ in Corollary 3.2.2 using point processes.

3.6.1 Conditional moments of bivariate α-stable random vectors

The conditional moments stated in Theorem 3.2.1 for the particular AR(1) case originate from the broader bivariate α-stable framework that was much studied in a series of papers in the 90s [30, 31, 32, 33, 64, 116, 117, 128] (see also [21, 23, 49, 103]). In this section, we give formulae for the conditional moments up to order four of arbitrary (not necessarily symmetric) α-stable bivariate vectors (X_1, X_2), that is, up to the maximum admissible integer order under the most favourable dispositions of Proposition 3.2.1. Only the first and second order moments received attention in the literature, the later besides, mostly in the Symmetric α-Stable (SαS) case. The conditional expectation of arbitrary α-stable bivariate vectors is the most comprehensively understood (see for instance [64, 116]. See also [23]). We suppose in the rest of this section that the shift vector $\boldsymbol{\mu}^0 = (\mu_1^0, \mu_2^0)$ is zero. This can be done without loss of generality because, assuming the conditional moment of order p exists,

$$\begin{aligned}\mathbb{E}\Big[X_2^p\Big|X_1 = x\Big] &= \mathbb{E}\Big[(X_2 - \mu_2^0 + \mu_2^0)^p\Big|X_1 - \mu_1^0 = x - \mu_1^0\Big]\\ &= \sum_{j=0}^{p} \mathrm{C}_p^j(\mu_2^0)^{p-j}\mathbb{E}\Big[\tilde{X}_2^j\Big|\tilde{X}_1 = \tilde{x}\Big],\end{aligned}$$

where $\tilde{x} = x - \mu_1^0$, and $(\tilde{X}_1, \tilde{X}_2) = (X_1 - \mu_1^0, X_2 - \mu_2^0)$ has the same spectral measure as (X_1, X_2) and zero shift parameter. For α-stable bivariate vectors with arbitrary spectral measure Γ, the constants of Theorem 3.2.1 will be replaced by the following quantities

$$\sigma_1^\alpha = \int_{S_2} |s_1|^\alpha \Gamma(d\boldsymbol{s}), \qquad \beta_1 = \frac{\int_{S_2} s_1^{<\alpha>}\Gamma(d\boldsymbol{s})}{\sigma_1^\alpha}, \tag{3.24}$$

$$\kappa_p = \frac{\int_{S_2}(s_2/s_1)^p|s_1|^\alpha\Gamma(d\boldsymbol{s})}{\sigma_1^\alpha}, \qquad \lambda_p = \frac{\int_{S_2}(s_2/s_1)^p s_1^{<\alpha>}\Gamma(d\boldsymbol{s})}{\sigma_1^\alpha}, \tag{3.25}$$

for $p \in \{1, 2, 3, 4\}$, when they exist. We will assume $\sigma_1^\alpha > 0$ so that the random variable X_1 is not degenerate. Notice that κ_1 is also known as the covariation of two stable random variables in the literature. We start with the conditional expectation in the case $\alpha \neq 1$.

Theorem 3.6.1 (Samorodnistky and Taqqu (Theorem 5.2.2, 1994)) *Let (X_1, X_2) be α-stable, $\alpha \in (0, 2) \setminus \{1\}$, with spectral representation $(\Gamma, \mathbf{0})$. If $0 < \alpha < 1$, let Γ satisfy (3.3) for some $\nu > 1 - \alpha$. Then, for almost every x,*

$$\mathbb{E}\Big[X_2\Big|X_1 = x\Big] = \kappa_1 x + \frac{a(\lambda_1 - \beta_1\kappa_1)}{1 + a^2\beta_1^2}\left[a\beta_1 x + \frac{1 - xH(x)}{\pi f_{X_1}(x)}\right], \tag{3.26}$$

where $a = tg\left(\frac{\pi\alpha}{2}\right)$ and σ_1, β_1, the κ_p's and the λ_p's are as in (3.24) and (3.25).

If $\alpha < 1$ and $\beta_1 = 1$, Relation (3.26) is well defined only for $x \geq 0$, and if $\alpha < 1$ and $\beta_1 = -1$, it is well defined only for $x \leq 0$.

The conditional expectation in the case $\alpha = 1$ has also been considered in the literature and is more intricate.

Theorem 3.6.2 (Samorodnistky and Taqqu (Theorem 5.2.3, 1994)) *Let (X_1, X_2) be α-stable with $\alpha = 1$ and spectral representation $(\Gamma, \mathbf{0})$ satisfying (3.3) with $\nu > 0$. Then, for almost every x,*

$$\mathbb{E}\Big[X_2\Big|X_1 = x\Big] = -\frac{2\sigma_1}{\pi}q_0 + \kappa_1(x - \mu_1) + \frac{\lambda_1 - \beta_1\kappa_1}{\beta_1}\left[(x - \mu_1) - \sigma_1\frac{U(x)}{\pi f_{X_1}(x)}\right], \tag{3.27}$$

if $\beta_1 \neq 0$, and

$$\mathbb{E}\Big[X_2\Big|X_1 = x\Big] = -\frac{2\sigma_1}{\pi}q_0 + \kappa_1(x - \mu_1) - \frac{2\sigma_1}{\pi}\lambda_1\frac{V(x)}{\pi f_{X_1}(x)}, \tag{3.28}$$

if $\beta_1 = 0$. Here $a = 2/\pi$, σ_1, β_1, the κ_p's and the λ_p's are as in (3.24) *and* (3.25), *and*

$$U(x) = \int_0^{+\infty} e^{-\sigma_1 t} \sin\Big(t(x-\mu_1) + a\sigma_1\beta_1 t \ln t\Big) dt,$$
$$V(x) = \int_0^{+\infty} e^{-\sigma_1 t}(1+\ln t) \cos\Big(t(x-\mu_1) + a\sigma_1\beta_1 t \ln t\Big) dt,$$
$$q_0 = \frac{1}{\sigma_1}\int_{S_2} s_2 \ln|s_1|\Gamma(d\boldsymbol{s}), \qquad \mu_1 = -a\int_{S_2} s_1 \ln|s_1|\Gamma(d\boldsymbol{s}).$$

If $\alpha < 1$ and $\beta_1 = 1$ (resp. $\beta_1 = -1$), Relation (3.26) *is well defined only for $x \geq 0$ (resp. $x \leq 0$).*

Regarding the conditional variance, studies have focused most exclusively on the SαS case (see [21, 49, 128]). One notable exception is Theorem 3.1 in [31] which states without proof the functional form of the conditional variance for an arbitrary, skewed bivariate α-stable vector for $\alpha \neq 1$. We therefore provide a proof for the second moment as well and fill the gap for $\alpha = 1$. We start with the case $\alpha \neq 1$.

Theorem 3.6.3 *Let (X_1, X_2) be α-stable, $\alpha \in (1/2, 2) \setminus \{1\}$, with spectral representation $(\Gamma, \mathbf{0})$, where Γ satisfies* (3.3) *with $\nu > 2 - \alpha$. Then, for almost every x,*

$$\mathbb{E}\Big[X_2^2 \Big| X_1 = x\Big] = \kappa_2 x^2 + \frac{ax(\lambda_2 - \beta_1\kappa_2)}{1+(a\beta_1)^2}\left[a\beta_1 x + \frac{1 - xH(x)}{\pi f_{X_1}(x)}\right] - \frac{\alpha^2\sigma_1^{2\alpha}}{\pi f_{X_1}(x)}\mathcal{H}\Big(2, \boldsymbol{\theta}_1; x\Big), \tag{3.29}$$

where $a = tg\left(\frac{\pi\alpha}{2}\right)$, $\boldsymbol{\theta}_1 = (\theta_{11}, \theta_{12})$ with

$$\theta_{11} = \kappa_1^2 - a^2\lambda_1^2 + a^2\beta_1\lambda_2 - \kappa_2, \qquad \theta_{12} = a(\lambda_2 + \beta_1\kappa_2) - 2a\lambda_1\kappa_1,$$

and σ_1, β_1, the κ_p's and the λ_p's are as in (3.24) *and* (3.25).
If $\alpha < 1$ and $\beta_1 = 1$ (resp. $\beta_1 = -1$), Relation (3.29) *is well defined only for $x \geq 0$ (resp. $x \leq 0$).*

We now give the formulae for the second conditional moment when $\alpha = 1$. As for the conditional expectation when (X_1, X_2) is *not S1S*, two different results hold according to whether the marginal distribution of X_1 is skewed or symmetric.

Theorem 3.6.4 *Let (X_1, X_2) be α-stable, with $\alpha = 1$ and spectral representation $(\Gamma, \mathbf{0})$, where Γ satisfies* (3.3) *with $\nu > 1$. Then, for almost every x,*

$$\begin{aligned}\mathbb{E}\Big[X_2^2 \Big| X_1 = x\Big] &= \sigma_1^2(a^2q_0^2 - \kappa_1^2) + \frac{2\sigma_1\lambda_1}{\beta_1}\Big(\sigma_1\kappa_1 - aq_0(x-\mu_1)\Big) + \frac{\lambda_2}{\beta_1}\Big((x-\mu_1)^2 - \sigma_1^2\Big) \\ &\quad + \Big(a\sigma_1 q_0(\lambda_1 - \beta_1\kappa_1) + (\kappa_1\lambda_1 - \lambda_2)(x-\mu_1)\Big)\frac{2\sigma_1 U(x)}{\beta_1\pi f_{X_1}(x)} \\ &\quad + \Big(\lambda_2 + \beta_1\kappa_2 - 2\kappa_1\lambda_1 + a^2\sigma_1\beta_1(\lambda_1^2 - \beta_1\lambda_2)W(x)\Big)\frac{\sigma_1}{\beta_1\pi f_{X_1}(x)},\end{aligned}$$

if $\beta_1 \neq 0$, and

$$\begin{aligned}\mathbb{E}\Big[X_2^2 \Big| X_1 = x\Big] &= \sigma_1^2(\kappa_2 + a^2q_0^2 - \kappa_1^2) - 2a\sigma_1\kappa_1 q_0(x-\mu_1) + \kappa_2(x-\mu_1)^2 \\ &\quad + a\sigma_1(\lambda_2 - 2\lambda_1\kappa_1)\frac{F_{X_1}(x) - 1/2}{f_{X_1}(x)} + \frac{a\sigma_1\lambda_1}{\pi f_{X_1}(x)}\left[2\Big(a\sigma_1 q_0 - \kappa_1(x-\mu_1)\Big)V(x) + a\sigma_1\lambda_1 W(x)\right],\end{aligned}$$

if $\beta_1 = 0$. *Here,* $a = 2/\pi$, σ_1, β_1, *the* κ_p*'s and the* λ_p*'s are as in* (3.24) *and* (3.25), U, V, q_0 *and* μ_1 *are as in Theorem 3.6.2 and*

$$W(x) = \int_0^{+\infty} e^{-\sigma_1 t}(1+\ln t)^2 \cos\Big(t(x-\mu_1) + a\sigma_1\beta_1 t \ln t\Big)dt.$$

Remark 3.6.1 Note that when $\alpha = 1$, $\pi f_{X_1}(x) = \int_0^{+\infty} e^{-\sigma_1 t} \cos\Big(t(x-\mu_1) + a\sigma_1\beta_1 t \ln t\Big)dt$. If in addition $\beta_1 = 0$, then X_1 is marginally Cauchy distributed and its density and cumulative distribution function are known explicitly.

Remark 3.6.2 The conditional variance when (X_1, X_2) is S1S (derived in [31]) is encompassed by the second statement of the Theorem. Indeed, when (X_1, X_2) is S1S, its spectral measure satisfies $\Gamma(-A) = \Gamma(A)$ for any $A \in S_2$ and it can be shown that $\beta_1 = \mu_1 = q_0 = \lambda_i = 0$, for $i = 1, 2$. This then yields

$$\mathbb{V}\Big(X_2\Big|X_1 = x\Big) = (\kappa_2 - \kappa_1^2)\Big(x^2 + \sigma_1^2\Big).$$

We now provide the analytical form for the third conditional moment.

Theorem 3.6.5 *Let* (X_1, X_2) *be* α*-stable,* $\alpha \in (1, 2)$*, with spectral representation* $(\Gamma, \mathbf{0})$*, where* Γ *satisfies* (3.3) *with* $\nu > \alpha - 3$*. Then, for almost every* x*,*

$$\begin{aligned}\mathbb{E}\Big[X_2^3\Big|X_1 = x\Big] = \kappa_3 x^3 + \frac{ax^2(\lambda_3 - \beta_1\kappa_3)}{1+(a\beta_1)^2}\left[a\beta_1 x + \frac{1 - xH(x)}{\pi f_{X_1}(x)}\right] \\ - \frac{\alpha^2\sigma_1^{2\alpha}}{2\pi f_{X_1}(x)}\left[x\mathcal{H}\Big(2, \boldsymbol{\theta}_2; x\Big) + \alpha\sigma_1^\alpha \mathcal{H}\Big(3, \boldsymbol{\theta}_3; x\Big)\right],\end{aligned} \tag{3.30}$$

where the $\boldsymbol{\theta}_i$*'s are given in* (3.72)-(3.75) *in Section 3.7.12 with* σ_1, β_1, *the* κ_p*'s and the* λ_p*'s as in* (3.24) *and* (3.25).

Finally, under the most favourable dispositions of Proposition 3.2.1, the fourth conditional moment exists and its analytical form is given in the following Theorem.

Theorem 3.6.6 *Let* (X_1, X_2) *be* α*-stable,* $\alpha \in (3/2, 2)$*, with spectral representation* $(\Gamma, \mathbf{0})$*, where* Γ *satisfies* (3.3) *with* $\nu > \alpha - 4$*. Then, for almost every* x*,*

$$\begin{aligned}\mathbb{E}\Big[X_2^4\Big|X_1 = x\Big] = \kappa_4 x^4 + \frac{ax^3(\lambda_4 - \beta_1\kappa_4)}{1+(a\beta_1)^2}\left[a\beta_1 x + \frac{1 - xH(x)}{\pi f_{X_1}(x)}\right] \\ - \frac{\alpha^2\sigma_1^{2\alpha}}{\pi f_{X_1}(x)}\left[\frac{x^2}{2}\mathcal{H}\Big(2, \boldsymbol{\theta}_4; x\Big) + \frac{\alpha x\sigma^\alpha}{6}\mathcal{H}\Big(3, \boldsymbol{\theta}_5; x\Big) + \frac{\alpha^2\sigma^{2\alpha}}{3}\mathcal{H}\Big(4, \boldsymbol{\theta}_6; x\Big)\right],\end{aligned} \tag{3.31}$$

where the $\boldsymbol{\theta}_i$*'s are given in* (3.66)-(3.71) *in Section 3.7.12 with* σ_1, β_1, *the* κ_p*'s and the* λ_p*'s as in* (3.24) *and* (3.25).

The previous expressions of the conditional moments simplify when one considers the asymptotics with respect to the conditioning variable, as $X_1 = x$ becomes large.

Proposition 3.6.1 *Let $p \in \{1,2,3,4\}$ and let (X_1, X_2) be α-stable with $\alpha \in (0,2)$, and spectral representation $(\Gamma, \mathbf{0})$ such that the conditional moment of order p exists. If $|\beta_1| \neq 1$, then*

$$x^{-p}\,\mathbb{E}\Big[X_2^p\Big|X_1 = x\Big] \underset{x\to+\infty}{\longrightarrow} \frac{\kappa_p + \lambda_p}{1+\beta_1},$$
$$x^{-p}\,\mathbb{E}\Big[X_2^p\Big|X_1 = x\Big] \underset{x\to-\infty}{\longrightarrow} \frac{\kappa_p - \lambda_p}{1-\beta_1},$$

and if $|\beta_1| = 1$ and $\beta_1 x \to +\infty$, then,

$$x^{-p}\,\mathbb{E}\Big[X_2^p\Big|X_1 = x\Big] \longrightarrow \kappa_p.$$

Remark 3.6.3 The difference between the cases $|\beta_1| = 1$ and $|\beta_1| \neq 1$ can be seen as a consequence of the different tail behaviours that prevail. When $|\beta_1| \neq 1$, both the left and right tail of the density of X_1 display power law decay as $O(|x|^{-\alpha-1})$. However, when $\beta_1 = -1$ for instance, the distribution of X_1 is said to be *totally skewed to the left.* The left tail still decays as $O(|x|^{-\alpha-1})$, but the right tail decays much faster and another asymptotics holds.[11]

3.6.2 Complementary results for $\alpha = 1$

Theorems 3.2.1, 3.2.2 and Proposition 3.3.1 give the conditional moments of the three considered processes in the case $\alpha \neq 1$. We provide here the remaining more intricate case $\alpha = 1$. As can be seen in Propositions 3.2.2, 3.2.3 and Lemma 3.3.1, the bivariate vectors (X_t, X_{t+h}) of each process are α-stable and their spectral representations display a non-zero shift parameter $\boldsymbol{\mu}^0$. For the sake of simplicity, we cancel this shift by considering the vector $(\tilde{X}_t, \tilde{X}_{t+h}) := (X_t, X_{t+h}) - \boldsymbol{\mu}^0$. We start with the anticipative AR(1).

Theorem 3.6.7 *Let (X_t) be the anticipative α-stable AR(1) solution of (3.1) with $\alpha = 1$, $\beta \in [-1,1]$ and $0 < |\rho| < 1$. Let $h \geq 1$. Then, $\mathbb{E}\Big[\tilde{X}_{t+h}\Big|\tilde{X}_t = x\Big]$ and $\mathbb{E}\Big[\tilde{X}^2_{t+h}\Big|\tilde{X}_t = x\Big]$ are given respectively by Theorem 3.6.2 and 3.6.4 with,*

$$\sigma_1 = \frac{\sigma}{1-|\rho|}, \qquad \beta_1 = \beta\frac{1-|\rho|}{1-\rho}, \qquad \mu_1 = \frac{1}{\pi}\sigma_1\beta_1\rho^h \ln(1+\rho^{-2h}), \qquad q_0 = -\frac{1}{2}\beta_1 \ln(1+\rho^{-2h})$$
$$\kappa_p = |\rho|^h \rho^{-hp}, \qquad \lambda_p = \beta_1 \rho^{h(1-p)}, \qquad \textit{for} \quad p \in \{1,2\}.$$

Theorem 3.6.8 *Let $\{X_t, t \in \mathbb{R}\}$ be the anticipative α-stable OU process, with $\alpha = 1$, defined by (3.19) with $\lambda > 0$ and M an α-stable random measure with Lebesgue control measure and constant skewness intensity*

[11] If $X_1 \sim \mathcal{S}(\alpha, -1, 1, 0)$, and $x \to +\infty$, then by Theorem 5.2.2 in [131]

$$f_{X_1}(x) \sim \frac{(x/\alpha)^{(\alpha-2)/2(\alpha-1)}}{\sqrt{2\pi\alpha|1-\alpha|}} \exp\Big\{ -|1-\alpha|(x/\alpha)^{\alpha/(\alpha-1)}\Big\}, \quad \text{if} \quad \alpha > 1,$$
$$f_{X_1}(x) \sim \frac{1}{\sqrt{2\pi}} \exp\Big\{\frac{x-1}{2} - e^{x-1}\Big\}, \quad \text{if} \quad \alpha = 1.$$

If $\alpha < 1$, the support of f_{X_1} is $\mathbb{R}_-$ and conditioning by $x > 0$ makes no sense. Note however that when $x \to 0$, a formula similar to the case $\alpha > 1, x \to +\infty$ holds.

$\beta \in [-1,1]$. Let $h \in \mathbb{R}_+^*$. Then, $\mathbb{E}\Big[\tilde{X}_{t+h}\Big|\tilde{X}_t = x\Big]$ and $\mathbb{E}\Big[\tilde{X}^2_{t+h}\Big|\tilde{X}_t = x\Big]$ are given respectively by Theorem 3.6.2 and 3.6.4 with,

$$\sigma_1 = \frac{1}{\lambda}, \qquad \beta_1 = \beta, \qquad \mu_1 = \frac{\beta}{\lambda\pi}e^{-\lambda h}\ln(1+e^{2\lambda h}), \qquad q_0 = -\frac{1}{2}\beta\ln(1+e^{2\lambda h})$$
$$\kappa_p = e^{-\lambda h(1-p)}, \qquad \lambda_p = \beta\kappa_p, \qquad \textit{for} \quad p \in \{1,2\}.$$

In addition to $\sigma_{1,j}$, $\beta_{1,j}$, $\kappa_{p,j}$ and $\lambda_{p,j}$, denote for each $j = 1,\dots,J$, the quantities $q_{0,j}$ defined at Theorem 3.6.7 where ρ, σ and β are replaced by ρ_j, σ_j and β_j.

Theorem 3.6.9 *Let (X_t) be the aggregated anticipative AR(1) defined according to (3.20) with $\alpha = 1$. Let $h \geq 1$. Then, $\mathbb{E}\Big[\tilde{X}_{t+h}\Big|\tilde{X}_t = x\Big]$ and $\mathbb{E}\Big[\tilde{X}^2_{t+h}\Big|\tilde{X}_t = x\Big]$ are given respectively by Theorem 3.6.2 and 3.6.4 with,*

$$\sigma_1 = c\sum_{j=1}^{J}\pi_j\sigma_{1,j}, \qquad \beta_1 = \mathbb{E}(B), \qquad \mu_1 = c\sum_{j=1}^{J}\pi_j\mu_{1,j}, \qquad q_0 = \mathbb{E}(Q_0)$$
$$\kappa_p = \mathbb{E}(K_p), \qquad \lambda_p = \mathbb{E}(L_p),$$

for $p \in \{1,2\}$, where B, K_p, L_p and K_0 are discrete random variables such that $\mathbb{P}\Big((B, K_p, L_p, Q_0) = (\beta_{1,j}, \kappa_{p,j}, \lambda_{p,j}, q_{0,j})\Big) = w_j$ and $w_j = \dfrac{\pi_j\sigma_{1,j}}{\sum_{i=1}^{J}\pi_i\sigma_{1,i}}$ for $j = 1,\dots,J$. stable random noise

3.6.3 Interpreting $\rho^{\alpha h}$ using point processes

The quantity $\rho^{\alpha h}$ appearing in Corollary 3.2.2 has the intuitive interpretation of a survival probability at horizon h of a bubble generated by (3.1). This conclusion can also be reached using point processes under the less restrictive assumption that the errors of (3.1) belong to the domain of attraction of an α-stable distribution. Consider n observations $X_1,\dots,X_n$ of (3.1) where now (ε_t) is an i.i.d. sequence of random variables such that:

$$\mathbb{P}(|\varepsilon_0| > x) = x^{-\alpha}L(x), \qquad \text{and} \qquad \lim_{x\to\infty}\frac{\mathbb{P}(\varepsilon_0 > x)}{\mathbb{P}(|\varepsilon_0| > x)} \to c \in [0,1],$$

with L a slowly varying function at infinity. Let $a_n = \inf\{u : \mathbb{P}(|\varepsilon_0| > u) \leq n^{-1}\}$. Then, adapting Section 3.D in [40], we can study the time indexes $k \in \{1,\dots,n\}$ for which $a_n^{-1}X_k$ falls outside the interval $(-x,x)$, for $x > 0$, that is, the time indexes for which (X_t) undergoes extreme events. The corresponding point process converges as the number of observations n grows to infinity:

$$\sum_{k=1}^{n}\delta_{\left(k/n, a_n^{-1}X_k\right)}\Big(\ \cdot\ \cap B_x\Big) \xrightarrow{d} \sum_{k=1}^{+\infty}\xi_k\delta_{\Upsilon_k},$$

where δ is the Dirac measure, $B_x = (0,+\infty)\times\Big((-\infty,-x)\cup(x,+\infty)\Big)$, $\{\Upsilon_k, k \geq 1\}$ are the points of a homogeneous Poisson Random Measure (PRM) on $(0,+\infty)$ with rate $x^{-\alpha}$,[12] and $\xi_k = \text{Card}\{i \in \mathbb{Z} : J_k|\rho^i| >$

[12] See [37]: $\{\Upsilon_k, k \geq 1\}$ are the points of a homogeneous PRM on $(0,+\infty)$ with rate $x^{-\alpha}$ if and only if, for any $\ell \geq 1$, nonnegative integers $a_1,\dots,a_\ell$ and $b_1,\dots,b_\ell$ such that $a_i < b_i \leq a_{i+1}$, $i = 1,\dots,\ell$, and any nonnegative integers $n_1,\dots,n_\ell$:

$$\mathbb{P}\Big(N(a_i,b_i] = n_i, i = 1,\dots,\ell\Big) = \prod_{i=1}^{\ell}\frac{[x^{-\alpha}(b_i-a_i)]^{n_i}}{n_i!}\exp\Big\{-x^{-\alpha}(b_i-a_i)\Big\},$$

1} where $\{J_k, k \geq 1\}$ are i.i.d. on $(1, +\infty)$, independent of $\{\Upsilon_k\}$, with common density:

$$f(z) = \alpha z^{-\alpha-1} \mathbb{1}_{(1,+\infty)}(z). \tag{3.32}$$

The sequences $\{\Upsilon_k\}$ and $\{\xi_k\}$ are interpreted (see [89]) as describing respectively the occurrence dates of clusters of extreme events and the size of these clusters (i.e. the number of co-occurring extreme events, which here corresponds to the duration of bubble episodes). Since $\xi_k = \text{Card}\{i \in \mathbb{Z} : J_k|\rho^i| > 1\} = \arg\max_{i\geq 1}\{J_k > |\rho|^{-i}\}$, we can obtain explicitly the distribution of the bubble duration using (3.32). For any $h \geq 1$,

$$\mathbb{P}\Big(\xi_k \geq h\Big) = \mathbb{P}\Big(J_k > |\rho|^{-h}\Big) = |\rho|^{\alpha h},$$

which as announced, is precisely the probability parameter of the Bernoulli variable intervening in the suggested interpretation of Corollary 3.2.2.

3.7 Postponed proofs

3.7.1 Proof of Proposition 3.2.2

To prove Proposition 3.2.2, we begin with a Lemma that gives the forms of the spectral measure and shift vector for more general, discrete time vectors of linear moving averages driven by α-stable noise. Let $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, \beta, \sigma, \mu)$, m an integer such that $m \geq 2$ and let $\{d_{k,i}, k \in \mathbb{Z}, i = 1, \ldots, m\}$ be a real deterministic sequence verifying

$$\text{for any } i = 1, \ldots, m, \quad \sum_{k\in\mathbb{Z}} |d_{k,i}|^s < +\infty, \quad \text{for some} \quad s < \alpha, \quad s \leq 1. \tag{3.33}$$

Consider the vector

$$\boldsymbol{X_t} = (X_{1,t}, \ldots, X_{m,t}), \quad \text{with} \quad X_{i,t} = \sum_{k\in\mathbb{Z}} d_{k,i}\varepsilon_{t+k}, \quad \text{for} \quad i = 1, \ldots, m. \tag{3.34}$$

It follows from Proposition 13.3.1 in [19] that the infinite series converge almost surely and $\boldsymbol{X}_t$ is well defined. Denote $\boldsymbol{d}_k = (d_{k,1}, \ldots, d_{k,m})$ for any $k \in \mathbb{Z}$ and $\boldsymbol{u} = (u_1, \ldots, u_m) \in \mathbb{R}^m$.

Lemma 3.7.1 *Let $0 < \alpha < 2$ and let $\boldsymbol{X_t}$ satisfy (3.33) and (3.34). Then, $\boldsymbol{X_t}$ is an α-stable random vector in $\mathbb{R}^m$, with spectral measure Γ on the unit sphere S_m and location vector $\boldsymbol{\mu}^0 \in \mathbb{R}^m$ such that*

$$\begin{aligned}\Gamma &= \sigma^\alpha \frac{1+\beta}{2} \sum_{k\in\mathbb{Z}} \|\boldsymbol{d}_k\|^\alpha \delta_{\left\{\frac{\boldsymbol{d}_k}{\|\boldsymbol{d}_k\|}\right\}} + \sigma^\alpha \frac{1-\beta}{2} \sum_{k\in\mathbb{Z}} \|\boldsymbol{d}_k\|^\alpha \delta_{\left\{\frac{-\boldsymbol{d}_k}{\|\boldsymbol{d}_k\|}\right\}}, \\ \boldsymbol{\mu}^0 &= \sum_{k\in\mathbb{Z}} \boldsymbol{d}_k \mu - \mathbb{1}_{\{\alpha=1\}} \frac{2}{\pi} \sigma\beta \sum_{k\in\mathbb{Z}} \boldsymbol{d}_k \ln \|\boldsymbol{d}_k\|,\end{aligned} \tag{3.35}$$

where $\delta_{\{x\}}$ is the dirac measure at point $x \in \mathbb{R}$ and by convention, if for some $k \in \mathbb{Z}$, $\boldsymbol{d}_k = \boldsymbol{0}$, i.e. $\|\boldsymbol{d}_k\| = 0$, then the kth term vanishes from the sums.

where $N(a_i, b_i]$ denotes the number of terms of $\{\Upsilon_k, k \geq 1\}$ falling in the half-open interval $(a_i, b_i]$, $i = 1, \ldots, \ell$.

Proof. The characteristic function of $\boldsymbol{X_t}$ reads, for any $\boldsymbol{u} \in \mathbb{R}^m$:

$$\varphi_{\boldsymbol{X_t}}(\boldsymbol{u}) = \mathbb{E}\left(\exp\left\{i\sum_{j=1}^m u_j X_{j,t}\right\}\right) = \prod_{k\in\mathbb{Z}} \mathbb{E}\left[i\left(\sum_{j=1}^m u_j d_{k,j}\right)\varepsilon_{t+k}\right].$$

We obtain that for $\alpha \neq 1$,

$$\varphi_{\boldsymbol{X_t}}(\boldsymbol{u}) = \exp\left\{-\sum_{k\in\mathbb{Z}}\sigma^\alpha|\sum_{j=1}^m u_j d_{k,j}|^\alpha\left(1 - i\beta\text{sign}\Big(\sum_{j=1}^m u_j d_{k,j}\Big)\text{tg}\frac{\pi\alpha}{2}\right) + i\sum_{j=1}^m u_j\sum_{k\in\mathbb{Z}} d_{k,j}\mu\right\}. \tag{3.36}$$

If $\alpha = 1$,

$$\begin{aligned}\varphi_{\boldsymbol{X_t}}(\boldsymbol{u}) = \exp\Bigg\{-\sum_{k\in\mathbb{Z}}\sigma|\sum_{j=1}^m u_j d_{k,j}|\left(1 + i\beta\frac{2}{\pi}\text{sign}\Big(\sum_{j=1}^m u_j d_{k,j}\Big)\ln\Big|\sum_{j=1}^m u_j d_{k,j}\Big|\right)&\\ +i\sum_{j=1}^m u_j\sum_{k\in\mathbb{Z}} d_{k,j}\mu\Bigg\}.&\end{aligned} \tag{3.37}$$

Replacing (3.35) in (3.2) allows to retrieve the two above formulae. □

Let us now prove Proposition 3.2.2. From Lemma 3.7.1, taking $\boldsymbol{X_t} = (X_t, X_{t+h})$, we have for the anticipative AR(1), $\boldsymbol{d}_k = (\rho^k \mathbb{1}_{k\geq 0}, \rho^{k-h}\mathbb{1}_{k\geq h})$ for any $h \in \mathbb{Z}$ and $h \geq 1$, and

$$\begin{aligned}
\Gamma_h &= \sigma^\alpha\sum_{\vartheta\in S_1}\frac{1+\vartheta\beta}{2}\left[\sum_{k=0}^{h-1}|\rho^k|^\alpha\delta_{\left\{\vartheta\,\frac{(\rho^k,0)}{|\rho|^k}\right\}} + \sum_{k\geq h}\left|\rho^{2k}+\rho^{2(k-h)}\right|^{\alpha/2}\delta_{\left\{\vartheta\,\frac{(\rho^k,\rho^{k-h})}{\sqrt{|\rho|^{2k}+|\rho|^{2(k-h)}}}\right\}}\right]\\
&= \sigma^\alpha\sum_{\vartheta\in S_1}\frac{1+\vartheta\beta}{2}\left[\sum_{k=0}^{h-1}|\rho|^{\alpha k}\delta_{\{\vartheta\,\text{sign}(\rho)^k(1,0)\}} + \Big(1+|\rho|^{-2h}\Big)^{\alpha/2}\sum_{k\geq h}|\rho|^{\alpha k}\delta_{\{\vartheta\,\text{sign}(\rho)^{k+h}\boldsymbol{s}_h\}}\right]\\
&= \frac{\sigma^\alpha}{2}\Bigg[\sum_{k=0}^{h-1}|\rho|^{\alpha k}\sum_{\vartheta\in S_1}\Big(1+\vartheta\,\text{sign}(\rho)^k\beta\Big)\delta_{\{(\vartheta,0)\}}\\
&\qquad + \Big(1+|\rho|^{-2h}\Big)^{\alpha/2}\sum_{k\geq h}|\rho|^{\alpha k}\sum_{\vartheta\in S_1}\Big(1+\vartheta\,\text{sign}(\rho)^{k+h}\beta\Big)\delta_{\{\vartheta\,\boldsymbol{s}_h\}}\Bigg]\\
&= \frac{\sigma^\alpha}{2}\sum_{\vartheta\in S_1}\Bigg[\left(\frac{1-|\rho|^{\alpha h}}{1-|\rho|^\alpha} + \frac{1-\big(\rho^{<\alpha>}\big)^h}{1-\rho^{<\alpha>}}\vartheta\beta\right)\delta_{\{(\vartheta,0)\}}\\
&\qquad + \Big(1+|\rho|^{-2h}\Big)^{\alpha/2}\left(\frac{|\rho|^{\alpha h}}{1-|\rho|^\alpha} + \text{sign}(\rho)^h\frac{\big(\rho^{<\alpha>}\big)^h\vartheta\beta}{1-\rho^{<\alpha>}}\right)\delta_{\{\vartheta\,\boldsymbol{s}_h\}}\Bigg]\\
&= \frac{\sigma^\alpha}{2}\sum_{\vartheta\in S_1}\Bigg[\left(\frac{1-|\rho|^{\alpha h}}{1-|\rho|^\alpha} + \frac{1-\big(\rho^{<\alpha>}\big)^h}{1-\rho^{<\alpha>}}\vartheta\beta\right)\delta_{\{(\vartheta,0)\}}\\
&\qquad + \Big(1+|\rho|^{2h}\Big)^{\alpha/2}\left(\frac{1}{1-|\rho|^\alpha} + \frac{\vartheta\beta}{1-\rho^{<\alpha>}}\right)\delta_{\{\vartheta\,\boldsymbol{s}_h\}}\Bigg].
\end{aligned}$$

From Proposition 3.7.1, we also have $\boldsymbol{\mu}^0 = \boldsymbol{0}$ for $\alpha \neq 1$ (since $\mu = 0$ in (3.1)). For $\alpha = 1$, we have

$$\mu_1^0 = -\frac{2}{\pi}\sigma\beta A_1, \qquad\qquad \mu_2^0 = -\frac{2}{\pi}\sigma\beta A_2,$$

with

$$A_1 = \ln|\rho| \sum_{k=0}^{+\infty} k\rho^k + \frac{1}{2}\ln\left(1+|\rho|^{-2h}\right)\sum_{k=h}^{+\infty}\rho^k, \qquad A_2 = \rho^{-h}\left[\ln|\rho|\sum_{k=h}^{+\infty} k\rho^k + \frac{1}{2}\ln\left(1+|\rho|^{-2h}\right)\sum_{k=h}^{+\infty}\rho^k\right].$$

It is easily shown that $\sum_{k=h}^{+\infty} k\rho^k = \dfrac{h\rho^h}{1-\rho} + \dfrac{\rho^{h+1}}{(1-\rho)^2}$ for $h \geq 0$. Substituting in A_1 and A_2 yields the conclusion.

3.7.2 Proof of Theorem 3.2.1

From Proposition 3.2.2, we know that (X_t, X_{t+h}) is an α-stable vector with spectral representation denoted $(\Gamma_h, \mathbf{0})$. Corollary 3.2.1 gives a sufficient condition for the existence of conditionals moments and Theorems 3.6.1, 3.6.3, 3.6.5 and 3.6.6 give their analytical forms in terms of the spectral measure. Given Γ_h as in (3.4), the constants σ_1, β_1, the κ_p's and λ_p's simplify. For instance:

$$\begin{aligned}
\sigma_1^\alpha &= \int_{S_2} |s_1|^\alpha \Gamma_h(d\boldsymbol{s}) \\
&= \frac{\bar{\sigma}^\alpha}{2}\sum_{\vartheta\in S_1}\left[\left(1-|\rho|^{\alpha h} + \left(1-\left(\rho^{<\alpha>}\right)^h\right)\vartheta\bar{\beta}\right)|\vartheta|^\alpha + \left(1+|\rho|^{2h}\right)^{\alpha/2}(1+\vartheta\bar{\beta})\left|\frac{\vartheta\rho^h}{\sqrt{1+|\rho|^{2h}}}\right|^\alpha\right] = \bar{\sigma}^\alpha.
\end{aligned}$$

3.7.3 Proof of Corollary 3.2.2

We will give the proof for the excess kurtosis. The other limits and equivalents are obtained in a similar manner. Letting $\alpha \in (3/2, 2)$ ensures the existence of the fourth order moment.

Since we assume $\rho > 0$, it follows that $\lambda_p = \beta_1\kappa_p$ for $p = 1, 2, 3, 4$. Using Proposition 3.6.1, it is straightforward to show that as x tends to infinity

$$\gamma_2(x,h) \longrightarrow \frac{\kappa_4 - 4\kappa_1\kappa_3 + 6\kappa_1^2\kappa_2 - 3\kappa_1^4}{\left(\kappa_2-\kappa_1^2\right)^2} - 3.$$

Substituting the κ_p's by the expressions in Theorem 3.2.1 and rearranging terms yields the conclusion.

3.7.4 Proof of Proposition 3.2.3

The α-stable random vector (X_t, X_{t+h}) admits the integral representation

$$(X_t, X_{t+h}) = \left(\int_{\mathbb{R}} f_1(x-t)M(dx), \int_{\mathbb{R}} f_2(x-t)M(dx)\right), \tag{3.38}$$

where $f_1(x) = e^{-\lambda x}\mathbb{1}_{\{x\geq 0\}}$ and $f_2(x) = f_1(x-h)$, for $x \in \mathbb{R}$. Let $\boldsymbol{u} = (u_1, u_2) \in \mathbb{R}^2$. For $\alpha \neq 1$, by Proposition 3.4.1(i) in [117], its characteristic function reads

$$
\begin{aligned}
&\mathbb{E}\bigg[\exp\Big\{i\sum_{j=1}^{2} u_j \int_{\mathbb{R}} f_j M(dx)\Big\}\bigg] \\
&\qquad = \exp\bigg\{\int_{\mathbb{R}}\bigg[-\Big|\sum_{j=1}^{2} u_j f_j\Big|^{\alpha} + ia\beta\Big(\sum_{j=1}^{2} u_j f_j\Big)^{<\alpha>}\bigg]dx\bigg\} \\
&\qquad = \exp\bigg\{\int_0^h \bigg[-\Big|u_1 e^{-\lambda x}\Big|^{\alpha} + ia\beta\Big(u_1 e^{-\lambda x}\Big)^{<\alpha>}\bigg]dx \\
&\qquad\qquad + \int_h^{+\infty}\bigg[-\Big|u_1 e^{-\lambda x} + u_2 e^{-\lambda(x-h)}\Big|^{\alpha} + ia\beta\Big(u_1 e^{-\lambda x} + u_2 e^{-\lambda(x-h)}\Big)^{<\alpha>}\bigg]dx\bigg\} \\
&\qquad = \exp\bigg\{\Big(-|u_1|^{\alpha} + ia\beta u_1^{<\alpha>}\Big)\frac{1-e^{-\alpha\lambda h}}{\alpha\lambda} \\
&\qquad\qquad + \Big(|u_1 + u_2 e^{\lambda h}|^{\alpha} + ia(u_1 + u_2 e^{\lambda h})^{<\alpha>}\Big)\frac{e^{-\alpha\lambda h}}{\alpha\lambda}\bigg\} \\
&\qquad = \exp\bigg\{\frac{1}{\alpha\lambda}\sum_{\vartheta\in S_1}\frac{1+\vartheta\beta}{2}\bigg[\Big(-|\vartheta u_1|^{\alpha} + ia(\vartheta u_1)^{<\alpha>}\Big)(1-e^{-\alpha\lambda h}) \\
&\qquad\qquad + \bigg(-\Big|u_1\frac{\vartheta e^{-\lambda h}}{\sqrt{1+e^{-2\lambda h}}} + u_2\frac{\vartheta}{\sqrt{1+e^{-2\lambda h}}}\Big|^{\alpha} \\
&\qquad\qquad\qquad + ia\Big(u_1\frac{\vartheta e^{-\lambda h}}{\sqrt{1+e^{-2\lambda h}}} + u_2\frac{\vartheta}{\sqrt{1+e^{-2\lambda h}}}\Big)^{<\alpha>}\bigg)\Big(1+e^{-2\lambda h}\Big)^{\alpha/2}\bigg]\bigg\} \\
&\qquad = \exp\bigg\{-\int_{S_2}|\langle\boldsymbol{u},\boldsymbol{s}\rangle|^{\alpha} - ia(\langle\boldsymbol{u},\boldsymbol{s}\rangle)^{<\alpha>}\Gamma_h(d\boldsymbol{s})\bigg\},
\end{aligned}
$$

with $a = \mathrm{tg}(\pi\alpha/2)$ and Γ_h as in the Proposition. If $\alpha = 1$, then with $a = 2/\pi$, we have by Proposition 3.4.1(ii) in [117]

$$
\begin{aligned}
&\mathbb{E}\Bigg[\exp\Big\{i\sum_{j=1}^{2} u_j \int_{\mathbb{R}} f_j M(dx)\Big\}\Bigg] \\
&\quad = \exp\Bigg\{-\int_{\mathbb{R}}\Big[\Big|\sum_{j=1}^{2} u_j f_j\Big| + ia\beta\Big(\sum_{j=1}^{2} u_j f_j\Big)\ln\Big|\sum_{j=1}^{2} u_j f_j\Big|\Big]dx\Bigg\} \\
&\quad = \exp\Bigg\{-\int_0^h\Big[\Big|u_1e^{-\lambda x}\Big| + ia\beta\Big(u_1e^{-\lambda x}\Big)\ln\Big|u_1e^{-\lambda x}\Big|\Big]dx \\
&\qquad\qquad -\int_h^{+\infty}\Big[\Big|u_1e^{-\lambda x} + u_2e^{-\lambda(x-h)}\Big| \\
&\qquad\qquad\qquad + ia\beta\Big(u_1e^{-\lambda x} + u_2e^{-\lambda(x-h)}\Big)\ln\Big|u_1e^{-\lambda x} + u_2e^{-\lambda(x-h)}\Big|\Big]dx\Bigg\} \\
&\quad = \exp\Bigg\{-\Big(|u_1| + ia\beta(u_1)\ln|u_1|\Big)\int_0^h e^{-\lambda x}dx + ia\lambda\beta u_1\int_0^h xe^{-\lambda x}dx \\
&\qquad\qquad -\Big(|u_1 + u_2e^{\lambda h}| + ia\beta(u_1 + u_2e^{\lambda h})\ln|u_1 + u_2e^{\lambda h}|\Big)\int_h^{+\infty} e^{-\lambda x}dx \\
&\qquad\qquad + ia\lambda\beta(u_1 + u_2e^{\lambda h})\int_h^{+\infty} xe^{-\lambda x}dx\Bigg\} \\
&\quad = \exp\Bigg\{-\frac{1}{\lambda}\sum_{\vartheta\in S_1}\frac{1+\vartheta\beta}{2}\Bigg[\Big(|\vartheta u_1| + ia(\vartheta u_1)\ln|\vartheta u_1|\Big)\Big(1 - e^{-\lambda h}\Big) \\
&\qquad + \Bigg(\Big|u_1\frac{\vartheta e^{-\lambda h}}{\sqrt{1+e^{-2\lambda h}}} + u_2\frac{\vartheta}{\sqrt{1+e^{-2\lambda h}}}\Big| \\
&\qquad + ia\Big(u_1\frac{\vartheta e^{-\lambda h}}{\sqrt{1+e^{-2\lambda h}}} + u_2\frac{\vartheta}{\sqrt{1+e^{-2\lambda h}}}\Big)\ln\Big|u_1\frac{\vartheta e^{-\lambda h}}{\sqrt{1+e^{-2\lambda h}}} + u_2\frac{\vartheta}{\sqrt{1+e^{-2\lambda h}}}\Big|\Bigg)\sqrt{1+e^{-2\lambda h}}\Bigg] \\
&\qquad - ia\beta(u_1e^{-\lambda h} + u_2)\Big(h + \frac{\ln(1+e^{-2\lambda h})}{2\lambda}\Big) + ia\lambda\beta\Big(u_1\int_0^{+\infty} xe^{-\lambda dx} + u_2e^{\lambda h}\int_h^{+\infty} xe^{-\lambda dx}\Big)\Bigg\},
\end{aligned}
$$

and

$$
\begin{aligned}
\lambda\Big(u_1\int_0^{+\infty} xe^{-\lambda dx} + u_2e^{\lambda h}\int_h^{+\infty} xe^{-\lambda dx}\Big) &= u_1\lambda^{-1} + u_2(h + \lambda^{-1}), \\
h + \frac{\ln(1+e^{-2\lambda h})}{2\lambda} &= \ln(1+e^{2\lambda h}).
\end{aligned}
$$

Hence,

$$
\mathbb{E}\Bigg[\exp\Big\{i\Big(u_1X_t + u_2X_{t+h}\Big)\Big\}\Bigg] = \exp\Bigg\{-\int_{S_2}|\langle\boldsymbol{u},\boldsymbol{s}\rangle| + ia(\langle\boldsymbol{u},\boldsymbol{s}\rangle)\ln|\langle\boldsymbol{u},\boldsymbol{s}\rangle|\Gamma_h(d\boldsymbol{s}) + i\langle\boldsymbol{u},\boldsymbol{\mu}^0\rangle\Bigg\},
$$

with Γ_h and $\boldsymbol{\mu}^0$ as claimed for $\alpha = 1$.

3.7.5 Proof of Theorem 3.2.2

Let $\{X_t, t \in \mathbb{R}\}$ be the anticipative OU defined in (3.19) with $\lambda > 0$ and let $h > 0$. Since, the Markov property of the anticipative OU has already been discussed in the relevant section, let us focus on the points two last points, namely the existence and the form of the conditional moments. The proof could be done similarly to those of the discrete time AR(1) using the expression of the spectral measure obtained at Proposition 3.2.3. However, we propose another proof which has the advantage of illustrating how α-stable vectors with different integrand functions f_1 and f_2 in (3.38) could be considered. Formulae are given by Proposition 3.1 in [116] for expressing the constants like the ones we introduced in Equations (3.24) and (3.25) in terms of these integrand functions. The condition (3.3) can be translated in terms of f_1 and f_2 as

$$\int_{S_2} |s_1|^{-\nu} \Gamma_h(d\boldsymbol{s}) < +\infty \iff \int_{\mathbb{R}_+} \frac{|f_2(x)|^{\alpha+\nu}}{|f_1(x)|^{\nu}} dx = e^{\lambda h(\alpha+\nu)} \int_h^{+\infty} e^{-\alpha\lambda x} dx = \frac{e^{\lambda h \nu}}{\alpha\lambda} < +\infty,$$

which is satisfied for any $\nu \geq 0$, hence $(\iota\iota)$. Let us turn to point $(\iota\iota\iota)$. The conditional moments are given by Theorems 3.6.1, 3.6.3, 3.6.5 and 3.6.6 for an arbitrary spectral measure Γ. From the proof of Proposition 3.1 in [116], we know how we can rewrite the constants σ_1^α, β_1, κ_p and λ_p, for $p \in \{1, 2, 3, 4\}$ which are expressed in terms of an integral of s_1, s_2 and Γ into expressions involving f_1, f_2 and the Lebesgue measure. It can be shown that

$$\begin{aligned}
\sigma_1^\alpha &= \int_{\mathbb{R}} |f_1(x)|^\alpha dx = \int_t^{+\infty} e^{-\alpha\lambda x} dx = \frac{1}{\alpha\lambda}, \\
\beta_1 &= \frac{\int_{\mathbb{R}} f_1(x)^{<\alpha>} \beta(x) dx}{\sigma_1^\alpha} = \beta \frac{\int_{\mathbb{R}} f_1(x)^\alpha dx}{\sigma_1^\alpha} = \beta, \\
\kappa_p &= \frac{\int_{\mathbb{R}_+} (f_2(x)/f_1(x))^p |f_1(x)|^\alpha dx}{\sigma_1^\alpha} = \frac{1}{\sigma_1^\alpha} \int_h^{+\infty} \left(\frac{e^{-\lambda(x-h)}}{e^{-\lambda x}} \right)^p e^{-\alpha\lambda x} dx = e^{\lambda h(\alpha-p)}, \\
\lambda_p &= \frac{\int_{\mathbb{R}_+} (f_2(x)/f_1(x))^p |f_1(x)|^{<\alpha>} \beta(x) dx}{\sigma_1^\alpha} = \beta \frac{\int_{\mathbb{R}_+} (f_2(x)/f_1(x))^p |f_1(x)|^\alpha dx}{\sigma_1^\alpha} = \beta\kappa_p, \\
q_0 &= \frac{\int_{\mathbb{R}_+} f_2(x)\beta(x) \ln \left| \frac{f_1(x)}{\sqrt{f_1^2(x)+f_2^2(x)}} \right| dx}{\sigma_1} = -\frac{1}{2}\beta \int_h^{+\infty} e^{-\lambda(x-h)} \ln(1+e^{2\lambda h}) dx = -\frac{1}{2}\beta \ln(1+e^{2\lambda h}).
\end{aligned}$$

3.7.6 Proof of Lemma 3.3.1

Using the independence between the $X_{j,t}$'s and denoting $\boldsymbol{X}_j = (X_{j,t}, X_{j,t+h})$,

$$\begin{aligned}
\mathbb{E}\Big[e^{iuX_t + ivX_{t+h}}\Big] &= \mathbb{E}\left[\exp\left\{iuc\sum_{j=1}^J \pi_j X_{j,t} + ivc\sum_{j=1}^J \pi_j X_{j,t+h}\right\}\right] \\
&= \prod_{j=1}^J \mathbb{E}\left[\exp\left\{i\langle \boldsymbol{u}c\pi_j, \boldsymbol{X}_j\rangle\right.\right] \\
&= \prod_{j=1}^J \exp\Bigg\{ -\int_{S_2} |\langle \boldsymbol{u}c\pi_j, \boldsymbol{s}\rangle|^\alpha \Big(1 - i\,\mathrm{sign}(\langle \boldsymbol{u}c\pi_j, \boldsymbol{s}\rangle) w(\alpha, \langle \boldsymbol{u}c\pi_j, \boldsymbol{s}\rangle)\Big)\Gamma_{j,h}(d\boldsymbol{s}) \\
&\qquad\qquad + i\,\langle \boldsymbol{u}c\pi_j, \boldsymbol{\mu}^0\rangle \Bigg\},
\end{aligned}$$

When $\alpha \neq 1$, then $w(\alpha, \cdot) = \mathrm{tg}(\pi\alpha/2)$ and

$$\begin{aligned}\mathbb{E}\Big[e^{iuX_t+ivX_{t+h}}\Big] &= \exp\bigg\{-c^\alpha \sum_{j=1}^J \pi_j^\alpha \int_{S_2} |\langle \boldsymbol{u}, \boldsymbol{s}\rangle|^\alpha \Big(1 - i\,\mathrm{sign}(\langle \boldsymbol{u}, \boldsymbol{s}\rangle) w(\alpha, \langle \boldsymbol{u}, \boldsymbol{s}\rangle)\Big)\Gamma_{j,h}(d\boldsymbol{s})\bigg\} \\ &= \exp\bigg\{-\int_{S_2} |\langle \boldsymbol{u}, \boldsymbol{s}\rangle|^\alpha \Big(1 - i\,\mathrm{sign}(\langle \boldsymbol{u}, \boldsymbol{s}\rangle) w(\alpha, \langle \boldsymbol{u}, \boldsymbol{s}\rangle)\Big)\Gamma_h(d\boldsymbol{s})\bigg\}.\end{aligned}$$

When $\alpha = 1$, with $a = 2/\pi$,

$$\begin{aligned}\mathbb{E}\Big[e^{iuX_t+ivX_{t+h}}\Big] &= \prod_{j=1}^J \exp\bigg\{-\int_{S_2} |\langle \boldsymbol{u}c\pi_j, \boldsymbol{s}\rangle| + ia\langle \boldsymbol{u}c\pi_j, \boldsymbol{s}\rangle \ln|\langle \boldsymbol{u}c\pi_j, \boldsymbol{s}\rangle|\Gamma_{j,h}(d\boldsymbol{s}) + i\,\langle \boldsymbol{u}c\pi_j, \boldsymbol{\mu}_j^0\rangle)\bigg\} \\ &= \exp\bigg\{-c\int_{S_2} |\langle \boldsymbol{u}, \boldsymbol{s}\rangle| + ia\langle \boldsymbol{u}, \boldsymbol{s}\rangle \ln|\langle \boldsymbol{u}, \boldsymbol{s}\rangle| \sum_{j=1}^J \pi_j \Gamma_{j,h}(d\boldsymbol{s}) \\ &\qquad + i\sum_{j=1}^J \Big(\langle \boldsymbol{u}, c\pi_j \boldsymbol{\mu}_j^0\rangle - ac\pi_j \ln|c\pi_j| \int_{S_2} \langle \boldsymbol{u}, \boldsymbol{s}\rangle \Gamma_{j,h}(d\boldsymbol{s})\Big)\bigg\},\end{aligned}$$

and

$$\begin{aligned}i\sum_{j=1}^J \Big(\langle \boldsymbol{u}, c\pi_j \boldsymbol{\mu}_j^0\rangle - ac\pi_j \ln|c\pi_j| \int_{S_2} \langle \boldsymbol{u}, \boldsymbol{s}\rangle \Gamma_{j,h}(d\boldsymbol{s})\Big) &= i\langle \boldsymbol{u}, c\sum_{j=1}^J \pi_j \Big(\boldsymbol{\mu}_j^0 - a\ln|c\pi_j| \int_{S_2} \boldsymbol{s}\Gamma_{j,h}(d\boldsymbol{s})\Big)\rangle \\ &= i\langle \boldsymbol{u}, c\sum_{j=1}^J \pi_j \Big(\boldsymbol{\mu}_j^0 - a\sigma_{1,j} \ln|c\pi_j| \begin{pmatrix}\beta_{1,j} \\ \lambda_{1,j}\end{pmatrix}\Big)\rangle.\end{aligned}$$

3.7.7 Preliminary elements for the proofs of the main results

Notations for the proofs of Theorems 3.6.3-3.6.6 and Proposition 3.6.1

Let $\boldsymbol{X} = (X_1, X_2)$ be an α-stable vector, with $0 < \alpha < 2$, $\alpha \neq 1$, and spectral representation $(\Gamma, \boldsymbol{0})$. Its characteristic function will be denoted $\varphi_{\boldsymbol{X}}(t, r)$ for any $(t, r) \in \mathbb{R}^2$, and reads

$$\varphi_{\boldsymbol{X}}(t, r) = \exp\left\{-\int_{S_2} g_1(ts_1 + rs_2)\Gamma(ds)\right\}, \tag{3.39}$$

where $g_1(z) = |z|^\alpha - iaz^{<\alpha>}$ for $z \in \mathbb{R}$, and $a = \mathrm{tg}(\pi\alpha/2)$. As we assume $\sigma_1 > 0$ so that X_1 is not degenerate, the conditional characteristic function of X_2 given $X_1 = x$, denoted $\phi_{X_2|x}(r)$ for $r \in \mathbb{R}$, equals

$$\phi_{X_2|x}(r) := 1 + \frac{1}{2\pi f_{X_1}(x)} \int_{\mathbb{R}} e^{-itx}\Big(\varphi_{\boldsymbol{X}}(t, r) - \varphi_{\boldsymbol{X}}(t, 0)\Big)dt. \tag{3.40}$$

where f_{X_1} denotes the density of $X_1 \sim \mathcal{S}(\alpha, \beta_1, \sigma_1, 0)$. The following notation of the $\mathcal{H}$ family function will be more handy than that in (3.10): for any $y > -1$ and $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \mathbb{R}^2$, define the function $\mathcal{H}(y, \boldsymbol{\theta}; \cdot)$ for $x \in \mathbb{R}$ as

$$\mathcal{H}(y, \boldsymbol{\theta}; x) = \int_0^{+\infty} e^{-\sigma_1^\alpha u^\alpha} u^y \Big(\theta_1 \cos(ux - a\beta_1\sigma_1^\alpha u^\alpha) + \theta_2 \sin(ux - a\beta_1\sigma_1^\alpha u^\alpha)\Big)du, \tag{3.41}$$

For $z \in \mathbb{R}$, denote also,

$$g_2(z) = z^{<\alpha-1>} - ia|z|^{\alpha-1}, \tag{3.42}$$

$$g_3(z) = |z|^{\alpha-2} - iaz^{<\alpha-2>}. \tag{3.43}$$

Often, we shall invoke functions of the form

$$r \longmapsto \int_{\mathbb{R}} e^{-itx} \varphi_{\boldsymbol{X}}(t,r) f_1^{p_1}(t,r) \dots f_m^{p_m}(t,r) dt, \tag{3.44}$$

where $m \leq 3$ and the f_i's will be functions of the type $f_i(t,r) = \int_{S_2} g_{j_i}(ts_1 + rs_2) s_1^{k_i} s_2^{\ell_i} \Gamma(d\boldsymbol{s})$, for $j_i = 2, 3$, $k_i, \ell_i \in \mathbb{Z}$ for which f_i is well defined and positive integer exponents p_i's. As a shorthand when no ambiguity is possible, we shall denote functions like (3.44) by

$$\Lambda\left(\int_{S_2} g_{j_1} s_1^{k_1} s_2^{\ell_1}\right)^{p_1} \left(\int_{S_2} g_{j_2} s_1^{k_2} s_2^{\ell_2}\right)^{p_2} \dots$$

up to the m^{th} term.

Lemma 3.7.2 *Let (X_1, X_2) be an α-stable vector, $0 < \alpha < 2, \alpha \neq 1$, with conditional characteristic function $\phi_{X_2|x}$ as given in* (3.40)*. Let $r \in \mathbb{R}$. If $1 < \alpha < 2$, or if $0 < \alpha < 1$ and* (3.3) *holds with $\nu > 1 - \alpha$, the first derivative of $\phi_{X_2|x}$ is given by*

$$\phi^{(1)}_{X_2|x}(r) = \frac{-\alpha}{2\pi f_{X_1}(x)} \Lambda\left(\int_{S_2} g_2 s_2\right). \tag{3.45}$$

If $1/2 < \alpha < 2$ and (3.3) *holds with $\nu > 2 - \alpha$, the second derivative is given by*

$$\phi^{(2)}_{X_2|x}(r) = \frac{-\alpha}{2\pi f_{X_1}(x)} \left[ix\Lambda\left(\int_{S_2} g_2 s_2^2 s_1^{-1}\right) + \alpha\left\{\Lambda\left(\int_{S_2} g_2 s_2^2 s_1^{-1}\right)\left(\int_{S_2} g_2 s_1\right) - \Lambda\left(\int_{S_2} g_2 s_2^2\right)^2\right\}\right], \tag{3.46}$$

If $1 < \alpha < 2$ and (3.3) *holds with $\nu > 3 - \alpha$, the third derivative is given by*

$$\phi^{(3)}_{X_2|x}(r) = \frac{-\alpha}{2\pi f_{X_1}(x)} \left(ix\Big((\alpha - 1)I_1 - \alpha I_2\Big) + \alpha^2(I_3 - I_4) + \alpha(\alpha - 1)(I_5 + I_6 - 2I_7)\right), \tag{3.47}$$

with

$$\begin{aligned}
I_1 &= \Lambda\left(\int_{S_2} g_3 s_2^3 s_1^{-1}\right), & I_5 &= \Lambda\left(\int_{S_2} g_2 s_2^2 s_1^{-1}\right)\left(\int_{S_2} g_3 s_2 s_1\right), \\
I_2 &= \Lambda\left(\int_{S_2} g_2 s_2\right)\left(\int_{S_2} g_2 s_2^2 s_1^{-1}\right), & I_6 &= \Lambda\left(\int_{S_2} g_2 s_1\right)\left(\int_{S_2} g_3 s_2^3 s_1^{-1}\right), \\
I_3 &= \Lambda\left(\int_{S_2} g_2 s_2\right)^3, & I_7 &= \Lambda\left(\int_{S_2} g_2 s_2\right)\left(\int_{S_2} g_3 s_2^2\right), \\
I_4 &= \Lambda\left(\int_{S_2} g_2 s_1\right)\left(\int_{S_2} g_2 s_2\right)\left(\int_{S_2} g_2 s_2^2 s_1^{-1}\right).
\end{aligned}$$

If $3/2 < \alpha < 2$ and (3.3) *holds with $\nu > 4 - \alpha$, the fourth derivative is given by*

$$\begin{aligned}
\phi^{(4)}_{X_2|x}(r) = \frac{-\alpha}{2\pi f_{X_1}(x)} \Bigg[\quad & i\alpha x\left(\alpha\Big(3J_1 - 2J_2\Big) + (\alpha - 1)\Big(2J_3 - 3J_4 + J_5\Big)\right) \\
& + \alpha x^2 J_6 - (\alpha - 1)x^2 J_7 \\
& + \alpha^2(\alpha - 1)\left(J_8 + J_9 + J_{10} - 3\Big(2J_{11} + J_{12} - J_{13}\Big)\right) \\
& + \alpha(\alpha - 1)^2\Big(4J_{14} - 3J_{15} - J_{16}\Big) \\
& + \alpha^3\Big(3J_{17} - J_{18} - J_{19}\Big)\Bigg],
\end{aligned} \tag{3.48}$$

with

$$
\begin{aligned}
J_1 &= \Lambda\Big(\int_{S_2} g_2 s_2^2 s_1^{-1}\Big)\Big(\int_{S_2} g_2 s_2\Big)^2, & J_{11} &= \Lambda\Big(\int_{S_2} g_2 s_2^2 s_1^{-1}\Big)\Big(\int_{S_2} g_3 s_2 s_1\Big)\Big(\int_{S_2} g_2 s_2\Big),\\
J_2 &= \Lambda\Big(\int_{S_2} g_2 s_2^3 s_1^{-2}\Big)\Big(\int_{S_2} g_2 s_1\Big)\Big(\int_{S_2} g_2 s_2\Big), & J_{12} &= \Lambda\Big(\int_{S_2} g_2 s_2^2 s_1^{-1}\Big)\Big(\int_{S_2} g_2 s_1\Big)\Big(\int_{S_2} g_2 s_2\Big),\\
J_3 &= \Lambda\Big(\int_{S_2} g_3 s_2^4 s_1^{-2}\Big)\Big(\int_{S_2} g_2 s_1\Big), & J_{13} &= \Lambda\Big(\int_{S_2} g_3 s_2^2\Big)\Big(\int_{S_2} g_2 s_2\Big)^2,\\
J_4 &= \Lambda\Big(\int_{S_2} g_3 s_2^3 s_1^{-1}\Big)\Big(\int_{S_2} g_2 s_2\Big), & J_{14} &= \Lambda\Big(\int_{S_2} g_3 s_2^3 s_1^{-1}\Big)\Big(\int_{S_2} g_3 s_2 s_1\Big),\\
J_5 &= \Lambda\Big(\int_{S_2} g_2 s_2^3 s_1^{-2}\Big)\Big(\int_{S_2} g_3 s_2 s_1\Big), & J_{15} &= \Lambda\Big(\int_{S_2} g_3 s_2^2\Big)^2,\\
J_6 &= \Lambda\Big(\int_{S_2} g_2 s_2^3 s_1^{-2}\Big)\Big(\int_{S_2} g_2 s_2\Big), & J_{16} &= \Lambda\Big(\int_{S_2} g_3 s_2^4 s_1^{-2}\Big)\Big(\int_{S_2} g_3 s_1^2\Big),\\
J_7 &= \Lambda\Big(\int_{S_2} g_3 s_2^4 s_1^{-2}\Big), & J_{17} &= \Lambda\Big(\int_{S_2} g_2 s_2^2 s_1^{-1}\Big)\Big(\int_{S_2} g_2 s_1\Big)\Big(\int_{S_2} g_2 s_2\Big)^2,\\
J_8 &= \Lambda\Big(\int_{S_2} g_2 s_2^3 s_1^{-2}\Big)\Big(\int_{S_2} g_3 s_1^2\Big)\Big(\int_{S_2} g_2 s_2\Big), & J_{18} &= \Lambda\Big(\int_{S_2} g_2 s_2\Big)^4,\\
J_9 &= \Lambda\Big(\int_{S_2} g_2 s_2^3 s_1^{-2}\Big)\Big(\int_{S_2} g_3 s_2 s_1\Big)\Big(\int_{S_2} g_2 s_1\Big), & J_{19} &= \Lambda\Big(\int_{S_2} g_2 s_2^3 s_1^{-2}\Big)\Big(\int_{S_2} g_2 s_1\Big)^2\Big(\int_{S_2} g_2 s_2\Big),\\
J_{10} &= \Lambda\Big(\int_{S_2} g_3 s_2^4 s_1^{-2}\Big)\Big(\int_{S_2} g_2 s_1\Big)^2.
\end{aligned}
$$

3.7.8 Proof of Lemma 3.7.2

For each of the derivatives, the proof involves two main steps: 1) computation of the derivative 2) justifying inversion of integral and derivation signs. Regarding computation, we detail only the case of the second derivative, whereas for the justification, we detail only the case of the third. Those cases are representative of the main techniques employed for the others.

Computation: second derivative Note that if $f(x) = |x|^b$, for $x, b \in \mathbb{R}$, $b \neq 0$, then for $x \neq 0$, $f'(x) = bx^{<b-1>}$ and if $f : x \longmapsto x^{<b>}$, then $f'(x) = b|x|^{b-1}$. This can be shown by distinguishing the cases $x > 0$ and $x < 0$. Formal computation of the second derivative yields divergent terms when $1/2 < \alpha < 1$ and a special manipulation called «appropriate integration by parts» in [30] (p.106) is needed.

$$
\begin{aligned}
\phi^{(2)}_{X_2|x}(r) &= \frac{\partial}{\partial r}\phi^{(1)}_{X_2|x}(r) \\
&= \frac{-\alpha}{2\pi f_{X_1}(x)} \lim_{h\to 0}\frac{1}{h}\Bigg[\int_{\mathbb{R}}\int_{S_2} e^{-itx}\varphi_{\boldsymbol{X}}(t, r+h) g_2(ts_1+(r+h)s_2)s_2\Gamma(d\boldsymbol{s})dt \\
&\qquad\qquad - \int_{\mathbb{R}}\int_{S_2} e^{-itx}\varphi_{\boldsymbol{X}}(t,r)g_2(ts_1+rs_2)s_2\Gamma(d\boldsymbol{s})dt\Bigg] \\
&= \frac{-\alpha}{2\pi f_{X_1}(x)} \lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}\int_{S_2} e^{-itx}\Big[\varphi_{\boldsymbol{X}}(t,r+h)-\varphi_{\boldsymbol{X}}(t,r)\Big]g_2(ts_1+(r+h)s_2)s_2\Gamma(d\boldsymbol{s})dt \\
&\quad + \frac{-\alpha}{2\pi f_{X_1}(x)} \lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}\int_{S_2} e^{-itx}\varphi_{\boldsymbol{X}}(t,r)\Big[g_2(ts_1+(r+h)s_2)-g_2(ts_1+rs_2)\Big]s_2\Gamma(d\boldsymbol{s})dt \\
&:= A_1 + A_2.
\end{aligned}
$$

The first limit can be straightforwardly obtained:

$$
\begin{aligned}
A_1 &= \frac{\alpha^2}{2\pi f_{X_1}(x)}\int_{\mathbb{R}} e^{-itx}\varphi_{\boldsymbol{X}}(t,r)\bigg(\int_{S_2} g_2(ts_1+rs_2)s_2\Gamma(d\boldsymbol{s})\bigg)^2 dt \\
&= \frac{\alpha^2}{2\pi f_{X_1}(x)}\Lambda\bigg(\int_{S_2} g_2 s_2\bigg)^2.
\end{aligned}
$$

The second one requires «appropriate integration by parts». With the change of variable $t' = t + \dfrac{hs_2}{s_1}$,

$$
\begin{aligned}
A_2 &= \frac{-\alpha}{2\pi f_{X_1}(x)} \lim_{h\to 0}\frac{1}{h}\Bigg[\int_{S_2}\int_{\mathbb{R}} e^{-itx}\varphi_{\boldsymbol{X}}(t,r)g_2(ts_1+(r+h)s_2)s_2 dt\Gamma(d\boldsymbol{s}) \\
&\qquad\qquad - \int_{S_2}\int_{\mathbb{R}} e^{-itx}\varphi_{\boldsymbol{X}}(t,r)g_2(ts_1+rs_2)s_2 dt\Gamma(d\boldsymbol{s})\Bigg] \\
&= \frac{-\alpha}{2\pi f_{X_1}(x)} \lim_{h\to 0}\frac{1}{h}\Bigg[\int_{S_2}\int_{\mathbb{R}} e^{-i\left(t-\frac{hs_2}{s_1}\right)x}\varphi_{\boldsymbol{X}}\Big(t-\frac{hs_2}{s_1}, r\Big)g_2(ts_1+rs_2)s_2 dt\Gamma(d\boldsymbol{s}) \\
&\qquad\qquad - \int_{S_2}\int_{\mathbb{R}} e^{-itx}\varphi_{\boldsymbol{X}}(t,r)g_2(ts_1+rs_2)s_2 dt\Gamma(d\boldsymbol{s})\Bigg] \\
&= \frac{\alpha}{2\pi f_{X_1}(x)}\int_{S_2}\int_{\mathbb{R}} s_2^2 s_1^{-1} g_2(ts_1+rs_2)\lim_{h\to 0}\frac{1}{-\frac{hs_2}{s_1}}\Bigg[e^{-i\left(t-\frac{hs_2}{s_1}\right)x}\varphi_{\boldsymbol{X}}\Big(t-\frac{hs_2}{s_1}, r\Big) \\
&\qquad\qquad - e^{-itx}\varphi_{\boldsymbol{X}}(t,r)\Bigg]dt\Gamma(d\boldsymbol{s}) \\
&= \frac{\alpha}{2\pi f_{X_1}(x)}\int_{S_2}\int_{\mathbb{R}} s_2^2 s_1^{-1} g_2(ts_1+rs_2)\Big[-ixe^{-itx}\varphi_{\boldsymbol{X}}(t,r)+e^{-itx}\frac{\partial}{\partial t}\varphi_{\boldsymbol{X}}(t,r)\Big]dt\Gamma(d\boldsymbol{s}) \\
&= \frac{-i\alpha x}{2\pi f_{X_1}(x)}\int_{\mathbb{R}} e^{-itx}\varphi_{\boldsymbol{X}}(t,r)\bigg(\int_{S_2} s_2^2 s_1^{-1} g_2(ts_1+rs_2)\Gamma(d\boldsymbol{s})\bigg)dt \\
&\quad - \frac{\alpha^2}{2\pi f_{X_1}(x)}\int_{\mathbb{R}} e^{-itx}\varphi_{\boldsymbol{X}}(t,r)\bigg(\int_{S_2} s_1 g_2(ts_1+rs_2)\Gamma(d\boldsymbol{s})\bigg)\bigg(\int_{S_2} s_2^2 s_1^{-1} g_2(ts_1+rs_2)\Gamma(d\boldsymbol{s})\bigg)dt \\
A_2 &= \frac{-i\alpha x}{2\pi f_{X_1}(x)}\Lambda\bigg(\int_{S_2} g_2 s_2^2 s_1^{-1}\bigg) - \frac{\alpha^2}{2\pi f_{X_1}(x)}\Lambda\bigg(\int_{S_2} g_2 s_2^2 s_1^{-1}\bigg)\bigg(\int_{S_2} g_2 s_1\bigg)
\end{aligned}
$$

Combining the expressions obtained for A_1 and A_2 yields the second derivative.

Justifying inversion of integral and derivation signs: First derivative Case $\alpha \in (0,1)$

Assume $\alpha \in (0,1)$. We begin with the first derivative of the imaginary part of $\phi_{X_2|x}$.

$$
\begin{aligned}
&\frac{d}{dr}\Big(\mathrm{Im}\phi_{X_2|x}(r)\Big) \\
&\quad= \frac{-1}{2\pi f_{X_1}(x)} \lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}\Bigg[e^{-\int_{S_2}|ts_1+(r+h)s_2|^{\alpha}\Gamma(ds)}\sin\Big(tx - a\int_{S_2}(ts_1+(r+h)s_2)^{<\alpha>}\Gamma(ds)\Big) \\
&\qquad\qquad - e^{-\int_{S_2}|ts_1+rs_2|^{\alpha}\Gamma(ds)}\sin\Big(tx - a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(ds)\Big)\Bigg]dt \\
&\quad= \frac{-1}{2\pi f_{X_1}(x)} \lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}\Bigg[\sin\Big(tx - a\int_{S_2}(ts_1+(r+h)s_2)^{<\alpha>}\Gamma(ds)\Big) \\
&\qquad\qquad - \sin\Big(tx - a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(ds)\Big)\Bigg] \\
&\qquad\qquad \times \exp\Big\{-\int_{S_2}|ts_1+rs_2|^{\alpha}\Gamma(ds)\Big\}dt \\
&\qquad - \frac{1}{2\pi f_{X_1}(x)} \lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}\Bigg[\exp\Big\{-\int_{S_2}|ts_1+(r+h)s_2|^{\alpha}\Gamma(ds)\Big\} \\
&\qquad\qquad - \exp\Big\{-\int_{S_2}|ts_1+rs_2|^{\alpha}\Gamma(ds)\Big\}\Bigg] \\
&\qquad\qquad \times \sin\Big(tx - a\int_{S_2}(ts_1+(r+h)s_2)^{<\alpha>}\Gamma(ds)\Big)\,dt \\
&\quad := I_1 + I_2.
\end{aligned} \tag{3.49}
$$

The integrand of I_1 converges to

$$
-\alpha a\cos\Big(tx - a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(ds)\Big)\times\int_{S_2}|ts_1+rs_2|^{\alpha-1}s_2\Gamma(ds)\times\exp\Big\{-\int_{S_2}|ts_1+rs_2|^{\alpha}\Gamma(ds)\Big\}
$$

Using the mean value theorem, the triangle inequality and the inequality $-|x+y|^{\alpha} \le -|x|^{\alpha}+|y|^{\alpha}$ when $0<\alpha<1$, the integrand of I_1 can be bounded for any h, $|h|<|r|$, by

$$
\begin{aligned}
&\Big|cos(y)\Big|\Big(\Big|\frac{a}{h}\Big|\int_{S_2}\Big|(ts_1+(r+h)s_2)^{<\alpha>}-(ts_1+rs_2)^{<\alpha>}\Big|\Gamma(ds)\Big)\exp\Big\{\int_{S_2}-|ts_1|^{\alpha}+|rs_2|^{\alpha}\Gamma(ds)\Big\} \\
&\le 2|a|e^{|r|^{\alpha}\sigma_2^{\alpha}}e^{-\sigma_1^{\alpha}|t|^{\alpha}}\int_{S_2}|ts_1+rs_2|^{\alpha-1}\Gamma(ds),
\end{aligned} \tag{3.50}
$$

where $\sigma_2 = \Big(\int_{S_2}|s_2|^{\alpha}\Gamma(d\boldsymbol{s})\Big)^{1/\alpha}$, $y\in\mathbb{R}$, and we used the bound

$$
\left|\frac{(ts_1+(r+h)s_2)^{<\alpha>}-(ts_1+rs_2)^{<\alpha>}}{h}\right| \le 2|ts_1+rs_2|^{\alpha-1}|s_2|, \tag{3.51}
$$

for $ts_1+rs_2\neq 0$, which is a consequence of $||1+z|^{<\alpha>}-1|\le 2|z|$, for $z\in\mathbb{R}$ (see Lemma 3.7.5 ($\iota\iota$) below). Bound (3.50) does not depend on h and is integrable with respect to t. Indeed, invoking Lemma 3.7.7 with

$\eta = \alpha - 1$, $b = p = 0$, and (3.3) with $\nu > 2 - \alpha > 1 - \alpha$

$$\begin{aligned}
&\left| \int_{\mathbb{R}} e^{-\sigma_1^\alpha |t|^\alpha} \int_{S_2} \left| t + r\frac{s_2}{s_1} \right|^{\alpha-1} |s_1|^{\alpha-1} \Gamma(ds) dt - \int_{\mathbb{R}} \int_{S_2} e^{-\sigma_1^\alpha |t|^\alpha} |t|^{\alpha-1} |s_1|^{\alpha-1} \Gamma(d\boldsymbol{s}) dt \right| \\
&\qquad \le \int_{S_2} |s_1|^{\alpha-1} \int_{\mathbb{R}} e^{-\sigma_1^\alpha |t|^\alpha} \left| \left| t + r\frac{s_2}{s_1} \right|^{\alpha-1} - |t|^{\alpha-1} \right| dt \Gamma(ds) \\
&\qquad \le \text{const} \int_{S_2} |s_1|^{\alpha-1+\nu} |s_1|^{-\nu} \Gamma(ds) \\
&\qquad \le \text{const} \int_{S_2} |s_1|^{-\nu} \Gamma(ds) \\
&\qquad < +\infty,
\end{aligned} \tag{3.52}$$

and the integrability with respect to t follows from the fact that $\int_{\mathbb{R}} e^{-\sigma_1^\alpha |t|^\alpha} |t|^{\alpha-1} dt < +\infty$. Hence the Lebesgue dominated convergence theorem applies to I_1 and we can invert integration and derivation. Focusing on I_2, its integrand tends to

$$-\alpha \int_{S_2} (ts_1 + rs_2)^{<\alpha-1>} s_2 \Gamma(ds) \exp\left\{ -\int_{S_2} |ts_1 + rs_2|^\alpha \Gamma(ds) \right\} \sin\left(tx - a \int_{S_2} |ts_1 + rs_2|^{<\alpha>} \Gamma(ds) \right).$$

Using the inequality

$$\left| \frac{(ts_1 + (r+h)s_2)^\alpha - (ts_1 + rs_2)^\alpha}{h} \right| \le |ts_1 + rs_2|^{\alpha-1} |s_2|,$$

for $ts_1 + rs_2 \neq 0$, which is a consequence of $||1+z|^\alpha - 1| \le |z|$, for $z \in \mathbb{R}$ (Lemma 3.7.5 (ι) below) and the inequality $|e^{-x} - e^{-y}| \le e^{-y} e^{|x-y|} |x - y|$, for $x, y \in \mathbb{R}$, we can bound the integrand of I_2 for any $|h| < |r|$ by

$$\begin{aligned}
&\exp\left\{ -\int_{S_2} |ts_1 + rs_2|^\alpha \Gamma(ds) \right\} \exp\left\{ \left| \int_{S_2} |ts_1 + (r+h)s_2|^\alpha - |ts_1 + rs_2|^\alpha \Gamma(ds) \right| \right\} \\
&\qquad\qquad\qquad\qquad \times \left| \frac{1}{h} \int_{S_2} |ts_1 + (r+h)s_2|^\alpha - |ts_1 + rs_2|^\alpha \Gamma(ds) \right| \\
&\le e^{2|r|^\alpha \sigma_2^\alpha} e^{-\sigma_1^\alpha |t|^\alpha} \int_{S_2} \left| t + r\frac{s_2}{s_1} \right|^{\alpha-1} |s_1|^{\alpha-1} \Gamma(ds).
\end{aligned}$$

The integrability with respect to t is deduced as for (3.52) using Lemma 3.7.7 with $\eta = \alpha - 1$, $b = p = 0$. Thus, the Lebesgue-dominated convergence theorem applies to I_2 and we can invert integration and derivation. The real part of $\phi_{X_2|x}(r)$ can be treated in a similar way, allowing us to derivate under the integral.

Case $\boldsymbol{\alpha \in (1,2)}$

Assume $\alpha \in (1,2)$. Just as for the case $\alpha \in (0,1)$, the imaginary part of $\phi_{X_2|x}$ is given by (3.49)

$$\frac{d}{dr}\Big(\mathrm{Im}\phi_{X_2|x}(r) \Big) = I_1 + I_2.$$

The integrands of I_1 and I_2 still converges to the same limits, however a different argument is needed to bound them. For $|h| < |r|$, the mean value theorem, the triangle inequality and the inequality of Lemma 3.7.6, yield the following bound for the integrand of I_1

$$\left(\left| \frac{a}{h} \right| \int_{S_2} \left| (ts_1 + (r+h)s_2)^{<\alpha>} - (ts_1 + rs_2)^{<\alpha>} \right| \Gamma(ds) \right) e^{|r|^\alpha \sigma_2^\alpha} e^{-2^{1-\alpha} \sigma_1^\alpha |t|^\alpha}, \tag{3.53}$$

where $y \in \mathbb{R}$. By the triangle inequality and the mean value theorem, we have for some $u \in \Bigg(\min\Big(ts_1 + (r+h)s_2, ts_1 + rs_2\Big), \max\Big(ts_1 + (r+h)s_2, ts_1 + rs_2\Big)\Bigg)$

$$\begin{aligned}\left|\int_{S_2}(ts_1+(r+h)s_2)^{<\alpha>} - (ts_1+rs_2)^{<\alpha>}\Gamma(d\boldsymbol{s})\right| &= \left|\int_{S_2}\alpha h s_2|u|^{\alpha-1}\Gamma(d\boldsymbol{s})\right| \\ &\leq \alpha|h|\left|\int_{S_2}|t|^{\alpha-1} + 2|r|^{\alpha-1}\Gamma(d\boldsymbol{s})\right. \\ &\leq \alpha|h|\Gamma(S_2)(|t|^{\alpha-1} + 2|r|^{\alpha-1}) \end{aligned} \tag{3.54}$$

Thus, (3.53) can be bounded by

$$\alpha|a|\Gamma(S_2)e^{|r|^\alpha\sigma_2^\alpha}e^{-2^{1-\alpha}\sigma_1^\alpha|t|^\alpha}(|t|^{\alpha-1} + 2|r|^{\alpha-1}),$$

which is certainly integrable with respect to t on $\mathbb{R}$ for $\alpha > 1$. Let us now turn to I_2. We have again by the mean value theorem,

$$\left|\frac{|ts_1+(r+h)s_2|^\alpha - |ts_1+rs_2|^\alpha}{h}\right| \leq \alpha(|t|^{\alpha-1} + 2|r|^{\alpha-1}),$$

if $|h| < |r|$, and thus

$$\begin{aligned}&\left|\frac{e^{-\int_{S_2}|ts_1+(r+h)s_2|^\alpha\Gamma(d\boldsymbol{s})} - e^{-\int_{S_2}|ts_1+rs_2|^\alpha\Gamma(d\boldsymbol{s})}}{h}\right| \\ &\qquad\leq \max\left(e^{-\int_{S_2}|ts_1+(r+h)s_2|^\alpha\Gamma(d\boldsymbol{s})}, e^{-\int_{S_2}|ts_1+rs_2|^\alpha\Gamma(d\boldsymbol{s})}\right) \\ &\qquad\quad\times\int_{S_2}\left|\frac{|ts_1+(r+h)s_2|^\alpha - |ts_1+rs_2|^\alpha}{h}\right|\Gamma(d\boldsymbol{s}) \\ &\qquad\leq \Gamma(S_2)e^{|2r|^\alpha\sigma_2^\alpha}e^{-2^{1-\alpha}\sigma_1^\alpha|t|^\alpha}\alpha(|t|^{\alpha-1} + 2|r|^{\alpha-1}),\end{aligned} \tag{3.55}$$

by Lemma 3.7.3 (3.77) and Lemma 3.7.6. The latter bound is again integrable with respect to t on $\mathbb{R}$. Hence the dominated convergence theorem applies to I_1, I_2 and therefore to $\frac{d}{dr}\Big(\mathrm{Im}\phi_{X_2|x}(r)\Big)$ and we can invert the integration and derivation signs. Similar arguments show the dominated convergence theorem applies to the real part of the conditional characteristic function as well.

Justifying inversion of integral and derivation signs: Second derivative   Case $\alpha \in (1/2, 1)$

In an expanded fashion, $\phi^{(1)}_{X_2|x}(r)$ can be written,

$$\phi^{(1)}_{X_2|x}(r) = \frac{-\alpha}{2\pi f_{X_1}(x)}\Big[J_1 - aJ_2 - i(J_3 + aJ_4)\Big], \tag{3.56}$$

with,

$$J_1(r) = \int_{\mathbb{R}} e^{-\int_{S_2} |ts_1+rs_2|^{\alpha}\Gamma(ds)} \cos\left(tx - a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(ds)\right)\int_{S_2}(ts_1+rs_2)^{<\alpha-1>}s_2\Gamma(ds)dt,$$

$$J_2(r) = \int_{\mathbb{R}} e^{-\int_{S_2} |ts_1+rs_2|^{\alpha}\Gamma(ds)} \sin\left(tx - a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(ds)\right)\int_{S_2}|ts_1+rs_2|^{\alpha-1}s_2\Gamma(ds)dt,$$

$$J_3(r) = \int_{\mathbb{R}} e^{-\int_{S_2} |ts_1+rs_2|^{\alpha}\Gamma(ds)} \sin\left(tx - a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(ds)\right)\int_{S_2}(ts_1+rs_2)^{<\alpha-1>}s_2\Gamma(ds)dt,$$

$$J_4(r) = \int_{\mathbb{R}} e^{-\int_{S_2} |ts_1+rs_2|^{\alpha}\Gamma(ds)} \cos\left(tx - a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(ds)\right)\int_{S_2}|ts_1+rs_2|^{\alpha-1}s_2\Gamma(ds)dt.$$

To obtain $\phi^{(2)}_{X_2|x}(r)$, we will show that the dominated convergence theorem applies to J_1'. Let us consider,

$$\begin{aligned}
J_1'(r) = \lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}\Bigg[&\exp\Big\{-\int_{S_2}|ts_1+(r+h)s_2|^{\alpha}\Gamma(d\boldsymbol{s})\Big\}\cos\Big(tx - a\int_{S_2}(ts_1+(r+h)s_2)^{<\alpha>}\Gamma(ds)\Big)\\
&\times\int_{S_2}(ts_1+(r+h)s_2)^{<\alpha-1>}s_2\Gamma(ds)\\
&-\exp\Big\{-\int_{S_2}|ts_1+rs_2|^{\alpha}\Gamma(d\boldsymbol{s})\Big\}\cos\Big(tx - a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(ds)\Big)\\
&\times\int_{S_2}(ts_1+rs_2)^{<\alpha-1>}s_2\Gamma(ds)\Bigg]dt\\
= \lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}&\Bigg[\exp\Big\{-\int_{S_2}|ts_1+(r+h)s_2|^{\alpha}\Gamma(d\boldsymbol{s})\Big\} - \exp\Big\{-\int_{S_2}|ts_1+rs_2|^{\alpha}\Gamma(d\boldsymbol{s})\Big\}\Bigg]\\
&\times\cos\Big(tx - a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(ds)\Big)\int_{S_2}(ts_1+rs_2)^{<\alpha-1>}s_2\Gamma(ds)dt\\
+\lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}&\exp\Big\{-\int_{S_2}|ts_1+(r+h)s_2|^{\alpha}\Gamma(d\boldsymbol{s})\Big\} \qquad (3.57)\\
&\times\Bigg[\cos\Big(tx - a\int_{S_2}(ts_1+(r+h)s_2)^{<\alpha>}\Gamma(ds)\Big)\\
&-\cos\Big(tx - a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(ds)\Big)\Bigg]\\
&\times\int_{S_2}(ts_1+rs_2)^{<\alpha-1>}s_2\Gamma(d\boldsymbol{s})dt\\
+\lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}&\exp\Big\{-\int_{S_2}|ts_1+(r+h)s_2|^{\alpha}\Gamma(d\boldsymbol{s})\Big\}\cos\Big(tx - a\int_{S_2}(ts_1+(r+h)s_2)^{<\alpha>}\Gamma(ds)\Big)\\
&\times\Bigg[\int_{S_2}(ts_1+(r+h)s_2)^{<\alpha-1>}s_2\Gamma(ds) - \int_{S_2}(ts_1+rs_2)^{<\alpha-1>}s_2\Gamma(ds)\Bigg]dt\\
:= K_1 + K_2 + K_3. & \qquad (3.58)
\end{aligned}$$

It can be shown that the dominated convergence theorem applies to K_1 following the proof in [30] (p.105) for I_1. Consider K_2. The integrand converges to

$$\alpha a\bigg(\int_{S_2}|ts_1+rs_2|^{\alpha-1}s_2\Gamma(ds)\bigg)\bigg(\int_{S_2}(ts_1+rs_2)^{<\alpha-1>}s_2\Gamma(ds)\bigg)$$
$$\times\sin\Big(tx-a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(ds)\Big)\exp\Big\{-\int_{S_2}|ts_1+rs_2|^{\alpha}\Gamma(ds)\Big\}.$$

Using the mean value theorem, (3.51) and the triangle inequality, we can bound the integrand for any $|h|<|r|$ by

$$\begin{aligned}
&\left|\frac{1}{h}\int_{S_2}(ts_1+(r+h)s_2)^{<\alpha>}-(ts_1+rs_2)^{<\alpha>}\Gamma(ds)\right|\\
&\qquad\times|\sin(y)|e^{2|r|^{\alpha}\sigma_2^{\alpha}}e^{-|t|^{\alpha}\sigma_1^{\alpha}}\int_{S_2}\left|t+r\frac{s_2}{s_1}\right|^{\alpha-1}|s_2||s_1|^{\alpha-1}\Gamma(ds)\\
&\le 2e^{2|r|^{\alpha}\sigma_2^{\alpha}}\bigg(\int_{S_2}\left|t+r\frac{s_2}{s_1}\right|^{\alpha-1}|s_1|^{\alpha-1}\Gamma(ds)\bigg)^2e^{-|t|^{\alpha}\sigma_1^{\alpha}}
\end{aligned}\tag{3.59}$$

where $y\in\mathbb{R}$. The bound (3.59) does not depend on h and is integrable with respect to t: invoking (2.9) Lemma 2.2 in [30],

$$\begin{aligned}
&\bigg|\int_{\mathbb{R}}\int_{S_2}\int_{S_2}e^{-\sigma_1^{\alpha}|t|^{\alpha}}\left|t+r\frac{s_2}{s_1}\right|^{\alpha-1}\left|t+r\frac{s_2'}{s_1'}\right|^{\alpha-1}|s_1'|^{\alpha-1}|s_1|^{\alpha-1}\Gamma(ds)\Gamma(ds')dt \qquad (3.60)\\
&\qquad\qquad-\int_{\mathbb{R}}\int_{S_2}\int_{S_2}e^{-\sigma_1^{\alpha}|t|^{\alpha}}|t|^{2\alpha-2}dt\Gamma(d\boldsymbol{s})\Gamma(d\boldsymbol{s}')\bigg|\\
&=\bigg|\int_{S_2}\int_{S_2}|s_1'|^{\alpha-1}|s_1|^{\alpha-1}\int_{\mathbb{R}}e^{-\sigma_1^{\alpha}|t|^{\alpha}}\bigg[\left|t+r\frac{s_2}{s_1}\right|^{\alpha-1}\left|t+r\frac{s_2'}{s_1'}\right|^{\alpha-1}-\left|t+r\frac{s_2}{s_1}\right|^{\alpha-1}\Big|t\Big|^{\alpha-1}\\
&\qquad\qquad+\left|t+r\frac{s_2}{s_1}\right|^{\alpha-1}\Big|t\Big|^{\alpha-1}-|t|^{2\alpha-2}\bigg]dt\Gamma(ds)\Gamma(ds')\bigg|\\
&\le\int_{S_2}\int_{S_2}|s_1'|^{\alpha-1}|s_1|^{\alpha-1}\int_{\mathbb{R}}e^{-\sigma_1^{\alpha}|t|^{\alpha}}\bigg[\bigg|\left|t+r\frac{s_2'}{s_1'}\right|^{\alpha-1}-|t|^{\alpha-1}\bigg|\left|t+r\frac{s_2}{s_1}\right|^{\alpha-1}\\
&\qquad\qquad+\bigg|\left|t+r\frac{s_2}{s_1}\right|^{\alpha-1}-|t|^{\alpha-1}\bigg||t|^{\alpha-1}\bigg]dt\Gamma(ds)\Gamma(ds')\\
&\le\text{const}\bigg(\int_{S_2}|s_1|^{\alpha-1}\Gamma(ds)\bigg)^2\\
&<+\infty,
\end{aligned}\tag{3.61}$$

where const is a constant depending only on α and σ_1^{α}. The integrability of (3.59) follows from (3.61), the fact that $\int_{\mathbb{R}}e^{-\sigma_1^{\alpha}|t|^{\alpha}}|t|^{2\alpha-2}dt<+\infty$ and (3.3) with $\nu>2-\alpha>1-\alpha$. Hence the dominated convergence theorem applies to K_2. Let us now turn to K_3: «this [a] case when appropriate "integration by part" is needed» ([30]). With the change of variable $t' = t + \dfrac{hs_2'}{s_1'}$,

$$
\begin{aligned}
K_3 = \lim_{h\to 0}\frac{1}{h}\Bigg[\int_{\mathbb{R}} &\exp\Big\{-\int_{S_2}|ts_1+(r+h)s_2|^\alpha\Gamma(d\boldsymbol{s})\Big\}\cos\Big(tx-a\int_{S_2}(ts_1+(r+h)s_2)^{<\alpha>}\Gamma(ds)\Big)\\
&\qquad\qquad\times\int_{S_2}(t+\frac{hs_2'}{s_1'}+\frac{rs_2'}{s_1'})^{<\alpha-1>}s_2's_1'^{<\alpha-1>}\Gamma(ds')dt\\
&-\int_{\mathbb{R}}\exp\Big\{-\int_{S_2}|ts_1+(r+h)s_2|^\alpha\Gamma(d\boldsymbol{s})\Big\}\cos\Big(tx-a\int_{S_2}(ts_1+(r+h)s_2)^{<\alpha>}\Gamma(ds)\Big)\\
&\qquad\qquad\times\int_{S_2}(t+\frac{rs_2'}{s_1'})^{<\alpha-1>}s_2's_1'^{<\alpha-1>}\Gamma(ds')dt\Bigg]\\
=\lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}\int_{S_2}&\Bigg[\exp\Bigg\{-\int_{S_2}\Big|\Big(t-\frac{hs_2'}{s_1'}\Big)s_1+(r+h)s_2\Big|^\alpha\Gamma(d\boldsymbol{s})\Bigg\}\\
&\times\cos\Bigg(\Big(t-\frac{hs_2'}{s_1'}\Big)x-a\int_{S_2}\Big(\Big(t-\frac{hs_2'}{s_1'}\Big)s_1+(r+h)s_2\Big)^{<\alpha>}\Gamma(ds)\Bigg)\\
&-\exp\Big\{-\int_{S_2}|ts_1+(r+h)s_2|^\alpha\Gamma(d\boldsymbol{s})\Big\}\cos\Big(tx-a\int_{S_2}(ts_1+(r+h)s_2)^{<\alpha>}\Gamma(ds)\Big)\Bigg]\\
&\times\Big(t+r\frac{s_2'}{s_1'}\Big)^{<\alpha-1>}s_2's_1'^{<\alpha-1>}\Gamma(ds')dt\\
=\lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}\int_{S_2}\frac{1}{\frac{hs_2'}{s_1'}}&\Bigg[\cos\Bigg(\Big(t-\frac{hs_2'}{s_1'}\Big)x-a\int_{S_2}\Big(\Big(t-\frac{hs_2'}{s_1'}\Big)s_1+(r+h)s_2\Big)^{<\alpha>}\Gamma(ds)\Bigg)\\
&-\cos\Big(tx-a\int_{S_2}(ts_1+(r+h)s_2)^{<\alpha>}\Gamma(ds)\Big)\Bigg]\\
&\times\exp\Big\{-\int_{S_2}|ts_1+(r+h)s_2|^\alpha\Gamma(d\boldsymbol{s})\Big\}\Big(t+r\frac{s_2'}{s_1'}\Big)^{<\alpha-1>}s_2'^2|s_1'|^{\alpha-2}\Gamma(ds')dt\\
+\lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}\int_{S_2}\frac{1}{\frac{hs_2'}{s_1'}}&\Bigg[\exp\Bigg\{-\int_{S_2}\Big|\Big(t-\frac{hs_2'}{s_1'}\Big)s_1+(r+h)s_2\Big|^\alpha\Gamma(d\boldsymbol{s})\Bigg\}-\exp\Big\{-\int_{S_2}|ts_1+(r+h)s_2|^\alpha\Gamma(d\boldsymbol{s})\Big\}\Bigg]\\
&\times\cos\Bigg(\Big(t-\frac{hs_2'}{s_1'}\Big)x-a\int_{S_2}\Big(\Big(t-\frac{hs_2'}{s_1'}\Big)s_1+(r+h)s_2\Big)^{<\alpha>}\Gamma(ds)\Bigg)\\
&\times\Big(t+r\frac{s_2'}{s_1'}\Big)^{<\alpha-1>}s_2'^2|s_1'|^{\alpha-2}\Gamma(ds')dt\\
=K_{31}+K_{32}.&
\end{aligned}
$$

The case of K_{32} is similar to that of I_{22} in [30] (p.106-108), the dominated convergence theorem applies. We focus on K_{31}. Its integrand converges to

$$
\begin{aligned}
\sin\Bigg(tx-a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(ds)\Bigg)&\exp\Bigg\{-\int_{S_2}|ts_1+rs_2|^\alpha\Gamma(ds)\Bigg\}\\
&\times\Bigg(x-\alpha a\int_{S_2}|ts_1+rs_2|^{\alpha-1}s_1\Gamma(ds)\Bigg)\Bigg(\int_{S_2}(ts_1'+rs_2')^{<\alpha-1>}s_2'^2s_1'^{-1}\Gamma(ds')\Bigg).
\end{aligned}
$$

Using the mean value theorem and Lemma 3.7.5 ($\iota\iota$), we can bound the integrand of K_{31} for any $|h|<|r|$

by

$$
\begin{aligned}
&|\sin(y)|e^{2|r|^\alpha\sigma_2^\alpha}e^{-|t|^\alpha\sigma_1^\alpha}\int_{S_2}\left|t+r\frac{s_2'}{s_1'}\right|^{\alpha-1}{s_2'}^2|s_1'|^{\alpha-2}\\
&\quad\times\left|\frac{1}{\frac{hs_2'}{s_1'}}\right|\left|-\frac{hs_2'}{s_1'}x-a\int_{S_2}\left(\left(\left(t-\frac{hs_2'}{s_1'}\right)s_1+(r+h)s_2\right)^{<\alpha>}-(ts_1+(r+h)s_2)^{<\alpha>}\Gamma(ds)\right)\right|\Gamma(ds')\\
&\quad\le e^{2|r|^\alpha\sigma_2^\alpha}e^{-|t|^\alpha\sigma_1^\alpha}\int_{S_2}\left|t+r\frac{s_2'}{s_1'}\right|^{\alpha-1}{s_2'}^2|s_1'|^{\alpha-2}\left(|x|+2a\int_{S_2}\left|t+(r+h)\frac{s_2}{s_1}\right|^{\alpha-1}|s_1|\Gamma(ds)\right)\Gamma(ds')\\
&\quad\le |x|e^{2|r|^\alpha\sigma_2^\alpha}e^{-|t|^\alpha\sigma_1^\alpha}\int_{S_2}\left|t+r\frac{s_2'}{s_1'}\right|^{\alpha-1}{s_2'}^2|s_1'|^{\alpha-2}\Gamma(ds')\\
&\qquad+2ae^{2|r|^\alpha\sigma_2^\alpha}e^{-|t|^\alpha\sigma_1^\alpha}\int_{S_2}\int_{S_2}\left|t+r\frac{s_2'}{s_1'}\right|^{\alpha-1}\left|t+(r+h)\frac{s_2}{s_1}\right|^{\alpha-1}|s_1|{s_2'}^2|s_1'|^{\alpha-2}\Gamma(ds)\Gamma(ds').
\end{aligned}
$$

The integrability with respect to t of the first (resp. second) term is obtained in the same way as for (3.52) (resp. (3.61)) and concluding using (3.3) with $\nu > 2-\alpha$. Thus, the dominated convergence theorem applies to K_{31}, which finally shows that the dominated convergence theorem applies to J_1'. The other J's can be treated in a similar fashion.

Case $\boldsymbol{\alpha \in (1,2)}$

After derivation, $\phi^{(1)}_{X_2|x}(r)$ is given by (3.56) with functions J's of the form

$$
\int_{\mathbb{R}} e^{-\int_{S_2}|ts_1+rs_2|^\alpha\Gamma(ds)}\mathrm{trig}\left(tx-a\int_{S_2}|ts_1+rs_2|^{<\alpha>}\Gamma(ds)\right)\int_{S_2}(ts_1+rs_2)^{<\alpha-1>\text{ or }\alpha-1}s_2\Gamma(ds)dt,
$$

which are similar to deal with. Consider for instance $J_1(r)$. It's derivative can be written as in (3.58)

$$
J_1'(r) = K_1 + K_2 + K_3.
$$

For the integrand of K_1, we can use (3.55) and the triangle inequality to bound it by

$$
\Gamma(S_2)e^{|2r|^\alpha\sigma_2^\alpha}e^{-2^{1-\alpha}\sigma_1^\alpha|t|^\alpha}\alpha(|t|^{\alpha-1}+2|r|^{\alpha-1})\int_{S_2}|ts_1+rs_2|^{\alpha-1}|s_2|\Gamma(ds).
$$

Since $0<\alpha-1<1$, we can further bound it by

$$
\Gamma(S_2)e^{|2r|^\alpha\sigma_2^\alpha}e^{-2^{1-\alpha}\sigma_1^\alpha|t|^\alpha}\alpha(|t|^{\alpha-1}+2|r|^{\alpha-1})^2,
$$

which is integrable with respect to t. The same bound can be obtained for the integrand of K_2 using the mean value theorem, (3.54) and Lemma 3.7.6. As for K_3, there is no need to perform "appropriate integration by parts" since $0<\alpha-1<1$. Its integrand converges to

$$
(\alpha-1)\exp\left\{-\int_{S_2}|ts_1+rs_2|^\alpha\Gamma(d\boldsymbol{s})\right\}\cos\left(tx-a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(ds)\right)\int_{S_2}|ts_1+rs_2|^{\alpha-2}s_2^2\Gamma(ds).
$$

Using Lemmas 3.7.6 and 3.7.5 ($\iota\iota$), it can be bounded for any $|h| < |r|$ by

$$\frac{2}{|h|}\Gamma(S_2)e^{|2r|^\alpha\sigma_2^\alpha}e^{-2^{1-\alpha}\sigma_1^\alpha|t|^\alpha}\int_{S_2}|ts_1+rs_2|^{\alpha-2}|hs_2|\Gamma(ds),$$
$$\leq \Gamma(S_2)e^{|2r|^\alpha\sigma_2^\alpha}e^{-2^{1-\alpha}\sigma_1^\alpha|t|^\alpha}\int_{S_2}\Big|t+\frac{rs_2}{s_1}\Big|^{\alpha-2}|s_1|^{\alpha-2}\Gamma(ds).$$

We can show that this bound is integrable with respect to t using Lemma 3.7.7 with $\eta=\alpha-2$, $b=0$ and $p=0$, the fact that $\int_{\mathbb{R}} e^{-2^{1-\alpha}\sigma_1^\alpha|t|^\alpha}|t|^{\alpha-2}dt < +\infty$ for $\alpha\in(1,2)$ and (3.3) with $\nu > 2-\alpha$. The dominated convergence theorem thus applies and we get

$$\phi^{(2)}_{X_2|x}(r) = \frac{-\alpha}{2\pi f_{X_1}(x)}\Bigg[-\alpha\int_{\mathbb{R}} e^{-itx}\varphi_{\boldsymbol{X}}(t,r)\Big(\int_{S_2} g_2(ts_1+rs_2)s_2\Gamma(ds)\Big)^2 dt$$
$$+(\alpha-1)\int_{\mathbb{R}} e^{-itx}\varphi_{\boldsymbol{X}}(t,r)\Big(\int_{S_2} g_3(ts_1+rs_2)s_2^2\Gamma(ds)\Big)dt\Bigg], \tag{3.62}$$

with $g_3(z)=|z|^{\alpha-2}-iaz^{<\alpha-2>}$ for $z\in\mathbb{R}$. Integrating by parts the terms $|ts_1+rs_2|^{<\alpha-2> \text{ or } \alpha-2}$ involved in the expression $\int_{\mathbb{R}} e^{-itx}\varphi_{\boldsymbol{X}}(t,r)\Big(\int_{S_2} g_3(ts_1+rs_2)s_2^2\Gamma(ds)\Big)dt$ yields the expression (3.46) obtained in the case $\alpha\in(1/2,1)$. Hence, the same representation for the second order conditional moment of Proposition 3.6.3 holds when $\alpha>1$.

Justifying inversion of integral and derivation signs: third derivative Let $\alpha\in(1,2)$ and let (3.3) hold with $\nu > 3-\alpha$. Starting from the second derivative of $\phi^{(2)}_{X_2|x}(r)$ given at (3.46), with obvious notations

$$\phi^{(2)}_{X_2|x}(r) = \frac{-\alpha}{2\pi f_{X_1}(x)}\Big[ixI_1(r)+\alpha(I_3(r)-I_2(r))\Big]$$

On the one hand, it can be shown that the dominated convergence theorem applies to I_1' using the usual arguments the fact that (3.3) holds with $\nu > 3-\alpha$. On the other hand, after some elementary manipulations, we get that

$$I_3-I_2 = \int_{\mathbb{R}} e^{-itx+ia\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(d\boldsymbol{s})}e^{-\int_{S_2}|ts_1+rs_2|^\alpha\Gamma(d\boldsymbol{s})}$$
$$\times\int_{S_2}\int_{S_2}\Bigg\{(ts_1+rs_2)^{<\alpha-1>}(ts_1'+rs_2')^{<\alpha-1>}-a^2|ts_1+rs_2|^{\alpha-1}|ts_1'+rs_2'|^{\alpha-1}$$
$$-ia\Big(|ts_1+rs_2|^{\alpha-1}(ts_1'+rs_2')^{<\alpha-1>}+(ts_1+rs_2)^{<\alpha-1>}|ts_1'+rs_2'|^{\alpha-1}\Big)\Bigg\}$$
$$\times\Big[s_2^2s_1^{-1}s_1'-s_2s_2'\Big]\Gamma(d\boldsymbol{s})\Gamma(d\boldsymbol{s}')dt$$

The previous expression can be decomposed into terms of the form

$$\int_{\mathbb{R}}\int_{S_2}\int_{S_2}\operatorname{trig}\Big(-tx+a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(d\boldsymbol{s})\Big)$$
$$\times e^{-\int_{S_2}|ts_1+rs_2|^\alpha\Gamma(d\boldsymbol{s})}$$
$$\times|ts_1+rs_2|^{<\alpha-1> \text{ or } \alpha-1}\quad\times\quad|ts_1'+rs_2'|^{<\alpha-1> \text{ or } \alpha-1}$$
$$\times\Big[s_2^2s_1^{-1}s_1'-s_2s_2'\Big]\Gamma(d\boldsymbol{s})\Gamma(d\boldsymbol{s}')dt,$$

where «trig» is to be replaced by a sine or cosine function. Each of these terms can be treated in a similar way to show that the dominated convergence theorem applies. We will consider

$$\begin{aligned}J(r) = \int_{\mathbb{R}}\int_{S_2}\int_{S_2} \cos\left(tx - a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(d\boldsymbol{s})\right)e^{-\int_{S_2}|ts_1+rs_2|^{\alpha}\Gamma(d\boldsymbol{s})} \\ \times |ts_1+rs_2|^{\alpha-1}(ts_1'+rs_2')^{<\alpha-1>}\left[s_2^2 s_1^{-1}s_1' - s_2 s_2'\right]\Gamma(d\boldsymbol{s})\Gamma(d\boldsymbol{s}')dt.\end{aligned}$$

We have

$$\begin{aligned}J'(r) &= \lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}\int_{S_2}\int_{S_2}\Bigg[\cos\left(tx - a\int_{S_2}(ts_1+(r+h)s_2)^{<\alpha>}\Gamma(d\boldsymbol{s})\right) \\ &\qquad - \cos\left(tx - a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(d\boldsymbol{s})\right)\Bigg] \\ &\qquad \times e^{-\int_{S_2}|ts_1+(r+h)s_2|^{\alpha}\Gamma(d\boldsymbol{s})}|ts_1+(r+h)s_2|^{\alpha-1}(ts_1'+(r+h)s_2')^{<\alpha-1>} \\ &\qquad \times \left[s_2^2 s_1^{-1}s_1' - s_2 s_2'\right]\Gamma(d\boldsymbol{s})\Gamma(d\boldsymbol{s}')dt \\ &+ \lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}\int_{S_2}\int_{S_2}\cos\left(tx - a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(d\boldsymbol{s})\right) \\ &\qquad \times\left[e^{-\int_{S_2}|ts_1+(r+h)s_2|^{\alpha}\Gamma(d\boldsymbol{s})} - e^{-\int_{S_2}|ts_1+rs_2|^{\alpha}\Gamma(d\boldsymbol{s})}\right] \\ &\qquad \times |ts_1+(r+h)s_2|^{\alpha-1}(ts_1'+(r+h)s_2')^{<\alpha-1>}\left[s_2^2 s_1^{-1}s_1' - s_2 s_2'\right]\Gamma(d\boldsymbol{s})\Gamma(d\boldsymbol{s}')dt \\ &+ \lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}\int_{S_2}\int_{S_2}\cos\left(tx - a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(d\boldsymbol{s})\right)e^{-\int_{S_2}|ts_1+rs_2|^{\alpha}\Gamma(d\boldsymbol{s})} \\ &\qquad \times\left[|ts_1+(r+h)s_2|^{\alpha-1} - |ts_1+rs_2|^{\alpha-1}\right] \\ &\qquad \times (ts_1'+(r+h)s_2')^{<\alpha-1>}\left[s_2^2 s_1^{-1}s_1' - s_2 s_2'\right]\Gamma(d\boldsymbol{s})\Gamma(d\boldsymbol{s}')dt \\ &+ \lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}\int_{S_2}\int_{S_2}\cos\left(tx - a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(d\boldsymbol{s})\right)e^{-\int_{S_2}|ts_1+rs_2|^{\alpha}\Gamma(d\boldsymbol{s})} \\ &\qquad \times\left[(ts_1'+(r+h)s_2')^{<\alpha-1>} - (ts_1'+rs_2')^{<\alpha-1>}\right] \\ &\qquad \times |ts_1+rs_2|^{\alpha-1}\left[s_2^2 s_1^{-1}s_1' - s_2 s_2'\right]\Gamma(d\boldsymbol{s})\Gamma(d\boldsymbol{s}')dt \\ &:= K_1+K_2+K_3+K_4.\end{aligned}$$

We will show that we can apply the dominated convergence theorem to the K_i's. Let us begin with K_1. Its integrand converges to

$$\begin{aligned}\alpha a\int_{S_2\times S_2\times S_2}\sin\left(tx - a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(d\boldsymbol{s})\right)e^{-\int_{S_2}|ts_1+rs_2|^{\alpha}\Gamma(d\boldsymbol{s})} \\ \times |ts_1+rs_2|^{\alpha-1}(ts_1'+rs_2')^{<\alpha-1>}|ts_1''+rs_2''|^{\alpha-1}s_2''\left[s_2^2 s_1^{-1}s_1' - s_2 s_2'\right]\Gamma(d\boldsymbol{s})\Gamma(d\boldsymbol{s}')\Gamma(d\boldsymbol{s}'').\end{aligned}$$

For any h, $|h| < |r|$, the integrand of K_1 can be bounded using the mean value theorem on the cosine and

Lemma 3.7.6 by

$$\frac{|a|}{|h|}\left|\int_{S_2}(ts_1+(r+h)s_2)^{<\alpha>}-(ts_1+rs_2)^{<\alpha>}\Gamma(d\boldsymbol{s})\right|e^{2^{\alpha}|r|^{\alpha}\sigma_2^{\alpha}}e^{-2^{1-\alpha}\sigma_1^{\alpha}|t|^{\alpha}}$$
$$\times\left|\int_{S_2}\int_{S_2}|ts_1+(r+h)s_2|^{\alpha-1}(ts_1'+(r+h)s_2')^{<\alpha-1>}\left[s_2^2s_1^{-1}s_1'-s_2s_2'\right]\Gamma(d\boldsymbol{s})\Gamma(d\boldsymbol{s}')\right|. \tag{3.63}$$

Hence, by inequality (3.54) and given that $0<\alpha-1<1$, the quantity (3.63) can be bounded by

$$\begin{aligned}
&\alpha|a|\Gamma(S_2)e^{2^{\alpha}|r|^{\alpha}\sigma_2^{\alpha}}e^{-2^{1-\alpha}\sigma_1^{\alpha}|t|^{\alpha}}(|t|^{\alpha-1}+2|r|^{\alpha-1})\\
&\qquad\times\left|\int_{S_2}\int_{S_2}|ts_1+(r+h)s_2|^{\alpha-1}(ts_1'+(r+h)s_2')^{<\alpha-1>}\left[s_2^2s_1^{-1}s_1'-s_2s_2'\right]\Gamma(d\boldsymbol{s})\Gamma(d\boldsymbol{s}')\right|\\
&\le\alpha|a|\Gamma(S_2)e^{2^{\alpha}|r|^{\alpha}\sigma_2^{\alpha}}e^{-2^{1-\alpha}\sigma_1^{\alpha}|t|^{\alpha}}(|t|^{\alpha-1}+2|r|^{\alpha-1})^3\left(\Gamma(S_2)+\int_{S_2}|s_1|^{-1}\Gamma(d\boldsymbol{s})\right)\\
&\le\text{const }e^{-2^{1-\alpha}\sigma_1^{\alpha}|t|^{\alpha}}(|t|^{\alpha-1}+2|r|^{\alpha-1})^3,
\end{aligned}$$

where const is a finite nonnegative constant because of (3.3) with $\nu>3-\alpha>1$ and the fact that Γ is a finite measure. This last bound, independent of h, is integrable with respect to t on $\mathbb{R}$. The dominated convergence theorem applies to K_1. Consider now K_2. Its integrand converges to

$$\alpha\int_{S_2\times S_2\times S_2}\cos\left(tx-a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(d\boldsymbol{s})\right)e^{-\int_{S_2}|ts_1+rs_2|^{\alpha}\Gamma(d\boldsymbol{s})} \tag{3.64}$$
$$\times|ts_1+rs_2|^{\alpha-1}(ts_1'+rs_2')^{<\alpha-1>}(ts_1''+rs_2'')^{<\alpha-1>}s_2''\left[s_2^2s_1^{-1}s_1'-s_2s_2'\right]\Gamma(d\boldsymbol{s})\Gamma(d\boldsymbol{s}')\Gamma(d\boldsymbol{s}'')$$

By (3.55), the integrand of K_2 can be bounded by

$$\Gamma(S_2)e^{|2r|^{\alpha}\sigma_2^{\alpha}}e^{-2^{1-\alpha}\sigma_1^{\alpha}|t|^{\alpha}}\alpha(|t|^{\alpha-1}+2|r|^{\alpha-1})$$
$$\left|\int_{S_2}\int_{S_2}|ts_1+(r+h)s_2|^{\alpha-1}(ts_1'+(r+h)s_2')^{<\alpha-1>}\left[s_2^2s_1^{-1}s_1'-s_2s_2'\right]\Gamma(d\boldsymbol{s})\Gamma(d\boldsymbol{s}')\right|$$

Which can be further bounded by an integrable function of t in a similar way as for the integrand of K_1. The dominated convergence theorem applies to K_2. Consider now K_3. Its integrand converges to

$$(\alpha-1)\int_{S_2}\int_{S_2}\cos\left(tx-a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(d\boldsymbol{s})\right)e^{-\int_{S_2}|ts_1+rs_2|^{\alpha}\Gamma(d\boldsymbol{s})}$$
$$\times(ts_1+rs_2)^{<\alpha-2>}(ts_1'+(r+h)s_2')^{<\alpha-1>}s_2\left[s_2^2s_1^{-1}s_1'-s_2s_2'\right]\Gamma(d\boldsymbol{s})\Gamma(d\boldsymbol{s}')$$

Using Lemmas 3.7.6, 3.7.5 (ι) and the triangle inequality, the integrand of K_3 can be bounded by

$$\begin{aligned}
&\frac{1}{|h|}e^{|r|^{\alpha}\sigma_2^{\alpha}}e^{-2^{1-\alpha}\sigma_1^{\alpha}|t|^{\alpha}}\int_{S_2}\int_{S_2}|hs_2||ts_1+rs_2|^{\alpha-2}|ts_1'+(r+h)s_2'|^{\alpha-1}\left|s_2^2s_1^{-1}s_1'-s_2s_2'\right|\Gamma(d\boldsymbol{s})\Gamma(d\boldsymbol{s}')\\
&\le e^{|r|^{\alpha}\sigma_2^{\alpha}}\Gamma(S_2)\int_{S_2}e^{-2^{1-\alpha}\sigma_1^{\alpha}|t|^{\alpha}}|ts_1+rs_2|^{\alpha-2}(|t|^{\alpha-1}+2|r|^{\alpha-1})\left|1+|s_1|^{-1}\right|\Gamma(d\boldsymbol{s})
\end{aligned}$$

To show the integrability with respect to t of the last bound we make use of Lemma 3.7.7 with $\eta=\alpha-2$, $b=0,\alpha-1$ and $p=0$ and the fact that with $1<\alpha<2$, $\int_{\mathbb{R}}e^{-2^{1-\alpha}\sigma_1^{\alpha}|t|^{\alpha}}|t|^{\alpha-2}dt<+\infty$ and

$\int_{\mathbb{R}} e^{-2^{1-\alpha}\sigma_1^\alpha|t|^\alpha}|t|^{2\alpha-3}dt < +\infty$

$$
\begin{aligned}
&e^{|r|^\alpha\sigma_2^\alpha}\Gamma(S_2)\int_{S_2}\Big|1+|s_1|^{-1}\Big|\int_{\mathbb{R}} e^{-2^{1-\alpha}\sigma_1^\alpha|t|^\alpha}|s_1|^{\alpha-2}\Big|t+r\frac{s_2}{s_1}\Big|^{\alpha-2}(|t|^{\alpha-1}+2|r|^{\alpha-1})dt\Gamma(d\boldsymbol{s})\\
&\quad\le e^{|r|^\alpha\sigma_2^\alpha}\Gamma(S_2)\int_{S_2}\Big|1+|s_1|^{-1}\Big||s_1|^{\alpha-2}\Bigg[\int_{\mathbb{R}} e^{-2^{1-\alpha}\sigma_1^\alpha|t|^\alpha}\Bigg|\Big|t+r\frac{s_2}{s_1}\Big|^{\alpha-2}-|t|^{\alpha-2}+|t|^{\alpha-2}\Bigg||t|^{\alpha-1}dt\\
&\qquad\qquad+2|r|^{\alpha-1}\int_{\mathbb{R}} e^{-2^{1-\alpha}\sigma_1^\alpha|t|^\alpha}\Bigg|\Big|t+r\frac{s_2}{s_1}\Big|^{\alpha-2}-|t|^{\alpha-2}+|t|^{\alpha-2}\Bigg|dt\Bigg]\Gamma(d\boldsymbol{s})\\
&\quad\le e^{|r|^\alpha\sigma_2^\alpha}\Gamma(S_2)\int_{S_2}\Big|1+|s_1|^{-1}\Big||s_1|^{\alpha-2}\Bigg[\int_{\mathbb{R}} e^{-2^{1-\alpha}\sigma_1^\alpha|t|^\alpha}\Bigg|\Big|t+r\frac{s_2}{s_1}\Big|^{\alpha-2}-|t|^{\alpha-2}\Bigg||t|^{\alpha-1}dt\\
&\qquad\qquad+2|r|^{\alpha-1}\int_{\mathbb{R}} e^{-2^{1-\alpha}\sigma_1^\alpha|t|^\alpha}\Bigg|\Big|t+r\frac{s_2}{s_1}\Big|^{\alpha-2}-|t|^{\alpha-2}\Bigg|dt\\
&\qquad\qquad+\int_{\mathbb{R}} e^{-2^{1-\alpha}\sigma_1^\alpha|t|^\alpha}|t|^{2\alpha-3}dt\\
&\qquad\qquad+2|r|^{\alpha-1}\int_{\mathbb{R}} e^{-2^{1-\alpha}\sigma_1^\alpha|t|^\alpha}|t|^{\alpha-2}dt\Bigg]\Gamma(d\boldsymbol{s})\\
&\quad\le \text{const}\int_{S_2}\Big|1+|s_1|^{-1}\Big||s_1|^{\alpha-2}\Gamma(d\boldsymbol{s})\\
&\quad\le \text{const}\Big(\int_{S_2}|s_1|^{\alpha-2}\Gamma(d\boldsymbol{s})+\int_{S_2}|s_1|^{\alpha-3}\Gamma(d\boldsymbol{s})\Big),
\end{aligned}
$$

which is finite because of (3.3) with $\nu > 3-\alpha$. Hence, the dominated convergence theorem applies to K_3. The case of K_4 is similar, using Lemma 3.7.5 ($\iota\iota$) instead of (ι) to bound the term $\Big|(ts_1'+(r+h)s_2')^{<\alpha-2>}-(ts_1'+rs_2')^{<\alpha-2>}\Big|$. The dominated convergence theorem applies to all the K_i's and we can invert the integration and derivation signs in J'.

Justifying inversion of integral and derivation signs: Fourth derivative Showing that the dominated convergence theorem holds when differentiating (3.65) is the most delicate for the terms: I_5, I_{63} and I_{73} -the terms involving the function g_3, that is, $|ts_1+rs_2|$ to the power $\alpha-2$. Arguments and bounds that have already been encountered can be used for the other ones.

Let us show the dominated convergence theorem applies to I_5. The cases of I_{63} and I_{73} are similar. We decompose I_5 into terms of the form

$$
\begin{aligned}
&\int_{\mathbb{R}}\int_{S_2}\int_{S_2}\text{trig}\Big(-tx+a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(d\boldsymbol{s})\Big)e^{-\int_{S_2}|ts_1+rs_2|^\alpha\Gamma(d\boldsymbol{s})}\\
&\qquad\times|ts_1+rs_2|^{\alpha-1\text{ or }<\alpha-1>}|ts_1'+rs_2'|^{\alpha-2\text{ or }<\alpha-2>}s_2^2s_1^{-1}s_2's_1'\Gamma(d\boldsymbol{s})\Gamma(d\boldsymbol{s}')dt.
\end{aligned}
$$

Consider for instance

$$
\begin{aligned}
J(r) := &\int_{\mathbb{R}}\int_{S_2}\int_{S_2}\cos\Big(-tx+a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(d\boldsymbol{s})\Big)e^{-\int_{S_2}|ts_1+rs_2|^\alpha\Gamma(d\boldsymbol{s})}\\
&\qquad\times|ts_1+rs_2|^{\alpha-1}|ts_1'+rs_2'|^{\alpha-2}s_2^2s_1^{-1}s_2's_1'\Gamma(d\boldsymbol{s})\Gamma(d\boldsymbol{s}')dt.
\end{aligned}
$$

We have

$$
\begin{aligned}
J'(r) = \lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}\int_{S_2}\int_{S_2}&\Big[|ts_1'+(r+h)s_2'|^{\alpha-2}-|ts_1'+rs_2'|^{\alpha-2}\Big]|ts_1+(r+h)s_2|^{\alpha-1}\\
&\qquad\qquad\times\cos\Big(-tx+a\int_{S_2}(ts_1+(r+h)s_2)^{<\alpha>}\Gamma(d\boldsymbol{s})\Big)\\
&\qquad\qquad\qquad\times e^{-\int_{S_2}|ts_1+(r+h)s_2|^{\alpha}\Gamma(d\boldsymbol{s})}s_2^2s_1^{-1}s_2's_1'\Gamma(d\boldsymbol{s})\Gamma(d\boldsymbol{s}')dt\\
&+\lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}\int_{S_2}\int_{S_2}|ts_1'+rs_2'|^{\alpha-2}\Big[|ts_1+(r+h)s_2|^{\alpha-1}-|ts_1+rs_2|^{\alpha-1}\Big]\\
&\qquad\qquad\times\cos\Big(-tx+a\int_{S_2}(ts_1+(r+h)s_2)^{<\alpha>}\Gamma(d\boldsymbol{s})\Big)\\
&\qquad\qquad\qquad\times e^{-\int_{S_2}|ts_1+(r+h)s_2|^{\alpha}\Gamma(d\boldsymbol{s})}s_2^2s_1^{-1}s_2's_1'\Gamma(d\boldsymbol{s})\Gamma(d\boldsymbol{s}')dt\\
&+\lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}\int_{S_2}\int_{S_2}|ts_1'+rs_2'|^{\alpha-2}|ts_1+rs_2|^{\alpha-1}\\
&\times\Big[\cos\Big(-tx+a\int_{S_2}(ts_1+(r+h)s_2)^{<\alpha>}\Gamma(d\boldsymbol{s})\Big)-\cos\Big(-tx+a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(d\boldsymbol{s})\Big)\Big]\\
&\qquad\qquad\qquad\times e^{-\int_{S_2}|ts_1+(r+h)s_2|^{\alpha}\Gamma(d\boldsymbol{s})}s_2^2s_1^{-1}s_2's_1'\Gamma(d\boldsymbol{s})\Gamma(d\boldsymbol{s}')dt\\
&+\lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}\int_{S_2}\int_{S_2}|ts_1'+rs_2'|^{\alpha-2}|ts_1+rs_2|^{\alpha-1}\\
&\qquad\times\cos\Big(-tx+a\int_{S_2}(ts_1+rs_2)^{<\alpha>}\Gamma(d\boldsymbol{s})\Big)\\
&\qquad\quad\times\Big[e^{-\int_{S_2}|ts_1+(r+h)s_2|^{\alpha}\Gamma(d\boldsymbol{s})}-e^{-\int_{S_2}|ts_1+rs_2|^{\alpha}\Gamma(d\boldsymbol{s})}\Big]s_2^2s_1^{-1}s_2's_1'\Gamma(d\boldsymbol{s})\Gamma(d\boldsymbol{s}')dt\\
&:= K_1+K_2+K_3+K_4
\end{aligned}
$$

The integrand of K_4 can be bounded using inequality (3.64), (3.55) and invoking Lemma 3.7.7 and (3.3) with $\nu > 4-\alpha$. The integrand of K_3 can be bounded using (3.54) Lemma 3.7.6, and concluding with Lemma 3.7.7 and (3.3) with $\nu > 4-\alpha$. Focus now on K_2. Using Lemmas 3.7.6 and 3.7.5 (ι), its integrand can be bounded by

$$
e^{|2r|^{\alpha}\sigma_2^{\alpha}}e^{-2^{1-\alpha}\sigma_1^{\alpha}|t|^{\alpha}}\Big|t+\frac{rs_2'}{s_1'}\Big|^{\alpha-2}\Big|t+\frac{rs_2}{s_1}\Big|^{\alpha-2}s_2^3|s_1|^{\alpha-3}|s_1'|^{\alpha-1}|s_2'|.
$$

The later bound does not depend on h and can be shown to be integrable with respect to t using (3.3) with $\nu > 4-\alpha$, Lemma 3.7.8 with $\eta = \alpha-2$, $z_2 = z_4 = 0$, $p = 0$ and the fact that $\int_{\mathbb{R}} e^{-c|t|^{\alpha}}|t|^{2(\alpha-2)} < +\infty$ for $\alpha \in (3/2, 2)$. Let us now turn to the term K_1 which is more intricate. Appropriate «integration by parts» is required. With the change of variable $t = t + \frac{hs_2'}{s_1'}$,

$$
\begin{aligned}
K_1 = \lim_{h\to 0}\frac{1}{h}\int_{S_2}\int_{S_2}\int_{\mathbb{R}} &\left[e^{-\int_{S_2}\left|\left(t-\frac{hs_2'}{s_1'}\right)s_1+(r+h)s_2\right|^{\alpha}\Gamma(d\boldsymbol{s})} - e^{-\int_{S_2}|ts_1+(r+h)s_2|^{\alpha}\Gamma(d\boldsymbol{s})}\right] \\
&\times \cos\left(\left(t-\frac{hs_2'}{s_1'}\right)x - a\int_{S_2}\left(\left(t-\frac{hs_2'}{s_1'}\right)s_1+(r+h)s_2\right)^{<\alpha>}\Gamma(d\boldsymbol{s})\right) \\
&\times \left|\left(t-\frac{hs_2'}{s_1'}\right)s_1+(r+h)s_2\right|^{\alpha-1}|ts_1'+rs_2'|^{\alpha-2}s_2^2s_1^{-1}s_2's_1'dt\Gamma(d\boldsymbol{s})\Gamma(d\boldsymbol{s}') \\
+ \lim_{h\to 0}\frac{1}{h}\int_{S_2}\int_{S_2}\int_{\mathbb{R}} & e^{-\int_{S_2}|ts_1+(r+h)s_2|^{\alpha}\Gamma(d\boldsymbol{s})} \\
&\times \cos\left(\left(t-\frac{hs_2'}{s_1'}\right)x - a\int_{S_2}\left(\left(t-\frac{hs_2'}{s_1'}\right)s_1+(r+h)s_2\right)^{<\alpha>}\Gamma(d\boldsymbol{s})\right) \\
&\times \left[\left|\left(t-\frac{hs_2'}{s_1'}\right)s_1+(r+h)s_2\right|^{\alpha-1} - \left|ts_1+(r+h)s_2\right|^{\alpha-1}\right] \\
&\times |ts_1'+rs_2'|^{\alpha-2}s_2^2s_1^{-1}s_2's_1'dt\Gamma(d\boldsymbol{s})\Gamma(d\boldsymbol{s}') \\
+ \lim_{h\to 0}\frac{1}{h}\int_{S_2}\int_{S_2}\int_{\mathbb{R}} & e^{-\int_{S_2}|ts_1+(r+h)s_2|^{\alpha}\Gamma(d\boldsymbol{s})} \\
&\times \left[\cos\left(\left(t-\frac{hs_2'}{s_1'}\right)x - a\int_{S_2}\left(\left(t-\frac{hs_2'}{s_1'}\right)s_1+(r+h)s_2\right)^{<\alpha>}\Gamma(d\boldsymbol{s})\right)\right. \\
&\left. - \cos\left(tx - a\int_{S_2}\left(ts_1+(r+h)s_2\right)^{<\alpha>}\Gamma(d\boldsymbol{s})\right)\right] \\
&\times \left|ts_1+(r+h)s_2\right|^{\alpha-1}|ts_1'+rs_2'|^{\alpha-2}s_2^2s_1^{-1}s_2's_1'dt\Gamma(d\boldsymbol{s})\Gamma(d\boldsymbol{s}') \\
:= K_{11}+K_{12}+K_{13}.&
\end{aligned}
$$

It can be shown that the generalised Lebesgue convergence theorem applies to the terms K_{11} and K_{12} following the proof in [33] (p.50-52). Regarding the integrand of K_{13}, using the mean value theorem on the cosine, Lemma 3.7.6 and (3.54), we get for $|h| < |r|$

$$
\begin{aligned}
&\frac{1}{\left|\frac{hs_2'}{s_1'}\right|}e^{|2r|^{\alpha}\sigma_2^{\alpha}}e^{-2^{1-\alpha}\sigma_1^{\alpha}|t|^{\alpha}}\left|ts_1+(r+h)s_2\right|^{\alpha-1}|ts_1'+rs_2'|^{\alpha-2}s_2^2|s_1|^{-1}|s_2'|^2 \\
&\qquad \times \left|\frac{hs_2'}{s_1'}x + a\int_{S_2}\left(\left(t-\frac{hs_2'}{s_1'}\right)s_1+(r+h)s_2\right)^{<\alpha>} - \left(ts_1+(r+h)s_2\right)^{<\alpha>}\Gamma(d\boldsymbol{s})\right| \\
&\le \frac{1}{\left|\frac{hs_2'}{s_1'}\right|}e^{|2r|^{\alpha}\sigma_2^{\alpha}}e^{-2^{1-\alpha}\sigma_1^{\alpha}|t|^{\alpha}}\left|ts_1+(r+h)s_2\right|^{\alpha-1}|ts_1'+rs_2'|^{\alpha-2}s_2^2|s_1|^{-1}|s_2'|^2 \\
&\qquad \times \left[\left|\frac{hs_2'}{s_1'}x\right| + \left|a\frac{hs_2'}{s_1'}\right|\int_{S_2}|s_1||ts_1+(r+h)s_2|^{\alpha-1}\Gamma(d\boldsymbol{s})\right] \\
&\le e^{|2r|^{\alpha}\sigma_2^{\alpha}}e^{-2^{1-\alpha}\sigma_1^{\alpha}|t|^{\alpha}}\left|t+\frac{rs_2'}{s_1'}\right|^{\alpha-2}s_2^2|s_1|^{-1}{s_2'}^2|s_1'|^{\alpha-2} \\
&\qquad \times \left(|t|^{\alpha-1}+|2r|^{\alpha-1}\right)\left[|x| + |a|\Gamma(S_2)(|t|^{\alpha-1}+|2r|^{\alpha-1})\right].
\end{aligned}
$$

The last bound can be shown to be integrable with respect to t using Lemma 3.7.9 with $\eta = \alpha - 2$, $b = 0, \alpha - 1, 2(\alpha - 1)$, $p = 0$ and (3.3) with $\nu > 4 - \alpha$. We established that we can invert the derivation and integration signs in all the K_i's, hence in J'.

A special manipulation to obtain the fourth derivative   Fourth derivative

Before derivating $\phi^{(3)}_{X_2|x}$, we follow the advice stated in [33] (p.48) and integrate by parts the terms containing $\int_{S_2} g_3(ts_1 + rs_2)s_2^3 s_1^{-1}\Gamma(d\boldsymbol{s})$ and $\int_{S_2} g_3(ts_1 + rs_2)s_2^2\Gamma(d\boldsymbol{s})$, namely I_1, I_6 and I_7. This is done in order to guarantee the validity of the representation of the fourth derivative when (3.3) holds for any $\nu > 4 - \alpha$. If we did not do this step first, the obtained fourth derivative would be valid only when (3.3) holds with $\nu > 5 - \alpha$. We obtain

$$\begin{aligned}\phi^{(3)}_{X_2|x}(r) = \frac{-\alpha}{2\pi f_{X_1}(x)}\Bigg[& i\alpha x\Big(I_{11} - I_2 + I_{62} - 2I_{72}\Big) - x^2 I_{12} \\ &+ \alpha^2\Big(I_3 - I_4 - 2I_{71} + I_{61}\Big) + \alpha(\alpha-1)\Big(I_5 - I_{63} + 2I_{73}\Big)\Bigg],\end{aligned} \tag{3.65}$$

where, in addition to I_2, I_3, I_4 and I_5 defined in the Lemma,

$$\begin{aligned}
I_{11} &= \Lambda\Big(\int_{S_2} g_2 s_2^3 s_1^{-2}\Big)\Big(\int_{S_2} g_2 s_1\Big), & I_{12} &= \Lambda\Big(\int_{S_2} g_2 s_2^3 s_1^{-2}\Big),\\
I_{61} &= \Lambda\Big(\int_{S_2} g_2 s_2^3 s_1^{-2}\Big)\Big(\int_{S_2} g_2 s_1\Big)^2, & I_{71} &= \Lambda\Big(\int_{S_2} g_2 s_2^2 s_1^{-1}\Big)\Big(\int_{S_2} g_2 s_1\Big)\Big(\int_{S_2} g_2 s_2\Big),\\
I_{62} &= \Lambda\Big(\int_{S_2} g_2 s_2^3 s_1^{-2}\Big)\Big(\int_{S_2} g_2 s_1\Big), & I_{72} &= \Lambda\Big(\int_{S_2} g_2 s_2^2 s_1^{-1}\Big)\Big(\int_{S_2} g_2 s_2\Big),\\
I_{63} &= \Lambda\Big(\int_{S_2} g_2 s_2^3 s_1^{-2}\Big)\Big(\int_{S_2} g_3 s_1^2\Big), & I_{73} &= \Lambda\Big(\int_{S_2} g_2 s_2^2 s_1^{-1}\Big)\Big(\int_{S_2} g_3 s_2 s_1\Big).
\end{aligned}$$

The fourth derivative is obtained from this representation by techniques similar to those used to get the first and second derivatives.

3.7.9 Proof of Theorem 3.6.3

The second order derivative of the characteristic function of $X_2|X_1 = x$ is given by (3.46) in Lemma 3.7.2. Evaluating it at $r = 0$ yields

$$
\begin{aligned}
&\mathbb{E}\Big[X_2^2\Big|X_1 = x\Big] \\
&\quad = -\phi^{(2)}_{X_2|x}(0) \\
&\quad = \frac{\alpha}{2\pi f_{X_1}(x)} \int_{\mathbb{R}} e^{-itx+ia\sigma_1^\alpha \beta_1 t^{<\alpha>}} e^{-\sigma_1^\alpha |t|^\alpha} \\
&\qquad\qquad \times \Big[ix\sigma_1^\alpha(\kappa_2 t^{<\alpha-1>} - ia\lambda_2|t|^{\alpha-1}) - \alpha\sigma_1^{2\alpha}(\kappa_1 t^{<\alpha-1>} - ia\lambda_1|t|^{\alpha-1})^2 \\
&\qquad\qquad\quad + \alpha\sigma_1^{2\alpha}(\kappa_2 t^{<\alpha-1>} - ia\lambda_2|t|^{\alpha-1})(t^{<\alpha-1>} - ia\beta_1|t|^{\alpha-1})\Big] dt \\
&\quad = \frac{\alpha\sigma_1^\alpha}{2\pi f_{X_1}(x)} \int_{\mathbb{R}} e^{-itx+ia\sigma_1^\alpha \beta_1 t^{<\alpha>}} e^{-\sigma_1^\alpha |t|^\alpha} \\
&\qquad\qquad \times \Big[xa\lambda_2|t|^{\alpha-1} + \alpha\sigma_1^\alpha |t|^{2(\alpha-1)}\Big(\kappa_2 - a^2\beta_1\lambda_2 - \kappa_1^2 + a^2\lambda_1^2\Big) \\
&\qquad\qquad\quad + ix\kappa_2 t^{<\alpha-1>} + i\alpha\sigma_1^\alpha t^{<2(\alpha-1)>}\Big(2a\lambda_1\kappa_1 - a(\lambda_2 + \beta_1\kappa_2)\Big)\Big] dt \\
&\quad = \frac{\alpha\sigma_1^\alpha}{\pi f_{X_1}(x)} \Big[ax\lambda_2 C_1(x) + \kappa_2 x S_1(x) \\
&\qquad\qquad - \alpha\sigma_1^\alpha\Big(\kappa_1^2 - a^2\lambda_1^2 + a^2\beta_1\lambda_2 - \kappa_2\Big)C_2(x) - \alpha\sigma_1^\alpha\Big(a(\lambda_2 + \beta_1\kappa_2) - 2a\lambda_1\kappa_1\Big)S_2(x)\Big],
\end{aligned}
$$

where the κ_i's and λ_i's are given at (3.25). Invoking Lemma 3.7.10 $(\iota\iota\iota)$ yields

$$
\begin{aligned}
\mathbb{E}\Big[X_2^2\Big|X_1 = x\Big] &= \frac{x}{1 + (a\beta_1)^2}\left[(a^2\lambda_2\beta_1 + \kappa_2)x + a(\lambda_2 - \kappa_2\beta_1)\frac{1 - xH(x)}{\pi f_{X_1}(x)}\right] \\
&\quad - \frac{\alpha^2\sigma_1^{2\alpha}}{\pi f_{X_1}(x)}\mathcal{H}\Big(2(\alpha - 1), \boldsymbol{\theta}_1; x\Big) \\
&= \kappa_2 x^2 + \frac{ax(\lambda_2 - \beta_1\kappa_2)}{1 + (a\beta_1)^2}\left[a\beta_1 x + \frac{1 - xH(x)}{\pi f_{X_1}(x)}\right] - \frac{\alpha^2\sigma_1^{2\alpha}}{\pi f_{X_1}(x)}\mathcal{H}\Big(2(\alpha - 1), \boldsymbol{\theta}_1; x\Big),
\end{aligned}
$$

where $\mathcal{H}$ is given in (3.41) with

$$
\theta_{11} = \kappa_1^2 - a^2\lambda_1^2 + a^2\beta_1\lambda_2 - \kappa_2, \qquad\qquad \theta_{12} = a(\lambda_2 + \beta_1\kappa_2) - 2a\lambda_1\kappa_1.
$$

3.7.10 Proof of Theorem 3.6.5

The third order derivative of the characteristic function of $X_2|X_1 = x$ is given by (3.47) in Lemma 3.7.2. It can be shown that the I's evaluated at $r = 0$ write

$$
\begin{aligned}
I_1 &= 2\sigma_1^{\alpha}\mathcal{H}\Big(\alpha - 2, \boldsymbol{\theta}_1^I; x\Big), & \boldsymbol{\theta}_1^I &= \Big(\kappa_3, -a\lambda_3\Big),\\
I_2 &= 2\sigma_1^{2\alpha}\mathcal{H}\Big(2(\alpha - 1), \boldsymbol{\theta}_2^I; x\Big), & \boldsymbol{\theta}_2^I &= \Big(L, -aK\Big),\\
iI_3 &= 2\sigma_1^{3\alpha}\mathcal{H}\Big(3(\alpha - 1), \boldsymbol{\theta}_3^I; x\Big), & \boldsymbol{\theta}_3^I &= \Big(a\lambda_1(3\kappa_1^2 - a^2\lambda_1^2), \kappa_1^3 - 3a^2\kappa_1\lambda_1^2\Big),\\
iI_4 &= 2\sigma_1^{3\alpha}\mathcal{H}\Big(3(\alpha - 1), \boldsymbol{\theta}_4^I; x\Big), & \boldsymbol{\theta}_4^I &= \Big(a\Big(K + \beta_1 L\Big), L - a^2\beta_1 K\Big),\\
iI_5 = iI_7 &= 2\sigma_1^{2\alpha}\mathcal{H}\Big(2\alpha - 3, \boldsymbol{\theta}_5^I; x\Big), & \boldsymbol{\theta}_5^I &= \Big(aK, L\Big),\\
iI_6 &= 2\sigma_1^{2\alpha}\mathcal{H}\Big(2\alpha - 3, \boldsymbol{\theta}_6^I; x\Big), & \boldsymbol{\theta}_6^I &= \Big(a(\lambda_3 + \beta_1\kappa_3), \kappa_3 - a^2\beta_1\lambda_3\Big),
\end{aligned}
$$

with $K = \kappa_1\lambda_2 + \lambda_1\kappa_2$ and $L = \kappa_1\kappa_2 - a^2\lambda_1\lambda_2$. Hence,

$$
\mathbb{E}\Big[X_2^3\Big|X_1 = x\Big] = -i\phi_{X_2|x}^{(3)}(0) = \frac{\alpha}{\pi f_{X_1}(x)}\Bigg[-x\Big((\alpha - 1)K_1 - \alpha K_2\Big) + \alpha^2 K_3 + \alpha(\alpha - 1)K_4\Bigg],
$$

with

$$
\begin{aligned}
K_1 &= \sigma_1^{\alpha}\mathcal{H}\Big(\alpha - 2, \boldsymbol{\theta}_1^K; x\Big), &&\text{with} \quad \boldsymbol{\theta}_1^K = \boldsymbol{\theta}_1^I,\\
K_2 &= \sigma_1^{2\alpha}\mathcal{H}\Big(2(\alpha - 1), \boldsymbol{\theta}_2^K; x\Big), &&\text{with} \quad \boldsymbol{\theta}_2^K = \boldsymbol{\theta}_2^I,\\
K_3 &= \sigma_1^{3\alpha}\mathcal{H}\Big(3(\alpha - 1), \boldsymbol{\theta}_3^K; x\Big), &&\text{with} \quad \boldsymbol{\theta}_3^K = \boldsymbol{\theta}_3^I - \boldsymbol{\theta}_4^I\\
K_4 &= \sigma_1^{2\alpha}\mathcal{H}\Big(2\alpha - 3, \boldsymbol{\theta}_4^K; x\Big), &&\text{with} \quad \boldsymbol{\theta}_4^K = \boldsymbol{\theta}_6^I - \boldsymbol{\theta}_5^I.
\end{aligned}
$$

Invoking Lemma 3.7.10 ($\iota\iota$) for $n = 1, 2$ and regrouping the terms, we get

$$
\begin{aligned}
\mathbb{E}\Big[X_2^3\Big|X_1 = x\Big] = &\frac{\alpha x^2\sigma_1^{\alpha}}{\pi f_{X_1}(x)}\Big(\theta_{12}^K C_1(x) - \theta_{11}^K S_1(x)\Big)\\
&+ \frac{\alpha}{\pi f_{X_1}(x)}\Bigg[\ \frac{\alpha x\sigma_1^{2\alpha}}{2}C_2(x)\Big(-2\Big(\theta_{11}^K + a\beta_1\theta_{12}^K\Big) + 2\theta_{21}^K - \theta_{42}^K\Big)\\
&\qquad + \frac{\alpha x\sigma_1^{2\alpha}}{2}S_2(x)\Big(-2\Big(\theta_{12}^K - a\beta_1\theta_{11}^K\Big) + 2\theta_{22}^K + \theta_{41}^K\Big)\\
&\qquad + \frac{\alpha^2\sigma_1^{3\alpha}}{2}C_3(x)\Big(2\theta_{31}^K + \theta_{41}^K + a\beta_1\theta_{42}^K\Big)\\
&\qquad + \frac{\alpha^2\sigma_1^{3\alpha}}{2}S_3(x)\Big(2\theta_{32}^K + \theta_{42}^K - a\beta_1\theta_{41}^K\Big)\Bigg].
\end{aligned}
$$

Using Lemma 3.7.10 ($\iota\iota\iota$) yields the conclusion with $\boldsymbol{\theta}_2 = (\theta_{21}, \theta_{22})$, $\boldsymbol{\theta}_3 = (\theta_{31}, \theta_{32})$ such that

$$
\begin{aligned}
\theta_{21} &= 3(L + a^2\beta_1\lambda_3 - \kappa_3), \\
\theta_{22} &= 3a(\lambda_3 + \beta_1\kappa_3 - K), \\
\theta_{31} &= a\Big(\lambda_3(1 - a^2\beta_1^2) + 2\beta_1\kappa_3 + 2\lambda_1(3\kappa_1^2 - a^2\lambda_1^2) - 3(K + \beta_1 L)\Big), \\
\theta_{32} &= \kappa_3(1 - a^2\beta_1^2) - 2a^2\beta_1\lambda_3 + 2(\kappa_1^3 - 3a^2\kappa_1\lambda_1^2) + 3(a^2\beta_1 K - L),
\end{aligned}
$$

with $K = \kappa_1\lambda_2 + \kappa_2\lambda_1$, $L = \kappa_1\kappa_2 - a^2\lambda_1\lambda_2$.

3.7.11 Proof of Theorem 3.6.6

The conditional moments are obtained by evaluating the derivatives of the conditional characteristic function at $r = 0$. We provide here the proof for the fourth order, which yields the expressions of the vectors $\boldsymbol{\theta}_4$, $\boldsymbol{\theta}_5$ and $\boldsymbol{\theta}_6$ appearing in Theorems 3.2.1 and 3.6.6. The fourth order derivative of the characteristic function of $X_2|X_1 = x$ is given by (3.48) in Lemma 3.7.2. It can be shown that the J's evaluated at $r = 0$ write

$$
\begin{aligned}
iJ_1 &= 2\sigma_1^{3\alpha}\mathcal{H}\Big(3(\alpha-1), \boldsymbol{\theta}_1^J; x\Big), & J_{11} = J_{13} &= 2\sigma_1^{3\alpha}\mathcal{H}\Big(3\alpha - 4, \boldsymbol{\theta}_{11}^J; x\Big), \\
iJ_2 &= 2\sigma_1^{3\alpha}\mathcal{H}\Big(3(\alpha-1), \boldsymbol{\theta}_2^J; x\Big), & J_{14} &= 2\sigma_1^{2\alpha}\mathcal{H}\Big(2\alpha - 4, \boldsymbol{\theta}_{14}^J; x\Big), \\
iJ_3 &= 2\sigma_1^{2\alpha}\mathcal{H}\Big(2\alpha - 3, \boldsymbol{\theta}_3^J; x\Big), & J_{15} &= 2\sigma_1^{2\alpha}\mathcal{H}\Big(2\alpha - 4, \boldsymbol{\theta}_{15}^J; x\Big), \\
iJ_4 = iJ_5 &= 2\sigma_1^{2\alpha}\mathcal{H}\Big(2\alpha - 3, \boldsymbol{\theta}_4^J; x\Big), & J_{16} &= 2\sigma_1^{2\alpha}\mathcal{H}\Big(2\alpha - 4, \boldsymbol{\theta}_{16}^J; x\Big), \\
J_6 &= 2\sigma_1^{2\alpha}\mathcal{H}\Big(2(\alpha-1), \boldsymbol{\theta}_6^J; x\Big), & J_{17} &= 2\sigma_1^{4\alpha}\mathcal{H}\Big(4(\alpha-1), \boldsymbol{\theta}_{17}^J; x\Big), \\
J_7 &= 2\sigma_1^{\alpha}\mathcal{H}\Big(\alpha - 2, \boldsymbol{\theta}_7^J; x\Big), & J_{18} &= 2\sigma_1^{4\alpha}\mathcal{H}\Big(4(\alpha-1), \boldsymbol{\theta}_{18}^J; x\Big), \\
J_8 = J_9 = J_{12} &= 2\sigma_1^{3\alpha}\mathcal{H}\Big(3\alpha - 4, \boldsymbol{\theta}_8^J; x\Big), & J_{19} &= 2\sigma_1^{4\alpha}\mathcal{H}\Big(4(\alpha-1), \boldsymbol{\theta}_{19}^J; x\Big), \\
J_{10} &= 2\sigma_1^{3\alpha}\mathcal{H}\Big(3\alpha - 4, \boldsymbol{\theta}_{10}^J; x\Big), & &
\end{aligned}
$$

where $\boldsymbol{\theta}_i^J = (\theta_{i1}^J, \theta_{i2}^J)$, for $i = 1, \dots, 19$,

$$
\begin{aligned}
\theta_{11}^J &= a\Big(\lambda_2(\kappa_1^2 - a^2\lambda_1^2) + 2\kappa_1\kappa_2\lambda_1\Big), & \theta_{12}^J &= \kappa_2(\kappa_1^2 - a^2\lambda_1^2) - 2a^2\kappa_1\lambda_1\lambda_2,\\
\theta_{21}^J &= a\Big(K + \beta_1 L\Big), & \theta_{22}^J &= L - a^2\beta_1 K,\\
\theta_{31}^J &= a\Big(\beta_1\kappa_4 + \lambda_4\Big), & \theta_{32}^J &= \kappa_4 - a^2\beta_1\lambda_4,\\
\theta_{41}^J &= aK, & \theta_{42}^J &= L,\\
\theta_{61}^J &= L, & \theta_{62}^J &= -aK,\\
\theta_{71}^J &= \kappa_4, & \theta_{72}^J &= -a\lambda_4,\\
\theta_{81}^J &= L - a^2\beta_1 K, & \theta_{82}^J &= -a\Big(K + \beta_1 L\Big),\\
\theta_{101}^J &= \kappa_4(1 - a^2\beta_1^2) - 2a^2\beta_1\lambda_4, & \theta_{102}^J &= -a\Big(\lambda_4(1 - a^2\beta_1^2) + 2\beta_1\kappa_4\Big),\\
\theta_{111}^J &= \theta_{12}^J, & \theta_{112}^J &= -\theta_{11}^J,\\
\theta_{141}^J &= L, & \theta_{142}^J &= -aK,\\
\theta_{151}^J &= \kappa_2^2 - a^2\lambda_2^2, & \theta_{152}^J &= -2a\kappa_2\lambda_2,\\
\theta_{161}^J &= \kappa_4 - a^2\beta_1\lambda_4, & \theta_{162}^J &= -a\Big(\lambda_4 + \beta_1\kappa_4\Big),\\
\theta_{171}^J &= \theta_{12}^J - a\beta_1\theta_{11}^J, & \theta_{172}^J &= -\theta_{11}^J + a\theta_{12}^J,\\
\theta_{181}^J &= \kappa_1^4 - 6a^2\kappa_1^2\lambda_1^2 + a^4\lambda_1^4, & \theta_{182}^J &= -4a\kappa_1\lambda_1(\kappa_1^2 - a^2\lambda_1^2),\\
\theta_{191}^J &= L(1 - a^2\beta_1^2) - 2a^2\beta_1 K, & \theta_{192} &= -a\Big(K(1 - a^2\beta_1^2) + 2\beta_1 L\Big),
\end{aligned}
$$

and $K = \kappa_1\lambda_3 + \lambda_1\kappa_3$, $L = \kappa_1\kappa_3 - a^2\lambda_1\lambda_3$. Hence,

$$
\begin{aligned}
\mathbb{E}\Big[X_2^4 \Big| X_1 = x\Big] &= \phi_{X_2|x}^{(4)}(0)\\
&= \frac{-\alpha}{\pi f_{X_1}(x)}\Bigg[\alpha x\Big(\alpha K_1 + (\alpha-1)K_2\Big) + \alpha x^2 K_6 - (\alpha-1)K_7 + \alpha^2(\alpha-1)K_3 + \alpha(\alpha-1)^2 K_4 + \alpha^3\Bigg],
\end{aligned}
$$

where

$$
\begin{aligned}
K_1 &= \sigma_1^{3\alpha}\mathcal{H}\Big(3(\alpha-1), \boldsymbol{\theta}_1^K; x\Big), && \text{with} \quad \boldsymbol{\theta}_1^K = 3\boldsymbol{\theta}_1^J - 2\boldsymbol{\theta}_2^J,\\
K_2 &= \sigma_1^{2\alpha}\mathcal{H}\Big(2\alpha - 3, \boldsymbol{\theta}_2^K; x\Big), && \text{with} \quad \boldsymbol{\theta}_2^K = 2(\boldsymbol{\theta}_3^J - \boldsymbol{\theta}_4^J),\\
K_3 &= \sigma_1^{3\alpha}\mathcal{H}\Big(3\alpha - 4, \boldsymbol{\theta}_3^K; x\Big), && \text{with} \quad \boldsymbol{\theta}_3^K = \boldsymbol{\theta}_{10}^J - 3\boldsymbol{\theta}_{11}^J - \boldsymbol{\theta}_8^J,\\
K_4 &= \sigma_1^{2\alpha}\mathcal{H}\Big(2\alpha - 4, \boldsymbol{\theta}_4^K; x\Big), && \text{with} \quad \boldsymbol{\theta}_4^K = 4\boldsymbol{\theta}_{14}^J - 3\boldsymbol{\theta}_{15}^J - \boldsymbol{\theta}_{16}^J,\\
K_5 &= \sigma_1^{4\alpha}\mathcal{H}\Big(4(\alpha-1), \boldsymbol{\theta}_5^K; x\Big), && \text{with} \quad \boldsymbol{\theta}_5^K = 3\boldsymbol{\theta}_{17}^J - \boldsymbol{\theta}_{18}^J - \boldsymbol{\theta}_{19}^J,\\
K_6 &= \sigma_1^{2\alpha}\mathcal{H}\Big(2(\alpha-1), \boldsymbol{\theta}_6^K; x\Big), && \text{with} \quad \boldsymbol{\theta}_6^K = \boldsymbol{\theta}_6^J,\\
K_7 &= \sigma_1^{\alpha}\mathcal{H}\Big(\alpha - 2, \boldsymbol{\theta}_7^K; x\Big), && \text{with} \quad \boldsymbol{\theta}_7^K = \boldsymbol{\theta}_7^J.
\end{aligned}
$$

Invoking Lemmas 3.7.10 ($\iota\iota$) for $n = 1, 2, 3$ and 3.7.11, we get

$$
\begin{aligned}
\mathbb{E}\Big[X_2^4\Big|X_1 = x\Big] = \frac{-\alpha}{\pi f_{X_1}(x)}\Bigg[& x^3\sigma_1^\alpha\Big(\theta_{72}^K C_1(x) - \theta_{71}^K S_1(x)\Big) \\
&+ \frac{\alpha x^2\sigma_1^{2\alpha}}{2}C_2(x)\Bigg(- \theta_{22}^K + 2\theta_{61}^K - 2\Big(\theta_{71}^K + a\beta_1\theta_{72}^K\Big) - \frac{\alpha-1}{2\alpha-3}\theta_{41}^K\Bigg) \\
&+ \frac{\alpha x^2\sigma_1^{2\alpha}}{2}S_2(x)\Bigg(\theta_{21}^K + 2\theta_{62}^K - 2\Big(\theta_{72}^K - a\beta_1\theta_{71}^K\Big) - \frac{\alpha-1}{2\alpha-3}\theta_{42}^K\Bigg)\Bigg] \\
&+ \frac{\alpha^2 x\sigma_1^{3\alpha}}{6}C_3(x)\Bigg(6\theta_{11}^K + 3\Big(\theta_{21}^K + a\beta_1\theta_{22}^K\Big) - 2\theta_{32}^K + 5\frac{\alpha-1}{2\alpha-3}\Big(a\beta_1\theta_{41}^K - \theta_{42}^K\Big)\Bigg) \\
&+ \frac{\alpha^2 x\sigma_1^{3\alpha}}{6}S_3(x)\Bigg(6\theta_{21}^K + 3\Big(\theta_{22}^K - a\beta_1\theta_{21}^K\Big) + 2\theta_{31}^K + 5\frac{\alpha-1}{2\alpha-3}\Big(\theta_{41}^K + a\beta_1\theta_{42}^K\Big)\Bigg) \\
&+ \frac{\alpha^3\sigma_1^{4\alpha}}{3}C_4(x)\Bigg(\theta_{31}^K + a\beta_1\theta_{32}^K + \frac{\alpha-1}{2\alpha-3}\Big(\theta_{41}^K(1-a^2\beta_1^2) + 2a\beta_1\theta_{42}^K\Big) + 3\theta_{51}^K\Bigg) \\
&+ \frac{\alpha^3\sigma_1^{4\alpha}}{3}S_4(x)\Bigg(\theta_{32}^K - a\beta_1\theta_{31}^K + \frac{\alpha-1}{2\alpha-3}\Big(\theta_{42}^K(1-a^2\beta_1^2) - 2a\beta_1\theta_{41}^K\Big) + 3\theta_{52}^K\Bigg)\Bigg].
\end{aligned}
$$

Using Lemma 3.7.10 ($\iota\iota\iota$) yields the conclusion. The coefficients $\boldsymbol{\theta}$'s in the expression of Proposition 3.6.6, are deduced from the $\boldsymbol{\theta}^K$'s and $\boldsymbol{\theta}^J$'s as follows:

$$\theta_{41} = -\theta_{22}^K + 2\theta_{61}^K - 2\Big(\theta_{71}^K + a\beta_1\theta_{72}^K\Big) - \frac{\alpha-1}{2\alpha-3}\theta_{41}^K, \tag{3.66}$$

$$\theta_{42} = \theta_{21}^K + 2\theta_{62}^K - 2\Big(\theta_{72}^K - a\beta_1\theta_{71}^K\Big) - \frac{\alpha-1}{2\alpha-3}\theta_{42}^K, \tag{3.67}$$

$$\theta_{51} = 6\theta_{11}^K + 3\Big(\theta_{21}^K + a\beta_1\theta_{22}^K\Big) - 2\theta_{32}^K + 5\frac{\alpha-1}{2\alpha-3}\Big(a\beta_1\theta_{41}^K - \theta_{42}^K\Big), \tag{3.68}$$

$$\theta_{52} = 6\theta_{21}^K + 3\Big(\theta_{22}^K - a\beta_1\theta_{21}^K\Big) + 2\theta_{31}^K + 5\frac{\alpha-1}{2\alpha-3}\Big(\theta_{41}^K + a\beta_1\theta_{42}^K\Big), \tag{3.69}$$

$$\theta_{61} = \theta_{31}^K + a\beta_1\theta_{32}^K + \frac{\alpha-1}{2\alpha-3}\Big(\theta_{41}^K(1-a^2\beta_1^2) + 2a\beta_1\theta_{42}^K\Big) + 3\theta_{51}^K, \tag{3.70}$$

$$\theta_{62} = \theta_{32}^K - a\beta_1\theta_{31}^K + \frac{\alpha-1}{2\alpha-3}\Big(\theta_{42}^K(1-a^2\beta_1^2) - 2a\beta_1\theta_{41}^K\Big) + 3\theta_{52}^K. \tag{3.71}$$

3.7.12 Vectors $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$ of Theorems 3.2.1 and 3.6.5

We provide here the expressions of $\boldsymbol{\theta}_2 = (\theta_{21}, \theta_{22})$, $\boldsymbol{\theta}_3 = (\theta_{31}, \theta_{32})$, which intervene in the form of the third conditional moments:

$$\theta_{21} = 3(L + a^2\beta_1\lambda_3 - \kappa_3), \tag{3.72}$$

$$\theta_{22} = 3a(\lambda_3 + \beta_1\kappa_3 - K), \tag{3.73}$$

$$\theta_{31} = a\Big(\lambda_3(1-a^2\beta_1^2) + 2\beta_1\kappa_3 + 2\lambda_1(3\kappa_1^2 - a^2\lambda_1^2) - 3(K + \beta_1 L)\Big), \tag{3.74}$$

$$\theta_{32} = \kappa_3(1-a^2\beta_1^2) - 2a^2\beta_1\lambda_3 + 2(\kappa_1^3 - 3a^2\kappa_1\lambda_1^2) + 3(a^2\beta_1 K - L), \tag{3.75}$$

with $K = \kappa_1\lambda_2 + \kappa_2\lambda_1$ and $L = \kappa_1\kappa_2 - a^2\lambda_1\lambda_2$.

3.7.13 Proof of Proposition 3.6.1 in the case $\alpha \neq 1$

First assume that $|\beta_1| \neq 1$. We will focus on the case $x \to +\infty$. The case $x \to -\infty$ can be obtained by considering the vector (X_1, X_2), whose parameter are $\beta_1^* = -\beta_1$, $\kappa_1^* = -\kappa_1$ and $\lambda_1^* = \lambda_1$ and noticing that $\mathbb{E}\Big[X_2^p\Big|X_1 = x\Big] = \mathbb{E}\Big[X_2^p\Big| - X_1 = -x\Big]$. For $p = 1$, the result is already known (see [64]). For $p = 2, 3, 4$, we have from the proofs of Propositions 3.29, 3.30 and 3.31, that

$$\mathbb{E}\Big[X_2^p\Big|X_1 = x\Big] = \frac{\alpha\sigma_1^\alpha}{\pi f_{X_1}(x)}\Big[x^{p-1}\mathcal{H}\Big(\alpha - 1, (a\lambda_p, \kappa_p); x\Big) + \sum_{i=2}^{p} b_{i,p} x^{p-i}\mathcal{H}\Big(i(\alpha - 1), \boldsymbol{\nu}_i; x\Big)\Big],$$

for some coefficients b's. From the proof of Corollary 3.2 in [64], we deduce the following limit:

$$x^\alpha\mathcal{H}\Big(\alpha - 1, (a\lambda_p, \kappa_p); x\Big) \underset{x\to+\infty}{\longrightarrow} \Big(\kappa_p + \lambda_p\Big)\sin\Big(\frac{\pi\alpha}{2}\Big)\Gamma(\alpha).$$

We also have

$$x^{\alpha+1} f_{X_1}(x) \underset{x\to+\infty}{\longrightarrow} \frac{1}{\pi}\sigma_1^\alpha(1 + \beta_1)\sin\Big(\frac{\pi\alpha}{2}\Big)\Gamma(1 + \alpha). \tag{3.76}$$

Hence,

$$x^{-p}\frac{\alpha\sigma_1^\alpha x^{p-1}}{\pi f_{X_1}(x)}\mathcal{H}\Big(\alpha - 1, (a\lambda_p, \kappa_p); x\Big) \longrightarrow \frac{\kappa_p + \lambda_p}{1 + \beta_1},$$

as $x \to +\infty$. It remains to be shown that $\dfrac{\sum_{i=2}^p b_{i,p}x^{p-i}\mathcal{H}\Big(i(\alpha-1), \boldsymbol{\nu}_i; x\Big)}{x^{p-1}\mathcal{H}\Big(\alpha - 1, (a\lambda_p, \kappa_p); x\Big)} \underset{x\to+\infty}{\longrightarrow} 0$. By Theorem 127 in [118], for $i = 2, 3, 4$,

$$\mathcal{H}\Big(i(\alpha - 1), \boldsymbol{\nu}_i; x\Big) \underset{x\to+\infty}{=} O\Big(x^{-i(\alpha-1)-1}\Big).$$

Hence,

$$\left|\frac{x^{p-i}\mathcal{H}\Big(i(\alpha - 1), \boldsymbol{\nu}_i; x\Big)}{x^{p-1}\mathcal{H}\Big(\alpha - 1, (a\lambda_p, \kappa_p); x\Big)}\right| \underset{x\to+\infty}{=} O\Big(x^{\alpha(1-i)}\Big) \longrightarrow 0.$$

Now assume that $|\beta_1| = 1$. For instance if $\beta_1 = 1$, the distribution of X_1 is *totally skewed to the right*. On the one hand, we have $\lambda_p = \beta_1\kappa_p$. On the other hand, the right tail of f_{X_1} still decays as (3.76), yielding the conclusion. □

The following elementary Lemmas, stated without proof, are used to establish Theorems 3.6.3-3.6.6.

Lemma 3.7.3 *For $x, y \in \mathbb{R}$,*

$$|e^{-x} - e^{-y}| \leq e^{-\min(x,y)}|x - y|, \tag{3.77}$$

$$|e^{-x} - e^{-y}| \leq e^{-y}e^{|x-y|}|x - y|. \tag{3.78}$$

Lemma 3.7.4 *For $\alpha > 1$ and $x, y \in \mathbb{R}$,*

$$\max\Big(2^{1-\alpha}|x|^\alpha - |y|^\alpha, 2^{1-\alpha}|y|^\alpha - |x|^\alpha\Big) \leq |x + y|^\alpha \leq 2^{\alpha-1}\Big(|x|^\alpha + |y|^\alpha\Big).$$

Lemma 3.7.5 *For $z \in \mathbb{R}$ and $0 < b \leq 1$,*

$$(\iota) \qquad \Big| |1+z|^b - 1 \Big| \leq |z|,$$
$$(\iota\iota) \qquad \Big| |1+z|^{<b>} - 1 \Big| \leq 2|z|.$$

Lemma 3.7.6 (Lemma 3.3, Cioszek-Georges and Taqqu (1998)) *For $\alpha > 1$ and $t, r \in \mathbb{R}$,*

$$\exp\Big\{ - \int_{S_2} |ts_1 + rs_2|^\alpha \Gamma(d\boldsymbol{s}) \Big\} \leq \exp\{|r|^\alpha \sigma_2^\alpha\} \exp\{-2^{1-\alpha}\sigma_1^\alpha |t|^\alpha\}.$$

Lemma 3.7.7 (Lemma 3.1, Cioszek-Georges and Taqqu (1998)) *The following inequality holds for $c > 0$, $0 < \alpha < 2$, $-1 < \eta < 0$ and $-1 - \eta < b$:*

$$\int_{\mathbb{R}} \exp(-c|t|^\alpha) \Big| |t+z|^\eta - |t|^\eta \Big| |t|^b dt \leq const. \, |z|^p$$

with

$$0 \leq p < b + \eta + 1 \quad for \quad -1 - \eta < b < 0,$$

and

$$0 \leq p < \eta + 1 \quad or \quad b \leq p < b + \eta + \eta + 1, p \leq 1 \quad for \quad 0 \leq b.$$

const. depends only on c, α, η, b and p.

Lemma 3.7.8 (Corollary 3.1, Cioszek-Georges and Taqqu (1998)) *The following inequality holds for $c > 0$, $0 < \alpha < 2$, $-1/2 < \eta < 0$ and $0 \leq p < 2\eta + 1$:*

$$\int_{\mathbb{R}} \exp(-c|t|^\alpha) \Big| |t+z_1|^\eta |t+z_3|^\eta - |t+z_2|^\eta |t+z_4|^\eta \Big| dt \leq const. \, (|z_1 - z_2|^p + |z_3 - z_4|^p),$$

where const depends only on c, α, η and p.

Lemma 3.7.9 (Lemma 3.12, Cioszek-Georges and Taqqu (1998)) *The following inequality holds for $c > 0$, $0 < \alpha < 2$, $-1 < \eta < 0$, $b \geq 0$ and $0 \leq p < \eta + 1$:*

$$\int_{\mathbb{R}} \exp(-c|t|^\alpha) \Big| |t+z_1|^\eta - |t+z_2|^\eta \Big| |t|^b dt \leq const. \, |z_1 - z_2|^p,$$

where const depends only on c, α, η, b and p.

Lemma 3.7.10 *Let $\alpha \in (1, 2)$, $b > 0$, $c \in \mathbb{R}$. Define for $n \geq 1$ and $x \in \mathbb{R}$*

$$C_n(x) = \int_0^{+\infty} e^{-bt^\alpha} t^{n(\alpha-1)} \cos(tx - ct^\alpha) dt, \qquad F_n(x) = \int_0^{+\infty} e^{-bt^\alpha} t^{n(\alpha-1)-1} \cos(tx - ct^\alpha) dt,$$
$$S_n(x) = \int_0^{+\infty} e^{-bt^\alpha} t^{n(\alpha-1)} \sin(tx - ct^\alpha) dt, \qquad G_n(x) = \int_0^{+\infty} e^{-bt^\alpha} t^{n(\alpha-1)-1} \sin(tx - ct^\alpha) dt.$$

ι) Then the following hold for any $n \geq 1$ and $x \in \mathbb{R}$

$$F_n(x) = \frac{1}{n(\alpha-1)} \Big[\alpha \Big(bC_{n+1}(x) - cS_{n+1}(x) \Big) + xS_n(x) \Big],$$
$$G_n(x) = \frac{1}{n(\alpha-1)} \Big[\alpha \Big(cC_{n+1}(x) + bS_{n+1}(x) \Big) - xC_n(x) \Big].$$

ιι) For any $n \geq 1$, $\theta_1, \theta_2 \in \mathbb{R}$ and $x \in \mathbb{R}$:

$$\theta_1 F_n(x) + \theta_2 G_n(x) = \frac{\alpha}{n(\alpha-1)}\Big[C_{n+1}(x)\Big(b\theta_1 + c\theta_2\Big) + S_{n+1}(x)\Big(b\theta_2 - c\theta_1\Big)\Big] + \frac{x}{n(\alpha-1)}\Big[-\theta_2 C_n(x) + \theta_1 S_n(x)\Big].$$

ιιι) We have for $x \in \mathbb{R}$, $b = \sigma_1^\alpha$ and $c = a\beta_1\sigma_1^\alpha$:

$$C_1(x) = \frac{1}{\alpha\sigma_1^\alpha(1+(a\beta_1)^2)}\Big[a\beta_1 x\pi f_{X_1}(x) + 1 - xH(x)\Big],$$
$$S_1(x) = \frac{1}{\alpha\sigma_1^\alpha(1+(a\beta_1)^2)}\Big[x\pi f_{X_1}(x) - a\beta_1(1 - xH(x))\Big].$$

Lemma 3.7.11 *Let $\alpha \in (3/2, 2)$, $b > 0$, $c \in \mathbb{R}$. Define for $x \in \mathbb{R}$*

$$h_c(x) = \int_0^{+\infty} e^{-bt^\alpha} t^{2\alpha-4}\cos(tx - ct^\alpha)dt, \qquad h_s(x) = \int_0^{+\infty} e^{-bt^\alpha} t^{2\alpha-4}\sin(tx - ct^\alpha)dt.$$

Then for any $\theta_1, \theta_2 \in \mathbb{R}$ and $x \in \mathbb{R}$,

$$\begin{aligned}\theta_1 h_c(x) + \theta_2 h_s(x) = &\frac{\alpha^2}{3(2\alpha-3)(\alpha-1)}\Big[C_4(x)\Big(\theta_1(b^2-c^2) + 2bc\theta_2\Big) + S_4(x)\Big(\theta_2(b^2-c^2) - 2bc\theta_1\Big)\Big]\\ &+ \frac{5\alpha x}{6(2\alpha-3)(\alpha-1)}\Big[C_3(x)\Big(c\theta_1 - b\theta_2\Big) + S_3(x)\Big(b\theta_1 + c\theta_2\Big)\Big]\\ &- \frac{x^2}{2(2\alpha-3)(\alpha-1)}\Big[\theta_1 C_2(x) + \theta_2 S_2(x)\Big].\end{aligned}$$

3.7.14 Proof Theorem 3.6.4

Let $\boldsymbol{X} = (X_1, X_2)$ be an α-stable vector with $\alpha = 1$ and spectral representation $(\Gamma, \mathbf{0})$. Its characteristic function, denoted $\varphi_{\boldsymbol{X}}(t,r)$ for any $(t,r) \in \mathbb{R}^2$, reads

$$\varphi_{\boldsymbol{X}}(t,r) = \exp\left\{-\int_{S_2} |ts_1 + rs_2| + ia(ts_1 + rs_2)\ln|ts_1 + rs_2|\Gamma(ds)\right\}, \tag{3.79}$$

with $a = 2/\pi$. The conditional characteristic function of X_2 given $X_1 = x$, denoted $\phi_{X_2|x}(r)$ for $r \in \mathbb{R}$, is still given by (3.40).

Lemma 3.7.12 *Let (X_1, X_2) be an α-stable random vector with $\alpha = 1$ and spectral representation $(\Gamma, \mathbf{0})$. If (3.3) holds with $\nu > 0$, the first derivative of $\phi_{X_2|x}$ is given by*

$$\phi^{(1)}_{X_2|x}(r) = \frac{-1}{2\pi f_{X_1}(x)}\Big(A_1 + iaA_2\Big),$$

with

$$A_1 = \int_{\mathbb{R}} e^{-itx}\varphi_{\boldsymbol{X}}(t,r)\Big(\int_{S_2} s_2(ts_1 + rs_2)^{<0>}\Gamma(d\boldsymbol{s})\Big)dt, \tag{3.80}$$

$$A_2 = \int_{\mathbb{R}} e^{-itx}\varphi_{\boldsymbol{X}}(t,r)\Big(\int_{S_2} s_2(1 + \ln|ts_1 + rs_2|)\Gamma(d\boldsymbol{s})\Big)dt \tag{3.81}$$

If (3.3) *holds with* $\nu > 1$, *the second derivative of* $\phi_{X_2|x}$ *is given by*

$$\phi^{(2)}_{X_2|x}(r) = \frac{-1}{2\pi f_{X_1}(x)}\Big(-B_1 + ixB_2 + B_3\Big), \tag{3.82}$$

where,

$$B_1 = \int_{\mathbb{R}} e^{-itx}\varphi_{\boldsymbol{X}}(t,r)\bigg(\int_{S_2} s_2(ts_1+rs_2)^{<0>} + ias_2(1+\ln|ts_1+rs_2|\Gamma(d\boldsymbol{s})\bigg)^2 dt,$$
$$B_2 = \int_{\mathbb{R}} e^{-itx}\varphi_{\boldsymbol{X}}(t,r)\bigg(\int_{S_2} \Big((ts_1+rs_2)^{<0>} + ia(1+\ln|ts_1+rs_2|\Big)s_2^2 s_1^{-1}\Gamma(d\boldsymbol{s})\bigg)dt,$$
$$B_3 = \int_{\mathbb{R}} e^{-itx}\varphi_{\boldsymbol{X}}(t,r)\bigg(\int_{S_2} s_1(ts_1+rs_2)^{<0>} + ias_1(1+\ln|ts_1+rs_2|\Gamma(d\boldsymbol{s})\bigg)$$
$$\times\bigg(\int_{S_2} \Big((ts_1+rs_2)^{<0>} + ia(1+\ln|ts_1+rs_2|\Big)s_2^2 s_1^{-1}\Gamma(d\boldsymbol{s})\bigg)dt.$$

Justifying inversion of integral and derivative signs First derivative

The terms depending on r in the right-hand side of (3.79) are of the form (omitting the factor $1/2\pi f_{X_1}(x)$)

$$\int_{\mathbb{R}} e^{-\int_{S_2}|ts_1+rs_2|\Gamma(d\boldsymbol{s})}\mathrm{trig}\bigg(-tx - a\int_{S_2}(ts_1+rs_2)\ln|ts_1+rs_2|\Gamma(d\boldsymbol{s})\bigg)dt.$$

Consider for instance the term obtained by replacing trig by the cosine function, denoted I_1.

$$\begin{aligned}I_1'(r) &= \lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}\Big[e^{-\int_{S_2}|ts_1+(r+h)s_2|\Gamma(d\boldsymbol{s})} - e^{-\int_{S_2}|ts_1+rs_2|\Gamma(d\boldsymbol{s})}\Big]\\ &\qquad\times\cos\bigg(tx + a\int_{S_2}(ts_1+(r+h)s_2)\ln|ts_1+(r+h)s_2|\Gamma(d\boldsymbol{s})\bigg)dt\\ &\quad+\lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}} e^{-\int_{S_2}|ts_1+rs_2|\Gamma(d\boldsymbol{s})}\bigg[\cos\bigg(tx + a\int_{S_2}(ts_1+(r+h)s_2)\ln|ts_1+(r+h)s_2|\Gamma(d\boldsymbol{s})\bigg)\\ &\qquad-\cos\bigg(tx + a\int_{S_2}(ts_1+rs_2)\ln|ts_1+rs_2|\Gamma(d\boldsymbol{s})\bigg)\bigg]dt\\ &\quad:= I_{11} + I_{12}\end{aligned}$$

The integrand of I_{11} converges to

$$-e^{-\int_{S_2}|ts_1+rs_2|\Gamma(d\boldsymbol{s})}\cos\bigg(tx + a\int_{S_2}(ts_1+rs_2)\ln|ts_1+rs_2|\Gamma(d\boldsymbol{s})\bigg)\int_{S_2}s_2(ts_1+rs_2)^{<0>}\Gamma(d\boldsymbol{s}).$$

Using (3.78) we can bound the integrand of I_{11} by

$$\frac{1}{|h|}\bigg|\int_{S_2}|ts_1+(r+h)s_2| - |ts_1+rs_2|\Gamma(d\boldsymbol{s})\bigg|e^{-\int_{S_2}|ts_1+rs_2|\Gamma(d\boldsymbol{s})}e^{\left|\int_{S_2}|ts_1+(r+h)s_2|-|ts_1+rs_2|\Gamma(d\boldsymbol{s})\right|}.$$

By Lemma 3.7.5 (ι) and the triangle inequality, we can further bound it for $|h| < |r|$ by

$$\sigma_2 e^{\sigma_2(1+|r|)-\sigma_1|t|},$$

which does not depend on h and is integrable with respect to t on $\mathbb{R}$. The dominated convergence theorem applies to I_{11}. Turning to I_{12}, its integrand converges to

$$-ae^{-\int_{S_2}|ts_1+rs_2|\Gamma(d\boldsymbol{s})}\sin\bigg(tx + a\int_{S_2}(ts_1+rs_2)\ln|ts_1+rs_2|\Gamma(d\boldsymbol{s})\bigg)\int_{S_2}s_2(1+\ln|ts_1+rs_2|)\Gamma(d\boldsymbol{s}).$$

Using the mean value theorem on the cosine, its integrand can be bounded by

$$\begin{aligned}
&\frac{a}{|h|}e^{-\int_{S_2}|ts_1+rs_2|\Gamma(d\boldsymbol{s})}\left|\int_{S_2}(ts_1+(r+h)s_2)\ln|ts_1+(r+h)s_2|-(ts_1+rs_2)\ln|ts_1+rs_2|\Gamma(d\boldsymbol{s})\right| \\
&\leq ae^{\sigma_2|r|-\sigma_1|t|}\frac{1}{|h|}\int_{S_2}\Big|(ts_1+(r+h)s_2)\ln|ts_1+(r+h)s_2|-(ts_1+rs_2)\ln|ts_1+rs_2|\Big|\Gamma(d\boldsymbol{s}) \\
&:= ae^{\sigma_2|r|-\sigma_1|t|}\Big(Q_1+Q_2\Big), \qquad (3.83)
\end{aligned}$$

where the two terms Q_1 and Q_2 involve integrals over $S_2\cap\{\boldsymbol{s}:|ts_1+rs_2|\geq 2|h|\}$ and $S_2\cap\{\boldsymbol{s}:|ts_1+rs_2|<2|h|\}$. Focus on Q_2. Introduce the function $f:\mathbb{R}_+\to\mathbb{R}_+$ defined for any $z\geq 0$ by $f(z)=z|\ln z|$. It is such that $f(0)=0$ and for z small enough $(0<z<e^{-1})$, f is monotone increasing. Since $|ts_1+rs_2|<2|h|$, we also have $|ts_1+(r+h)s_2|<3|h|$. Thus, for $0<|h|<(3e)^{-1}$, the integrand of Q_2 can be bounded by

$$|h|^{-1}\Big(\Big|f(|3h|)\Big|+\Big|f(|2h|)\Big|\Big)\leq 2|h|^{-1}\Big|f(|3h|)\Big|\leq 6\Big|ln|3h|\Big|$$

Using Lemma 3.7.13, we can bound the later quantity for any $v>0$ by

$$6v^{-1}\Big(2+|3h|^{v}+|3h|^{-v}\Big).$$

From $|ts_1+rs_2|/2<|h|<(3e)^{-1}$, we deduce that $|3h|^{-v}<\Big(3|ts_1+rs_2|/2\Big)^{-v}$ and

$$6v^{-1}\Big(2+|3h|^{v}+|3h|^{-v}\Big)\leq 6v^{-1}\Big(2+e^{-v}+\Big(3|ts_1+rs_2|/2\Big)^{-v}\Big)\leq \mathrm{const}_1+\mathrm{const}_2|ts_1+rs_2|^{-v},$$

for some nonnegative constants const_1 and const_2. Hence, the term involving Q_2 in 3.83 can be further bounded for any $v>0$ by

$$ae^{\sigma_2|r|-\sigma_1|t|}\Big(\mathrm{const}_1+\mathrm{const}_2\int_{S_2}\Big|t+\frac{rs_2}{s_1}\Big|^{-v}|s_1|^{-v}\Gamma(d\boldsymbol{s})\Big). \qquad (3.84)$$

The term with const_1 is clearly integrable with respect to t on $\mathbb{R}$. Letting (3.3) hold with $\nu>0$, choose some $v\in(0,\min(\nu,1))$. We show that the second term is bounded by an integrable function of t as we did in Equation (3.52) using Lemma 3.7.7 with $\eta=v$, $b=0$, $p=0$, the fact that $\int_{\mathbb{R}}e^{-\sigma_1|t|}|t|^{-v}dt<+\infty$ and (3.3) with $\nu>v>0$. There remains to be bounded the part involving Q_1 in (3.83). For this term, we apply the mean value theorem to the function $z\longmapsto z\ln|z|$ and get that

$$\begin{aligned}
|h|^{-1}&\Big|(ts_1+(r+h)s_2)\ln|ts_1+(r+h)s_2|-(ts_1+rs_2)\ln|ts_1+rs_2|\Big| \\
&\leq |h|^{-1}|hs_2|\Big|1+\ln|u|\Big| \\
&\leq 1+\Big|\ln|u|\Big|,
\end{aligned}$$

for some $u\in[ts_1+(r+h)s_2\wedge ts_1+rs_2, ts_1+(r+h)s_2\vee ts_1+rs_2]$. Since Q_1 is an integral over $S_2\cap\{\boldsymbol{s}:|ts_1+rs_2|\geq 2|h|\}$, we have $|u|\in\Big[\frac{|ts_1+rs_2|}{2},2|ts_1+rs_2|\Big]$, and because of the quasi-convexity of the function $z\longmapsto\Big|\ln|z|\Big|$, we can bound the above term by

$$1+\left|\ln\left|\frac{ts_1+rs_2}{2}\right|\right|+\Big|\ln|2(ts_1+rs_2)|\Big|\leq \mathrm{const}+2\Big|\ln|ts_1+rs_2|\Big|.$$

Using Lemma 3.7.13, we can bound this term for any $v > 0$ by

$$\text{const} + 2v^{-1}\Big(2 + |ts_1 + rs_2|^v + |ts_1 + rs_2|^{-v}\Big) \leq \text{const}_1 + \text{const}_2|t|^v + \text{const}_3\Big|t + \frac{rs_2}{s_1}\Big|^{-v}|s_1|^{-v}$$

Hence, the term in (3.83) involving Q_1 can be bounded for any $v > 0$ by

$$ae^{\sigma_2|r|-\sigma_1|t|}\Big(\text{const}_1 + \text{const}_2|t|^v + \text{const}_3\int_{S_2}\Big|t + \frac{rs_2}{s_1}\Big|^{-v}|s_1|^{-v}\Gamma(d\boldsymbol{s})\Big). \tag{3.85}$$

which can be shown to be integrable with respect to t on $\mathbb{R}$ as we did above for the term with Q_2. The dominated convergence theorem applies to I_{12} and thus to I_1. We can derivate $\phi_{X_2|x}$ under the integral sign.

Second derivative

Let us start with A_2, which is the most delicate. It is composed of terms of the form

$$\int_{\mathbb{R}} e^{-\int_{S_2}|ts_1+rs_2|\Gamma(d\boldsymbol{s})}\text{trig}\Big(-tx - a\int_{S_2}(ts_1 + rs_2)\ln|ts_1 + rs_2|\Gamma(d\boldsymbol{s})\Big) \\ \times\Big(\int_{S_2}s_2(1 + \ln|ts_1 + rs_2|)\Gamma(d\boldsymbol{s})\Big)dt,$$

where «trig» stands for sine or cosine. Denoting the one with cosine as K_2, we have

$$\begin{aligned} K_2 = \lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}&\Big[e^{-\int_{S_2}|ts_1+(r+h)s_2|\Gamma(d\boldsymbol{s})} - e^{-\int_{S_2}|ts_1+rs_2|\Gamma(d\boldsymbol{s})}\Big] \\ &\times\cos\Big(tx + a\int_{S_2}(ts_1 + (r+h)s_2)\ln|ts_1 + (r+h)s_2|\Gamma(d\boldsymbol{s})\Big) \\ &\times\Big(\int_{S_2}s_2(1 + \ln|ts_1 + (r+h)s_2|)\Gamma(d\boldsymbol{s})\Big)dt \\ + \lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}&e^{-\int_{S_2}|ts_1+rs_2|\Gamma(d\boldsymbol{s})}\Big[\cos\Big(tx + a\int_{S_2}(ts_1 + (r+h)s_2)\ln|ts_1 + (r+h)s_2|\Gamma(d\boldsymbol{s})\Big) \\ &- \cos\Big(tx + a\int_{S_2}(ts_1 + rs_2)\ln|ts_1 + rs_2|\Gamma(d\boldsymbol{s})\Big)\Big] \\ &\times\Big(\int_{S_2}s_2(1 + \ln|ts_1 + (r+h)s_2|)\Gamma(d\boldsymbol{s})\Big)dt \\ + \lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}&e^{-\int_{S_2}|ts_1+rs_2|\Gamma(d\boldsymbol{s})}\cos\Big(tx + a\int_{S_2}(ts_1 + rs_2)\ln|ts_1 + rs_2|\Gamma(d\boldsymbol{s})\Big) \\ &\times\Big[\int_{S_2}s_2\ln|ts_1 + (r+h)s_2| - s_2\ln|ts_1 + rs_2|\Gamma(d\boldsymbol{s})\Big]dt \\ := K_{21} &+ K_{22} + K_{23}. \end{aligned}$$

The integrand of K_{21} converges to

$$-e^{-\int_{S_2}|ts_1+rs_2|\Gamma(d\boldsymbol{s})}\cos\Big(tx + a\int_{S_2}(ts_1 + rs_2)\ln|ts_1 + rs_2|\Gamma(d\boldsymbol{s})\Big) \\ \times\Big(\int_{S_2}s_2(ts_1 + rs_2)^{<0>}\Gamma(d\boldsymbol{s})\Big)\Big(\int_{S_2}s_2(1 + \ln|ts_1 + rs_2|)\Gamma(d\boldsymbol{s})\Big).$$

Using (3.78), the triangle inequality and (3.7.6), it can be bounded by

$$\sigma_2e^{\sigma_2(1+|r|)-\sigma_1|t|}\int_{S_2}|s_2|\Big|1 + \ln|ts_1 + (r+h)s_2|\Big|\Gamma(d\boldsymbol{s}). \tag{3.86}$$

The integrand of the above expression can be bounded using Lemma 3.7.13 for any $v > 0$ by

$$1 + v^{-1}\Big(2 + |ts_1 + (r+h)s_2|^v + |ts_1 + (r+h)s_2|^{-v}\Big)$$
$$\leq \mathrm{const}_1 + \mathrm{const}_2|t|^v + \mathrm{const}_3\Big|t + \frac{(r+h)s_2}{s_1}\Big|^{-v}|s_1|^{-v},$$

hence, (3.86) is bounded by

$$\sigma_2 e^{\sigma_2(1+|r|)-\sigma_1|t|}\Big(\mathrm{const}_1 + \mathrm{const}_2|t|^v + \mathrm{const}_3\int_{S_2}\Big|t + \frac{(r+h)s_2}{s_1}\Big|^{-v}|s_1|^{-v}\Gamma(d\boldsymbol{s})\Big).$$

The terms involving const_1 and const_2 are clearly integrable with respect to t. The last term is more intricate as it still depends on h. We will show that the generalised Lebesgue dominated convergence theorem (Theorem 19, p.89 in [113]) applies. Denoting

$$T(h) = e^{-\sigma_1|t|}\Big|t + \frac{(r+h)s_2}{s_1}\Big|^{-v}|s_1|^{-v},$$

it can be shown that $T(0)$ is integrable with respect to t on $\mathbb{R}$ and Γ on S_2 invoking the usual arguments. Also, choosing some $v \in (0,1)$, with have by Lemma 3.7.9 with $\eta = -v$, $b = 0$ and $0 < p < 1 - v$,

$$\begin{aligned}\Big|\int T(h) - T(0)\Big| &\leq \int_{S_2}|s_1|^{-v}\int_{\mathbb{R}} e^{-\sigma_1|t|}\Big|\Big|t + \frac{(r+h)s_2}{s_1}\Big|^{-v} - \Big|t + \frac{rs_2}{s_1}\Big|^{-v}\Big|dt\Gamma(d\boldsymbol{s}) \\ &\leq \mathrm{const}\int_{S_2}|s_1|^{-v}\Big|\frac{hs_2}{s_1}\Big|^p\Gamma(d\boldsymbol{s}) \\ &\leq \mathrm{const}\ |h|^p\int_{S_2}|s_1|^{-v-p}\Gamma(d\boldsymbol{s}) \underset{h\to 0}{\longrightarrow} 0,\end{aligned}$$

because (3.3) holds with $\nu > 1$ and $v+p < v+1-v < 1$. Since $T(0)$ is integrable and $\lim_{h\to 0}\int T(h) = \int T(0)$, the generalised dominated convergence theorem applies to K_{21}. We turn to K_{22}. Its integrand converges to

$$-ae^{-\int_{S_2}|ts_1+rs_2|\Gamma(d\boldsymbol{s})}\sin\Big(tx + a\int_{S_2}(ts_1+rs_2)\ln|ts_1+rs_2|\Gamma(d\boldsymbol{s})\Big)$$
$$\times\Big(\int_{S_2}s_2(1+\ln|ts_1+rs_2|)\Gamma(d\boldsymbol{s})\Big)^2.$$

With the usual inequalities and Lemma 3.7.13, it can be bounded for any $v > 0$ by

$$\begin{aligned}&\frac{a}{|h|}e^{\sigma_2|r|-\sigma_1|t|}\Big|\int_{S_2}(ts_1+(r+h)s_2)\ln|ts_1+(r+h)s_2| - (ts_1+rs_2)\ln|ts_1+rs_2|\Gamma(d\boldsymbol{s})\Big| \\ &\qquad\qquad\times\Big|\int_{S_2}s_2(1+\ln|ts_1+(r+h)s_2|)\Gamma(d\boldsymbol{s})\Big| \\ &\leq ae^{\sigma_2|r|-\sigma_1|t|}\Big(Q_1+Q_2\Big)\Big(\sigma_2 + \int_{S_2}\Big|\ln|ts_1+(r+h)s_2|\Big|\Gamma(d\boldsymbol{s})\Big) \\ &\leq ae^{\sigma_2|r|-\sigma_1|t|}\Big(Q_1+Q_2\Big)\Big(\mathrm{const}_1 + \mathrm{const}_2|t|^v + \mathrm{const}_3\int_{S_2}\Big|t + \frac{(r+h)s_2}{s_1}\Big|^{-v}|s_1|^{-v}\Gamma(d\boldsymbol{s})\Big),\end{aligned}$$

where, similarly to (3.83), the two terms Q_1 and Q_2 involve integrals over $S_2 \cap \{\boldsymbol{s} : |ts_1 + rs_2| \geq 2|h|\}$ and $S_2 \cap \{\boldsymbol{s} : |ts_1 + rs_2| < 2|h|\}$. After expansion, the terms with const_1 and const_2 are readily dealt with by following the method developed for (3.83). Focus on the remaining term

$$a\int_{S_2}e^{\sigma_2|r|-\sigma_1|t|}(Q_1+Q_2)\Big|t + \frac{(r+h)s_2}{s_1}\Big||s_1|^{-v}\Gamma(d\boldsymbol{s}).$$

In view of the bounds (3.84) and (3.85), the integrand can be bounded (up to a multiplicative constant) by

$$U(h) = e^{-\sigma_1|t|}\Big|t+\frac{rs_2}{s_1}\Big|^{-v}\Big|t+\frac{(r+h)s_2'}{s_1'}\Big|^{-v}|s_1|^{-v}|s_1'|^{-v}.$$

Choosing some $v\in(0,1/2)$, we can invoke Lemma (3.7.8) with $\eta=-v$, $p=0$ and the fact that $\int_{\mathbb{R}}e^{-\sigma_1|t|}|t|^{-2v}dt<+\infty$ to show that $U(0)$ is integrable on the one hand. On the other hand we can again invoke Lemma (3.7.8), this time with $\eta=-v$, $0<p<1-2v$, and the fact that (3.3) holds with $\nu>1>v+1-2v>v+p$ to show that $\int U(h)\to\int U(0)$. The generalised dominated convergence theorem applies to K_{12}.

We turn to K_{23} for which «appropriate integration by parts» is required. After obvious manipulations,

$$\begin{aligned}
K_{23}=\lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}\int_{S_2}&s_2'\ln|ts_1'+rs_2'|\Bigg[e^{-\int_{S_2}\left|\left(t-\frac{hs_2'}{s_1'}\right)s_1+rs_2\right|\Gamma(d\boldsymbol{s})}-e^{-\int_{S_2}|ts_1+rs_2|\Gamma(d\boldsymbol{s})}\Bigg]\\
&\times\cos\Bigg(\Big(t-\frac{hs_2'}{s_1'}\Big)x+a\int_{S_2}\Bigg(\Big(t-\frac{hs_2'}{s_1'}\Big)s_1+rs_2\Bigg)\ln\Bigg|\Big(t-\frac{hs_2'}{s_1'}\Big)s_1+rs_2\Bigg|\Gamma(d\boldsymbol{s})\Bigg)\Gamma(d\boldsymbol{s}')\\
+\lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}}\int_{S_2}&s_2'\ln|ts_1'+rs_2'|e^{-\int_{S_2}|ts_1+rs_2|\Gamma(d\boldsymbol{s})}\\
&\times\Bigg[\cos\Bigg(\Big(t-\frac{hs_2'}{s_1'}\Big)x+a\int_{S_2}\Bigg(\Big(t-\frac{hs_2'}{s_1'}\Big)s_1+rs_2\Bigg)\ln\Bigg|\Big(t-\frac{hs_2'}{s_1'}\Big)s_1+rs_2\Bigg|\Gamma(d\boldsymbol{s})\Bigg)\\
&\qquad-\cos\Bigg(tx+a\int_{S_2}(ts_1+rs_2)\ln|ts_1+rs_2|\Gamma(d\boldsymbol{s})\Bigg)\Bigg]\Gamma(d\boldsymbol{s}')\\
:=L_1+L_2.&
\end{aligned}$$

Starting with L_1, its integrand converges to

$$\begin{aligned}
e^{-\int_{S_2}|ts_1+rs_2|\Gamma(d\boldsymbol{s})}&\cos\Bigg(tx+a\int_{S_2}(ts_1+rs_2)\ln|ts_1+rs_2|\Gamma(d\boldsymbol{s})\Bigg)\\
&\times\Bigg(\int_{S_2}s_1(ts_1+rs_2)^{<0>}\Gamma(d\boldsymbol{s})\Bigg)\Bigg(\int_{S_2}\ln|ts_1+rs_2|{s_2}^2{s_1}^{-1}\Gamma(d\boldsymbol{s})\Bigg)
\end{aligned}$$

It can be bounded using (3.77) and Lemma 3.7.5 (ι) by

$$\begin{aligned}
&\Bigg|\frac{s_2'\ln|ts_1'+rs_2'|}{h}\Bigg|\exp\Bigg\{-\min\Bigg(\int_{S_2}\Bigg|\Big(t-\frac{hs_2'}{s_1'}\Big)s_1+rs_2\Bigg|\Gamma(d\boldsymbol{s}),\int_{S_2}|ts_1+rs_2|\Gamma(d\boldsymbol{s})\Bigg)\Bigg\}\\
&\qquad\times\Bigg|\int_{S_2}\Bigg|\Big(t-\frac{hs_2'}{s_1'}\Big)s_1+rs_2\Bigg|-|ts_1+rs_2|\Gamma(d\boldsymbol{s})\Bigg|\\
&\le e^{\sigma_2|r|}\exp\Bigg\{-\sigma_1\min\Bigg(\Big|t-\frac{hs_2'}{s_1'}\Big|,|t|\Bigg)\Bigg\}\Big|s_2'\ln|ts_1'+rs_2'|\Big|\frac{1}{|h|}\int_{S_2}\Big|\frac{hs_2'}{s_1'}s_1\Big|\Gamma(d\boldsymbol{s})\\
&\le\sigma_1e^{\sigma_2|r|}\exp\Bigg\{-\sigma_1\min\Bigg(\Big|t-\frac{hs_2'}{s_1'}\Big|,|t|\Bigg)\Bigg\}\Big|\ln|ts_1'+rs_2'|\Big||s_2'|^2|s_1'|^{-1}\\
&:=V(h).
\end{aligned}$$

We follow a similar procedure as the one used in [33] (p.51) to deal with the min inside the exponential.

Focus on the case $\dfrac{hs_2}{s_1} > 0$ (the converse case is similar). We have

$$\min\left(\left|t - \frac{hs_2'}{s_1'}\right|, |t|\right) = \begin{cases} \left|t - \dfrac{hs_2'}{s_1'}\right|, & \text{if} \quad t \geq hs_2'/2s_1', \\ |t|, & \text{if} \quad t < hs_2'/2s_1'. \end{cases}$$

Thus, up to a multiplicative constant,

$$\begin{aligned}
\int_{\mathbb{R}} V(h)dt &= \int_{\frac{hs_2}{2s_1}}^{+\infty} e^{-\sigma_1|t-\frac{hs_2}{s_1}|}\Big|\ln|ts_1+rs_2|\Big||s_2|^2|s_1|^{-1}dt + \int_{-\infty}^{-\frac{hs_2}{2s_1}} e^{-\sigma_1|t|}\Big|\ln|ts_1+rs_2|\Big||s_2|^2|s_1|^{-1}dt \\
&= \int_{-\frac{hs_2}{2s_1}}^{+\infty} e^{-\sigma_1|t|}\left|\ln\left|ts_1+rs_2+\frac{hs_2}{s_1}\right|\right||s_2|^2|s_1|^{-1}dt + \int_{-\infty}^{-\frac{hs_2}{2s_1}} e^{-\sigma_1|t|}\Big|\ln|ts_1+rs_2|\Big||s_2|^2|s_1|^{-1}dt \\
&= \int_{\mathbb{R}} e^{-\sigma_1|t|}\left[\Big|\ln|ts_1+(r+h)s_2|\Big|\mathbb{1}_{\{t\geq -hs_2/2s_1\}} + \Big|\ln|ts_1+rs_2|\Big|\mathbb{1}_{\{t\leq -hs_2/2s_1\}}\right]|s_2|^2|s_1|^{-1}dt.
\end{aligned}$$

Thus, using Lemma 3.7.13, we can bound the integrand for any $v > 0$ and $|h| < |r|$ by

$$\begin{aligned}
&e^{-\sigma_1|t|}\left[\Big|\ln|ts_1+(r+h)s_2|\Big| + \Big|\ln|ts_1+rs_2|\Big|\right]|s_2|^2|s_1|^{-1} \\
&\quad\leq v^{-1}e^{-\sigma_1|t|}\Bigg[\text{const}_1 + \text{const}_2|t|^v \\
&\qquad\qquad + \text{const}_3\left|t+\frac{rs_2}{s_1}\right|^{-v}|s_1|^{-v} + \text{const}_4\left|t+\frac{(r+h)s_2}{s_1}\right|^{-v}|s_1|^{-v}\Bigg]|s_2|^2|s_1|^{-1}.
\end{aligned}$$

Clearly, the terms involving const_1 and const_2 are integrable with respect to t and Γ. Denoting the last term as $V_4(h) := e^{-\sigma_1|t|}\left|t+\dfrac{(r+h)s_2}{s_1}\right|^{-v}|s_2|^2|s_1|^{-1-v}$, we show that the generalised dominated convergence theorem applies. As (3.3) holds for some $\nu > 1$, choose $v = \dfrac{\nu-1}{2} > 0$ if $\nu < 2$, and some $v \in (0,1)$ if $\nu \geq 2$. The integrability of $V_4(0)$ (and at the same time, of the term involving const_3) is obtained from Lemma 3.7.7 with $\eta = -v$, $b = 0$, $p = 0$ and the fact that $\int_{\mathbb{R}} e^{-\sigma_1|t|}|t|^{-v}dt < +\infty$. Doing so indeed yields

$$\begin{aligned}
&\left|\int_{S_2}|s_2|^2|s_1|^{-1-v}\int_{\mathbb{R}} e^{-\sigma_1|t|}\left|t+\frac{rs_2}{s_1}\right|^{-v} - |t|^{-v}|s_2|^2|s_1|^{-1-v}dt\right|\Gamma(d\boldsymbol{s}) \\
&\qquad\leq \int_{S_2}\int_{\mathbb{R}} e^{-\sigma_1|t|}\left|\left|t+\frac{rs_2}{s_1}\right|^{-v} - |t|^{-v}\right|dt\Gamma(d\boldsymbol{s}) \\
&\qquad\leq \text{const}\int_{S_2}|s_1|^{-\nu}|s_1|^{\nu-1-v}\Gamma(d\boldsymbol{s}) \\
&\qquad\leq \text{const}\int_{S_2}|s_1|^{-\nu}\Gamma(d\boldsymbol{s}) \\
&\qquad< +\infty,
\end{aligned}$$

since $\nu - 1 - v = \dfrac{\nu-1}{2} > 0$ if $\nu \in (1,2)$ and $\nu - 1 - v > \nu - 2 > 0$ if $\nu \geq 2$. The convergence $\int V_4(h) \to \int V_4(0)$ can be obtained from Lemma 3.7.9 with $\eta = -v$, $b = 0$ and $0 < p < v$. The generalised dominated convergence hence applies to L_1.

We turn to L_2. Its integrand converges to

$$e^{-\int_{S_2}|ts_1+rs_2|\Gamma(d\boldsymbol{s})}\sin\Big(tx+a\int_{S_2}(ts_1+rs_2)\ln|ts_1+rs_2|\Gamma(d\boldsymbol{s})\Big)$$
$$\times\Big(x+a\int_{S_2}s_1(1+\ln|ts_1+rs_2|)\Gamma(d\boldsymbol{s})\Big)\ln|ts_1'+rs_2'|{s_2'}^2{s_1'}^{-1}.$$

Applying the mean value theorem to the cosine function and the usual bounds, we can bound it by

$$\begin{aligned}
&e^{\sigma_2|r|-\sigma_1|t|}\Big|s_2'^2s_1'^{-1}\ln|ts_1'+rs_2'|\Big|\\
&\frac{1}{\Big|\frac{hs_2'}{s_1'}\Big|}\Big|-\frac{hs_2'}{s_1'}x+a\int_{S_2}\Big(\Big(t-\frac{hs_2'}{s_1'}\Big)s_1+rs_2\Big)\ln\Big|\Big(t-\frac{hs_2'}{s_1'}\Big)s_1+rs_2\Big|-(ts_1+rs_2)\ln|ts_1+rs_2|\Gamma(d\boldsymbol{s})\Big|\\
&\le e^{\sigma_2|r|-\sigma_1|t|}\Big|s_2'^2s_1'^{-1}\ln|ts_1'+rs_2'|\Big|\\
&\Big(|x|+\frac{a}{\Big|\frac{hs_2'}{s_1'}\Big|}\int_{S_2}\Big|\Big(\Big(t-\frac{hs_2'}{s_1'}\Big)s_1+rs_2\Big)\ln\Big|\Big(t-\frac{hs_2'}{s_1'}\Big)s_1+rs_2\Big|-(ts_1+rs_2)\ln|ts_1+rs_2|\Big|\Gamma(d\boldsymbol{s})\Big).
\end{aligned}\tag{3.87}$$

The term involving $|x|$ can be treated using the usual arguments. The one with the integral is of course the most delicate. Let us split this integral into two parts as:

$$\begin{aligned}
&\int_{S_2}\frac{1}{\Big|\frac{hs_2'}{s_1'}\Big|}\Big|\Big(\Big(t-\frac{hs_2'}{s_1'}\Big)s_1+rs_2\Big)\ln\Big|\Big(t-\frac{hs_2'}{s_1'}\Big)s_1+rs_2\Big|-(ts_1+rs_2)\ln|ts_1+rs_2|\Big|\Gamma(d\boldsymbol{s})\\
&\qquad:=Q_1+Q_2,
\end{aligned}$$

where Q_1 and Q_2 involve integrals over $S_2\cap\{\boldsymbol{s}:|ts_1+rs_2|\ge 2|hs_2'/s_1'|\}$ and $S_2\cap\{\boldsymbol{s}:|ts_1+rs_2|<2|hs_2'/s_1'|\}$ respectively. We will first majorise Q_1 and Q_2, and then use these bounds in inequality (3.87). Consider Q_2 and define the function g such that for any $z>0$

$$g(z)=\begin{cases} f(z)=z|\ln z|, & \text{if } \ 0<z<e^{-1},\\ z(2+\ln z), & \text{if } \quad z\ge e^{-1}.\end{cases}$$

It is easily checked that g is continuous, strictly increasing and such that for any $z>0$, $0\le f(z)\le g(z)$. The integrand of Q_2 can be bounded as

$$\begin{aligned}
\frac{1}{\Big|\frac{hs_2'}{s_1'}\Big|}\Big(\Big|f\Big(\Big(t-\frac{hs_2'}{s_1'}\Big)s_1+rs_2\Big)\Big|+\Big|f\Big(ts_1+rs_2\Big)\Big|\Big)&\le\frac{1}{\Big|\frac{hs_2'}{s_1'}\Big|}\Big(\Big|g\Big(\Big(t-\frac{hs_2'}{s_1'}\Big)s_1+rs_2\Big)\Big|+\Big|g\Big(ts_1+rs_2\Big)\Big|\Big)\\
&\le\frac{1}{\Big|\frac{hs_2'}{s_1'}\Big|}\Big(\Big|g\Big(\Big|\frac{3hs_2'}{s_1'}\Big|\Big)\Big|+\Big|g\Big(\Big|\frac{2hs_2'}{s_1}\Big|\Big)\Big|\Big)\\
&\le\frac{2}{\Big|\frac{hs_2'}{s_1'}\Big|}g\Big(\frac{3hs_2'}{s_1'}\Big).
\end{aligned}$$

By Lemma (3.7.13), with bound further the right-hand side for any $v>0$ by

$$\frac{2}{\Big|\frac{hs_2'}{s_1'}\Big|}g\Big(\frac{3hs_2'}{s_1'}\Big)\le\text{const}_1+\text{const}_2\Big|\frac{3hs_2'}{s_1'}\Big|^{v}+\text{const}_3\Big|\frac{3hs_2'}{s_1'}\Big|^{-v}.$$

On the one hand if $\Big|\dfrac{3hs_2'}{s_1'}\Big| < e^{-1}$, given that $(3|ts_1+rs_2|/2)^{-v} > (3hs_2'/s_1')^{-v}$,

$$\mathrm{const}_1+\mathrm{const}_2\Big|\frac{3hs_2'}{s_1'}\Big|^{v}+\mathrm{const}_3\Big|\frac{3hs_2'}{s_1'}\Big|^{-v}\le \mathrm{const}_1+\mathrm{const}_2\Big|t+\frac{rs_2}{s_1}\Big|^{-v}|s_1|^{-v}.$$

On the other hand if $\Big|\dfrac{3hs_2'}{s_1'}\Big| \ge e^{-1}$, then for $|h|<|r|$,

$$\mathrm{const}_1+\mathrm{const}_2\Big|\frac{3hs_2'}{s_1'}\Big|^{v}+\mathrm{const}_3\Big|\frac{3hs_2'}{s_1'}\Big|^{-v}\le \mathrm{const}_1+\mathrm{const}_2|s_1'|^{-v}. \tag{3.88}$$

Focusing now on Q_1, we can use the mean value theorem to bound its integrand by

$$|s_1|\Big|1+\ln|u|\Big|,$$

for some $u\in\Big[ts_1+rs_2-hs_2's_1/s_1'\wedge ts_1+rs_2, ts_1+rs_2-hs_2's_1/s_1'\vee ts_1+rs_2\Big]$. Given that $|ts_1+rs_2|\ge 2|hs_2'/s_1'|$, we have $|u|\in\Big[\frac{|ts_1+rs_2|}{2}, 2|ts_1+rs_2|\Big]$ and thus, we further bound the above inequality using Lemma 3.7.13 for any $v>0$ by

$$\begin{aligned}|s_1|\Big(&\mathrm{const}_1+\mathrm{const}_2|ts_1+rs_2|^{v}+\mathrm{const}_3|ts_1+rs_2|^{-v}\Big)\\ &\le \mathrm{const}_1+\mathrm{const}_2|t|^{v}+\mathrm{const}_3\Big|t+\frac{rs_2}{s_1}\Big|^{-v}|s_1|^{1-v}.\end{aligned} \tag{3.89}$$

Hence, using (3.88) and (3.89) in (3.87), and making use again of Lemma (3.7.13) to bound $\Big|\ln|ts_1'+rs_2'|\Big|$, we can bound integrand of L_2 for any $v>0$ by

$$\begin{aligned}e^{-\sigma_1|t|}&\Big(\mathrm{const}_1+\mathrm{const}_2|t|^{v}+\mathrm{const}_3\Big|t+\frac{rs_2'}{s_1'}\Big|^{-v}\Big)|s_1'|^{-1-v}\\ &\times\Big(|x|+\mathrm{const}_4+\mathrm{const}_5|t|^{v}+\mathrm{const}_6|s_1'|^{-v}+\mathrm{const}_7\Big|t+\frac{rs_2}{s_1}\Big|^{-v}|s_1|^{1-v}\Big)\end{aligned}$$

It can be shown that all the terms obtained after expansion can be bounded by functions integrable with respect to t and Γ using the usual combinations of either Lemma 3.7.7 or Lemma 3.7.8 with $\eta=-v$, $b=0$, $p=0$, the fact that $\int_{\mathbb{R}} e^{-\sigma_1|t|}|t|^{-v}<+\infty$, $\int_{\mathbb{R}} e^{-\sigma_1|t|}|t|^{-2v}<+\infty$ for appropriately chosen values $v>0$, and (3.3) with $\nu>1$. The detail we have to pay attention to is precisely to chose an appropriate exponent $v>0$ so that it satisfies the constraint (3.3) and ensures the finiteness of the two integrals in t. The later imposes us to have $v\in(0,1/2)$. Regarding the former, we identify that the most negative power of which $|s_1|$ appears in the above bound after expansion is $-1-2v$. We need $\nu-1-2v>0$. Choosing $v=(\nu-1)/4$ if $1<\nu<3$ and any $v\in(0,1/2)$ if $\nu\ge 3$ enables to satisfy both constraints, validating the use of the dominated convergence theorem for L_2, and finally, for B_2 in (3.81).

The proof is essentially similar, somewhat easier, for B_1 in (3.80) for which the only difficulty is to perform the «appropriate integration by parts» when it comes to differentiating the term involving $(ts_1+rs_2)^{<0>}$.

Evaluating at $r=0$ Since $\mathbb{E}\Big[X_2^2\Big|X_1=x\Big]=-\phi^{(2)}_{X_2|x}(0)$, we evaluate (3.82) at $r=0$ and get

$$\begin{aligned}
\varphi_{\boldsymbol{X}}(t,0)&=\exp\{-\sigma_1|t|-ia\sigma_1\beta_1 t\ln|t|+it\mu_1\},\\
A_1/2&=\sigma_1^2\Big((\kappa_1^2-a^2q_0^2)H_c(0)+2a\kappa_1q_0H_s(0)\Big)\\
&\qquad\qquad+2a\lambda_1\sigma_1^2\Big(-aq_0H_c(1)+\kappa_1H_s(1)\Big)-a^2\lambda_1^2\sigma_1^2H_c(2),\\
iA_2/2&=\sigma_1\Big(-ak_1H_c(0)+\kappa_2H_s(0)\Big)-a\lambda_2\sigma_1H_c(1),\\
A_3/2&=\sigma_1\Big((\sigma_1\kappa_2+a\mu_1k_1)H_c(0)+(\sigma_1ak_1-\mu_1\kappa_2)H_s(0)\Big)\\
&\qquad+a\sigma_1\Big((\lambda_2\mu_1-a\sigma_1\beta_1k_1)H_c(1)+\sigma_1(\lambda_2+\beta_1\kappa_2)H_s(1)\Big)-a^2\sigma_1^2\beta_1\lambda_2H_c(2),
\end{aligned}$$

where $k_1=\sigma_1^{-1}\int_{S_2}(s_2/s_1)^2s_1\ln|s_1|\Gamma(d\boldsymbol{s})$, and the H_c's and H_s's are defined at Lemma 3.7.14. Using the result of the same Lemma under $\beta_1\neq 0$ and $\beta_1=0$, and regrouping the terms allows to retrieve the two formulae of Theorem 3.6.4.

3.7.15 Proof of Proposition 3.6.1 in the case $\alpha=1$

Case $\boldsymbol{\beta_1\neq 0}$ The conditional second order moment when $\alpha=1$ has a particular form. We only consider the case $|\beta_1|\neq 1$ and $x\longrightarrow+\infty$. Since $|x|\to+\infty$, we have $x-\mu_1\sim x$ and we may assume that $\mu_1=0$. From [64], we know that $U(x)\sim x^{-1}$. Notice that

$$\begin{aligned}
W(x)=&\int_0^{+\infty}e^{-\sigma_1t}(1+\ln t)^2\cos(a\sigma_1\beta_1t\ln t)\cos(tx)dt\\
&-\int_0^{+\infty}e^{-\sigma_1t}(1+\ln t)^2\sin(a\sigma_1\beta_1t\ln t)\sin(tx)dt.
\end{aligned}$$

Because the factors of $\cos(tx)$ and $\sin(tx)$ are integrable, we have by the Riemann-Lebesgue Lemma that $W(x)\underset{x\to+\infty}{\longrightarrow}0$. Having also

$$f_{X_1}(x)\sim\frac{\sigma_1(1+\beta_1)}{\pi}x^{-2},$$

we deduce the following limits

$$\begin{aligned}
\Big(2a\sigma_1q_0(\lambda_1-\beta_1\kappa_1)+2(\kappa_1\lambda_1-\lambda_2)x\Big)\frac{\sigma_1U(x)}{\beta_1\pi f_{X_1}(x)}x^{-2}&\underset{x\to+\infty}{\longrightarrow}\frac{2(\kappa_1\lambda_1-\lambda_2)}{(1+\beta_1)\beta_1},\\
\Big(\lambda_2+\beta_1\kappa_2-2\kappa_1\lambda_1+a^2\sigma_1\beta_1(\lambda_1^2-\beta_1\lambda_2)W(x)\Big)\frac{\sigma_1x^{-2}}{\pi f_{X_1}(x)}&\underset{x\longrightarrow+\infty}{\longrightarrow}\frac{\lambda_2+\beta_1\kappa_2-2\kappa_1\lambda_1}{(1+\beta_1)\beta_1}.
\end{aligned}$$

Hence,

$$x^{-2}\mathbb{E}\Big[X_2^2\Big|X_1=x\Big]\underset{x\to+\infty}{\longrightarrow}\frac{\lambda_2}{\beta_1}+\frac{2(\kappa_1\lambda_1-\lambda_2)}{(1+\beta_1)\beta_1}+\frac{\lambda_2+\beta_1\kappa_2-2\kappa_1\lambda_1}{(1+\beta_1)\beta_1}=\frac{\kappa_2+\lambda_2}{1+\beta_1}$$

Case $\boldsymbol{\beta_1=0}$ From [64],

$$V(x)\longrightarrow-\frac{\pi}{2x},$$

hence,

$$2a\sigma_1\lambda_1\Big(a\sigma_1 q_0 - \kappa_1(x-\mu_1)\Big)\frac{V(x)}{\pi f_{X_1}(x)}x^{-2} \longrightarrow a\pi\lambda_1\kappa_1.$$

Moreover,

$$a\sigma_1\frac{F_{X_1}(x)-1/2}{f_{X_1}(x)}x^{-2} \longrightarrow \frac{1}{2}a\pi(\lambda_2 - 2\kappa_1\lambda_1).$$

It can be shown that $W(x) \longrightarrow 0$. Therefore,

$$x^{-2}\mathbb{E}\Big[X_2^2\Big|X_1 = x\Big] \underset{x\to+\infty}{\longrightarrow} \kappa_2 + \frac{1}{2}a\pi(\lambda_2 - 2\kappa_1\lambda_1) + a\pi\kappa_1\lambda_1 = \kappa_2 + \lambda_2$$

□

Lemma 3.7.13 *For any $x > 0$ and $v > 0$*

$$|\ln x| \leq \frac{1}{v}\Big(2 + x^v + x^{-v}\Big).$$

We provide here two Lemmas which are used in the proof of Theorem 3.6.4.

Lemma 3.7.14 *Let for any $n \geq 0$,*

$$H_c(n) = \int_0^{+\infty} e^{-\sigma_1 t}(1+\ln t)^n \cos\Big(t(x-\mu_1) + a\sigma_1\beta_1 t\ln t\Big)dt,$$
$$H_s(n) = \int_0^{+\infty} e^{-\sigma_1 t}(1+\ln t)^n \sin\Big(t(x-\mu_1) + a\sigma_1\beta_1 t\ln t\Big)dt.$$

Then, if $\beta_1 \neq 0$,

$$H_c(1) = \frac{1}{a\sigma_1\beta_1}\Big(\sigma_1 H_s(0) - (x-\mu_1)H_c(0)\Big), \qquad H_s(1) = \frac{1}{a\sigma_1\beta_1}\Big(1 - \sigma_1 H_c(0) - (x-\mu_1)H_s(0)\Big).$$

If $\beta_1 = 0$,

$$H_c(0) = \pi f_{X_1}(x),$$
$$H_s(0) = \frac{x-\mu_1}{\sigma_1}\pi f_{X_1}(x),$$
$$H_s(1) - \frac{x-\mu_1}{\sigma_1}H_c(1) = \frac{\pi F_{X_1}(x)}{\sigma_1}.$$

Proof. The equalities of Lemmas 3.7.10-3.7.14 can be obtained by integrating by parts. We provide details for the last equality of Lemma 3.7.14 when $\beta_1 = 0$. Integrating the exponential by parts, we obtain

$$H_s(1) = \frac{1}{\sigma_1}\int_0^{+\infty} e^{-\sigma_1 t}t^{-1}\sin\Big(t(x-\mu_1)\Big)dt + \frac{x-\mu_1}{\sigma_1}H_c(1)$$

Denote $A(x) = \int_0^{+\infty} e^{-\sigma_1 t}t^{-1}\sin\Big(t(x-\mu_1)\Big)dt$ for $x \in \mathbb{R}$ (A is well defined since $e^{-\sigma_1 t}t^{-1}\sin\Big(t(x-\mu_1)\Big) \to x-\mu_1$ as $t \to 0$). It can be shown that we can derivate A under the integral sign and get

$$A'(x) = \int_0^{+\infty} e^{-\sigma_1 t}\cos\Big(t(x-\mu_1)\Big)dt = \pi f_{X_1}(x),$$

Since X_1 is Cauchy distributed when $\alpha = 1$ and $\beta_1 = 0$,

$$A(x) = \pi F_{X_1}(x) + \text{const} = \operatorname{Arctg}\Big(\frac{x-\mu_1}{\sigma_1}\Big) + \frac{\pi}{2} + \text{const},$$

and evaluating the integral form of A at μ_1, we deduce that const $= -\pi/2$. Thus, $A(x) = \pi\Big(F_{X_1}(x) - 1/2\Big)$.

$\square$

Chapter 4

Path prediction of aggregated α-stable moving averages using semi-norm representations

Sébastien Fries

Abstract For (X_t) a two-sided α-stable moving average, this paper studies the conditional distribution of future paths given a piece of observed trajectory when the process is far from its central values. Under this framework, vectors of the form $\boldsymbol{X_t} = (X_{t-m}, \dots, X_t, X_{t+1}, \dots, X_{t+h})$, $m \geq 0$, $h \geq 1$, are multivariate α-stable and the dependence between the past and future components is encoded in their spectral measures. A new representation of stable random vectors on unit cylinders –sets $\{\boldsymbol{s} \in \mathbb{R}^{m+h+1} : \quad \|\boldsymbol{s}\| = 1\}$ for $\|\cdot\|$ an adequate semi-norm– is proposed in order to describe the tail behaviour of vectors $\boldsymbol{X_t}$ when only the first $m+1$ components are assumed to be observed and large in norm. Not all stable vectors admit such a representation and (X_t) will have to be «anticipative enough» for $\boldsymbol{X_t}$ to admit one. The conditional distribution of future paths can then be explicitly derived using the regularly varying tails property of stable vectors and has a natural interpretation in terms of pattern identification. The approach extends to processes resulting from the linear combination of stable moving averages which feature much richer dynamics and applied to several examples.

Keywords: Anticipative processes, Noncausal processes, Stable processes, Stable random vectors, Spectral representation, Pattern identification, Prediction

MSC classes: 60G52, 60E07, 60G25

4.1 Introduction

Stochastic processes depending on the «future» values of an independent and identically distributed (i.i.d.) sequence, often referred to as *anticipative*, have witnessed a recent surge of attention from the statistical and econometric literatures. This gain of interest is driven in particular by their convenience for modelling exotic patterns in time series, such as explosive bubbles in financial prices [24, 35, 51, 52, 61, 63, 67, 68, 69, 70] (see also [3, 9, 10, 26, 59, 60, 87, 88, 115]). The attractive flexibility of anticipative processes cannot yet be fully leveraged however, as their dynamics, and especially the conditional distribution of future paths given the observed past trajectory, remains largely mysterious. A remarkable exception is that of the anticipative α-stable AR(1) for which partial results were obtained in [63] and further completed in [52]. Even in this simplest case within the family of anticipative processes however, future realisations feature a complex dependence on the observed past, which is reflected in the functional forms of the conditional moments obtained in [52]. Interestingly, the dynamics of the anticipative stable AR(1) simplifies during extreme events where it appears to follow an explosive exponential path with a determined killing probability. This naturally raises the question of whether and under which form such a behaviour could be found in more general stable processes.

For $X_t = \sum_{k\in\mathbb{Z}} d_k\varepsilon_{t+k}$ a two-sided moving average with (ε_t) an i.i.d. α-stable sequence and (d_k) a non-random coefficients sequence, this paper analyses the conditional distribution of future paths given the observed trajectory, say $(X_{t+1},\ldots,X_{t+h})$ given $(X_{t-m},\ldots,X_t)$, $m\geq 0$, $h\geq 1$, when the process is far from its central values. Only mild summability conditions are assumed on the sequence (d_k) and, in particular, we do not presume anything upfront on the anticipativeness or non-anticipativeness of (X_t).[1] Under this framework, any vector of the form $\boldsymbol{X_t} = (X_{t-m},\ldots,X_{t+h})$ is multivariate α-stable and its distribution is characterised by a unique finite measure Γ on the Euclidean unit sphere $S_{m+h+1} = \{\boldsymbol{s}\in\mathbb{R}^{m+h+1}: \ \|\boldsymbol{s}\|_e = 1\}$, where $\|\cdot\|_e$ denotes the Euclidean norm (Theorem 2.3.1 in [117]). The measure Γ in particular completely describes the conditional distribution of the normalised paths $\boldsymbol{X_t}/\|\boldsymbol{X_t}\|_e$, the «shape» of the trajectory, when $\boldsymbol{X_t}$ is large according to the Euclidean norm and given some information about the observed first $m+1$ components. A straightforward application of Theorem 4.4.8 by Samorodnitsky and Taqqu (1994) [117] indeed shows that

$$\mathbb{P}\Big(\boldsymbol{X}_t/\|\boldsymbol{X}_t\|_e\in A \ \Big| \ \|\boldsymbol{X}_t\|_e > x \ \text{ and } \ \boldsymbol{X}_t/\|\boldsymbol{X}_t\|_e\in B\Big) \underset{x\longrightarrow\infty}{\longrightarrow} \frac{\Gamma(A\cap B)}{\Gamma(B)}, \tag{4.1}$$

for any appropriately chosen Borel sets $A,B\subset S_{m+h+1}$. As such however, (4.1) is of little value for prediction purposes where only $X_{t-m},\ldots,X_t$ are assumed to be observed, given that the conditioning generally depends on the future realisations $X_{t+1},\ldots,X_{t+h}$, mainly through the Euclidean norm of $\boldsymbol{X}_t$. The idea developed here is to obtain a version of (4.1) where the Euclidean norm is replaced by a semi-norm $\|\cdot\|$ satisfying

$$\|(x_{-m},\ldots,x_0,x_1,\ldots,x_h)\| = \|(x_{-m},\ldots,x_0,0,\ldots,0)\|, \tag{4.2}$$

[1] That is, we do not presume anything on the zeros of (d_k), e.g., $d_k = 0$ for $k>0$ (purely non-anticipative case) or $k<0$ (purely anticipative case).

for any $(x_{-m},\dots,x_h) \in \mathbb{R}^{m+h+1}$. In this view, a new representation of stable random vectors on the «unit cylinder» $C^{\|\cdot\|}_{m+h+1} := \{\boldsymbol{s} \in \mathbb{R}^{m+h+1} : \|\boldsymbol{s}\| = 1\}$ is thus explored, where $\|\cdot\|$ is such a semi-norm. Contrary to representations involving norms (see Theorem 2.3.8 in [117]), not all stable random vectors admit representations on unit cylinders and a characterisation is provided. It is shown that only if (X_t) is «anticipative enough» will $\boldsymbol{X_t}$ admit a representation by a measure $\Gamma^{\|\cdot\|}$ on $C^{\|\cdot\|}_{m+h+1}$. Property (4.1) is then shown to hold with an adequate semi-norm and with Γ (resp. S_{m+h+1}) replaced by $\Gamma^{\|\cdot\|}$ (resp. $C^{\|\cdot\|}_{m+h+1}$). The problem finally boils down to choosing appropriate Borels B in (4.1) reflecting that only the past «shape» $(X_{t-m},\dots,X_t)/\|\boldsymbol{X_t}\|$ is observed.

The use of (4.1) to infer about the future paths of (X_t) has connections with the so-called *spectral process* introduced by Basrak and Segers (2009) [8] which has opened a fruitful line of research (see for instance [7, 44, 78, 79, 102, 107]). This spectral process is defined as the limit in distribution of a vector of observations of a multivariate regularly varying time series conditionnally on the first observation being large. The approach followed here differs in that it operates at the representation level of α-stable vectors, establishing a link between the spectral representation and the tail conditional distribution of stable linear processes and shedding light on the (un)predictability of their extremes. A natural interpretation of path prediction in terms of pattern identification emerges from Property (4.1) applied to stable linear processes, similar to what Janssen (2017) [78] pointed out in a framework close to that of Basrak and Segers [8]. The results are extended to encompass processes resulting from the linear combination of α-stable moving averages, coined *stable aggregates*, and illustrated on several examples. Contrary to non-aggregated moving averages, which trajectories recurrently feature the same pattern from one extreme episode to another, stable aggregates appear flexible enough to accomodate trajectories exhibiting various patterns through time.

Section 4.2 characterises the representation of general α-stable vectors on semi-norm unit cylinders and shows that Property (4.1) can be restated under this new representation. Focusing first on α-stable moving averages and then on linear combination thereof, Section 4.3 studies under which condition on the process (X_t) the vector $(X_{t-m},\dots,X_{t+h})$ admits a representation on the unit cylinder $C^{\|\cdot\|}_{m+h+1}$. The anticipativeness of (X_t) surprisingly arises as a necessary condition for such a representation to exist. Section 4.4 then exploits Property (4.1) to analyse the tail conditional distribution of general stable aggregates and of some particular processes: the aggregation of anticipative AR(1), the anticipative AR(2) and the anticipative fractionally integrated process. Section 4.5 finally considers a simple bivariate process to illustrate an extension to vector moving averages. New properties emerge in higher dimensions where, in particular, the presence of a non-anticipative component does not rule out the existence of adequate semi-norm representations. Section 4.6 concludes. Proofs are collected in Section 4.7.

4.2 Stable random vectors representation on unit cylinders

This section starts by recalling the characterisation of stable random vectors on the Euclidean unit sphere before exploring the case of unit cylinders relative to semi-norms and reformulating the regularly varying tails property.

Definition 4.2.1 *A random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ is said 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 non-random vector $\boldsymbol{D} \in \mathbb{R}^d$ such that*

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

where $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$ are independent copies of $\boldsymbol{X}$. Moreover, if $\boldsymbol{X}$ is stable, then there exists a constant $\alpha \in (0, 2]$ such that the above holds with $C = (A^\alpha + B^\alpha)^{1/\alpha}$, and $\boldsymbol{X}$ is then called α-stable.

The Gaussian case ($\alpha = 2$) is henceforth excluded. For $0 < \alpha < 2$, the vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ is an α-stable random vector if and only if there exists a unique pair $(\Gamma, \boldsymbol{\mu}^0)$, Γ a finite measure on S_d and $\boldsymbol{\mu}^0$ a non-random vector in $\mathbb{R}^d$, such that,

$$\mathbb{E}\Big[e^{i\langle \boldsymbol{u}, \boldsymbol{X}\rangle}\Big] = \exp\Bigg\{ - \int_{S_d} |\langle \boldsymbol{u}, \boldsymbol{s}\rangle|^\alpha \Big(1 - i\,\mathrm{sign}(\langle \boldsymbol{u}, \boldsymbol{s}\rangle) w(\alpha, \langle \boldsymbol{u}, \boldsymbol{s}\rangle)\Big)\Gamma(d\boldsymbol{s}) + i\,\langle \boldsymbol{u}, \boldsymbol{\mu}^0\rangle\Bigg\}, \quad \forall \boldsymbol{u} \in \mathbb{R}^d, \qquad (4.3)$$

where $\langle \cdot, \cdot \rangle$ denotes the canonical scalar product, $w(\alpha, s) = \mathrm{tg}\left(\frac{\pi\alpha}{2}\right)$, if $\alpha \neq 1$, and $w(1, s) = -\frac{2}{\pi}\ln|s|$ otherwise, for $s \in \mathbb{R}$. The pair $(\Gamma, \boldsymbol{\mu}^0)$ is called the *spectral representation* of 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 [117]), and in that case

$$\mathbb{E}\Big[e^{i\langle \boldsymbol{u}, \boldsymbol{X}\rangle}\Big] = \exp\Bigg\{ - \int_{S_d} |\langle \boldsymbol{u}, \boldsymbol{s}\rangle|^\alpha \Gamma(d\boldsymbol{s})\Bigg\}, \quad \forall \boldsymbol{u} \in \mathbb{R}^d. \qquad (4.4)$$

In the univariate case, (4.3) boils down to

$$\mathbb{E}\Big[e^{iuX}\Big] = \exp\Bigg\{ - \sigma^\alpha |u|^\alpha \Big(1 - i\beta\,\mathrm{sign}(u) w(\alpha, u)\Big) + iu\mu\Bigg\}, \quad \forall u \in \mathbb{R},$$

for some $\sigma > 0$, $\beta \in [-1, 1]$ and $\mu \in \mathbb{R}$. The representations (4.3) and (4.4) of a stable random vector involves integration over all directions of $\mathbb{R}^d$,[2] here parameterised by the unit sphere relative to the Euclidean norm. Proposition 2.3.8 in [117] shows that the unit sphere relative to any norm can be used instead, provided a change of spectral measure and shift vector. We study alternative representations where integration is performed over a unit cylinder relative to a semi-norm. For a given semi-norm, not all stable vectors admit such a representation, which motivates the following definition.

[2] By *direction* of $\mathbb{R}^d$, it is meant 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 $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^d$.

Definition 4.2.2 Let $\|\cdot\|$ be a semi-norm on $\mathbb{R}^d$, $C_d^{\|\cdot\|} := \{\boldsymbol{s} \in \mathbb{R}^d : \ \|\boldsymbol{s}\| = 1\}$ be the corresponding unit cylinder, and let $\boldsymbol{X} = (X_1, \dots, X_d)$ be an α-stable random vector.
(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 $\boldsymbol{u} \in \mathbb{R}^d$

$$\int_{C_d^{\|\cdot\|}} |\langle \boldsymbol{u}, \boldsymbol{s}\rangle|^{\alpha} \Gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty, \tag{4.5}$$

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

$$\int_{C_d^{\|\cdot\|}} |\langle \boldsymbol{u}, \boldsymbol{s}\rangle| \Big| \ln |\langle \boldsymbol{u}, \boldsymbol{s}\rangle| \Big| \Gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty, \tag{4.6}$$

such that the joint characteristic function of $\boldsymbol{X}$ can be written as in (4.3) with $(S_d, \Gamma, \boldsymbol{\mu}^0)$ replaced by $(C_d^{\|\cdot\|}, \Gamma^{\|\cdot\|}, \boldsymbol{\mu}^0_{\|\cdot\|})$.
(Symmetric case) In the case where $\boldsymbol{X}$ is symmetric α-stable (SαS), $0 < \alpha < 2$, 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 (4.5) such that the joint characteristic function of $\boldsymbol{X}$ can be written as in (4.4) with (S_d, Γ) replaced by $(C_d^{\|\cdot\|}, \Gamma^{\|\cdot\|})$.

Remark 4.2.1 As unit cylinders are unbounded sets, the integrability conditions (4.5)-(4.6) ensure the sanity of the above definition.

We start by characterising stable random vectors that are representable on a given semi-norm unit cylinder.

Proposition 4.2.1 *Let $\|\cdot\|$ be a semi-norm 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*

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

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

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

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

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

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

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

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

Remark 4.2.2 The representability condition in the case [$\alpha = 1$ and $\boldsymbol{X}$ not symmetric] is slightly stronger than that in the other cases. Indeed, $\int_{K^{\|\cdot\|}} \Big| \ln \|\boldsymbol{s}\| \Big| \Gamma(d\boldsymbol{s}) \leq \int_{S_d} \Big| \ln \|\boldsymbol{s}\| \Big| \Gamma(d\boldsymbol{s}) < +\infty$ necessarily implies that $\Gamma(K^{\|\cdot\|}) = 0$ since $\Big| \ln \|\boldsymbol{s}\| \Big| = +\infty$ for $\boldsymbol{s} \in K^{\|\cdot\|}$.

Remark 4.2.3 The case $d = 2$ is insightful. In view of (4.1), the spectral measure of the α-stable vector (X_1, X_2) describes its likelihood of being in any particular direction of $\mathbb{R}^2$ when it is large in norm. As unit spheres relative to norms span all the directions of $\mathbb{R}^2$, spectral measures on such spheres can describe any potential tail dependence of (X_1, X_2). Unit cylinders however do not span all directions of $\mathbb{R}^2$ and spectral measures thereon necessarily 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 semi-norm such that $\|(x_1, x_2)\| = |x_1|$ for all $(x_1, x_2) \in \mathbb{R}^2$. It is easy to see that $C_2^{\|\cdot\|}$ spans all directions of $\mathbb{R}^2$ but the ones of $(0, -1)$ and $(0, +1)$. A stable vector (X_1, X_2) will admit a representation on $C_2^{\|\cdot\|}$ provided these directions are irrelevant to characterise its distribution, that is, if $\Gamma\Big(\{(0, -1), (0, +1)\}\Big) = 0$. In terms of tail dependence, the latter condition intuitively means that realisations (X_1, X_2) where X_2 is extreme and X_1 is not almost never occur (i.e., occur with probability zero).[3]

Provided the adequate representation exists, Property (4.1) then holds with semi-norms instead of norms, providing the cornerstone for studying the tail conditional distribution of stable processes.

Proposition 4.2.2 *Let $\boldsymbol{X} = (X_1, \ldots, X_d)$ be an α-stable random vector and let $\|\cdot\|$ be a semi-norm 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 $\Gamma^{\|\cdot\|}\Big(\partial(A \cap B)\Big) = \Gamma^{\|\cdot\|}\big(\partial B\big) = 0$, and $\Gamma^{\|\cdot\|}(B) > 0$,*

$$\mathbb{P}_x^{\|\cdot\|}(\boldsymbol{X}, A|B) \underset{x \to +\infty}{\longrightarrow} \frac{\Gamma^{\|\cdot\|}(A \cap B)}{\Gamma^{\|\cdot\|}(B)}, \tag{4.7}$$

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

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

4.3 Unit cylinder representation for paths of stable linear processes

Given a semi-norm, Proposition 4.2.2 is only applicable to stable vectors that are representable on the corresponding unit cylinder. This section investigates under which condition on an stable moving average (X_t) vectors of the form $(X_{t-m}, \ldots, X_t, X_{t+1}, \ldots, X_{t+h})$ admit such representations. A characterisation is

[3]The conditions $\Gamma\Big(\{(0, -1), (0, +1)\}\Big) = 0$ and $\int_{S_2} \Big| \ln \|\boldsymbol{s}\| \Big| \Gamma(d\boldsymbol{s}) < +\infty$ can also be related to the stronger condition ensuring the existence of conditional moments of X_2 given X_1 obtained in [30, 33] (see also Theorem 5.1.3 in [117]) and which requires Γ not to be too concentrated around the points $(0, \pm 1)$. Namely, 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 $\gamma < \min(\alpha + \nu, 2\alpha + 1)$, despite the fact that $\mathbb{E}[|X_2|^\alpha] = +\infty$. If the previous holds for some $\nu > 0$, then necessarily both of the aforementioned conditions are satisfied.

proposed and is then extended to linear combination of stable moving averages. Any semi-norm satisfying (4.2) could be relevant for the prediction framework mentioned in introduction. However to fix ideas and avoid numerous cases with respect to all the possible kernels, we restrict to semi-norms such that

$$\|(x_{-m},\dots,x_0,x_1,\dots,x_h)\| = 0 \iff x_{-m} = \dots = x_0 = 0, \tag{4.8}$$

for any $(x_{-m},\dots,x_h) \in \mathbb{R}^{m+h+1}$, which in particular satisfy (4.2).

Example 4.3.1 Semi-norms on $\mathbb{R}^{m+h+1}$ satisfying (4.8) can be naturally obtained from norms on the $m+1$ first components of vectors. For any $p \in [1,+\infty]$, one can consider for instance semi-norms $\|\cdot\|$ 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 $\big(\sum_{i=-m}^{0}|x_i|^p\big)^{1/p} = \sup\limits_{-m\le i\le 0} |x_i|$ for $p=+\infty$.

4.3.1 The case of moving averages

Consider (X_t) the α-stable moving average defined by

$$X_t = \sum_{k\in\mathbb{Z}} d_k \varepsilon_{t+k}, \qquad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha,\beta,\sigma,0) \tag{4.9}$$

with (d_k) a real deterministic sequence such that

$$\text{if} \quad \alpha \neq 1 \quad \text{or} \quad (\alpha,\beta) = (1,0), \qquad 0 < \sum_{k\in\mathbb{Z}} |d_k|^s < +\infty, \qquad \text{for some} \quad s \in (0,\alpha)\cap[0,1], \tag{4.10}$$

and

$$\text{if} \quad \alpha = 1 \quad \text{and} \quad \beta \neq 0, \qquad 0 < \sum_{k\in\mathbb{Z}} |d_k| \Big| \ln|d_k| \Big| < +\infty. \tag{4.11}$$

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

$$\boldsymbol{X_t} = (X_{t-m},\dots,X_t,X_{t+1},\dots,X_{t+h}), \tag{4.12}$$

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. The random vector $\boldsymbol{X_t}$ is multivariate α-stable: denoting $\boldsymbol{d}_k := (d_{k+m},\dots,d_k,d_{k-1},\dots,d_{k-h})$ for $k \in \mathbb{Z}$, the spectral representation of $\boldsymbol{X_t}$ on the Euclidean sphere reads $(\Gamma,\boldsymbol{\mu}^0)$ with

$$\Gamma = \sigma^\alpha \sum_{\vartheta\in S_1}\sum_{k\in\mathbb{Z}} w_\vartheta \|\boldsymbol{d}_k\|_e^\alpha \delta_{\left\{\frac{\vartheta \boldsymbol{d}_k}{\|\boldsymbol{d}_k\|_e}\right\}}, \tag{4.13}$$

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

where $w_\vartheta = (1+\vartheta\beta)/2$, $S_1 = \{-1,+1\}$, δ is the dirac mass and by convention, if for some $k\in\mathbb{Z}$, $\boldsymbol{d}_k = \boldsymbol{0}$, i.e. $\|\boldsymbol{d}_k\|_e = 0$, then the kth term vanishes from the sums. 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 next result characterises the representability of $\boldsymbol{X_t}$ on a unit cylinder for fixed m and h.

Lemma 4.3.1 *Let $\boldsymbol{X_t}$ satisfy* (4.9)-(4.12) *and let* $\|\cdot\|$ *be a semi-norm on* $\mathbb{R}^{m+h+1}$ *satisfying* (4.8)*. For* $\alpha \neq 1$ *and* $(\alpha,\beta)=(1,0)$*, the vector* $\boldsymbol{X_t}$ *is representable on* $C^{\|\cdot\|}_{m+h+1}$ *if and only if*

$$\forall k \in \mathbb{Z}, \quad \Big[(d_{k+m},\dots,d_k)=\mathbf{0} \quad \Longrightarrow \quad \forall \ell \leq k-1, \quad d_\ell = 0\Big]. \tag{4.14}$$

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* (4.14)*, it holds that*

$$\sum_{k\in\mathbb{Z}} \|\boldsymbol{d}_k\|_e \left| \ln\Big(\|\boldsymbol{d}_k\|/\|\boldsymbol{d}_k\|_e\Big)\right| < +\infty. \tag{4.15}$$

In the cases $\alpha \neq 1$ and $(\alpha,\beta)=(1,0)$, the representability of $\boldsymbol{X_t}$ on a semi-norm unit cylinder depends on the number of observation $m+1$ but not on the prediction horizon h. Moreover, it is easy to see that if (4.14) is true for some $m \geq 0$, it then holds for any $m' \geq m$. The case $\alpha = 1$, $\beta \neq 0$ is more intricate, the roles of m and h in the validity of the additional requirement (4.15) not being as clear-cut.

A key distinction appears between moving averages according to whether finite length paths admit semi-norm representations. This distinction especially matters for the applicability of Proposition 4.7 when studying the conditional dynamics of a given process. The following definition thus introduces the notion of *past-representability* of a stable moving average.

Definition 4.3.1 Let (X_t) be an α-stable moving average satisfying (4.9)-(4.11). 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 $\boldsymbol{X_t} = (X_{t-m},\dots,X_t,X_{t+1},\dots,X_{t+h})$ is representable on $C^{\|\cdot\|}_{m+h+1}$ for some semi-norm satisfying (4.8). For any such pair (m,h), we will say that (X_t) is (m,h)-past-representable.

Remark 4.3.1 It can be noticed that if $\boldsymbol{X_t} = (X_{t-m},\dots,X_t,X_{t+1},\dots,X_{t+h})$ is representable on $C^{\|\cdot\|}_{m+h+1}$ for some semi-norm satisfying (4.8), then it is representable on unit cylinders relative to any other semi-norms satisfying (4.8). This holds because (4.8) ensures that all these semi-norms have the same kernel. The notion of past-representability can thus be defined independently of the particular choice of a semi-norm.[4]

The following proposition provides a characterisation of past-representability.

Proposition 4.3.1 *Let* (X_t) *be an* α*-stable moving average satisfying* (4.9)-(4.11).
(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*

$$m_0 = \begin{cases} \sup \mathcal{M}, & \textit{if} \quad \mathcal{M} \neq \emptyset, \\ 0, & \textit{if} \quad \mathcal{M} = \emptyset. \end{cases} \tag{4.16}$$

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

$$m_0 < +\infty. \tag{4.17}$$

[4] This will not be true in general under the weaker assumption (4.2) and different notions of representability of a process could emerge depending on the kernels of the semi-norms.

Moreover, letting $m \geq 0$, $h \geq 1$, the process (X_t) is (m,h)-past-representable if and only if (4.17) *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 (4.17), *there exist an $m \geq m_0$ and an $h \geq 1$ such that* (4.15) *holds. If such a pair (m,h) exists,* (X_t) *is then (m,h)-past-representable.*

$(\iota\iota)$ Let $\|\cdot\|$ a semi-norm satisfying (4.8) *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* (4.13) *with the Euclidean norm $\|\cdot\|_e$ replaced by the semi-norm $\|\cdot\|$.*

Remark 4.3.2 Note in particular 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 $k < k_0$.

Remark 4.3.3 Proposition 4.3.1 shows that for an α-stable moving average to be past-representable, sequences of consecutive zero values in the coefficients (d_k) have to be either of finite lengths, or infinite to the left. This surprisingly places the anticipativeness of a stable moving average as a necessary –and sufficient for $\alpha \neq 1$ and $(\alpha, \beta) = (1,0)$– condition for its past-representability. The less anticipative a moving average is, in the sense of the larger the gaps of zeros in its forward-looking side, then the higher m has to be chosen so as to have 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.

Corollary 4.3.1 Let (X_t) an α-stable moving average satisfying (4.9)-(4.11). If (X_t) is purely non-anticipative, i.e., $d_k = 0$ for all $k \geq 1$, then (X_t) is not past-representable.

Remark 4.3.4 This fault line between anticipativeness and non-anticipativeness sheds light on the predictability of extreme events in linear processes. Consider for illustration the two following α-stable AR(1) processes defined as the stationary solutions of

$$X_t = \rho X_{t+1} + \varepsilon_t, \qquad \forall t \in \mathbb{Z}, \tag{4.18}$$

$$Y_t = \rho Y_{t-1} + \eta_t, \qquad \forall t \in \mathbb{Z}, \tag{4.19}$$

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 stems 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) (4.19), a jump does not manifest any early visible sign 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)

(4.18), 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 more informative than their plain unconditional likelihood of occurrence. Building on the «information encoding» interpretation of spectral measures given in Remark 4.2.3, 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 condition for past-representability simplifies for ARMA processes and is equivalent to the autoregressive polynomial having at least one root located inside the unit circle.

Corollary 4.3.2 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 (4.16). 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 $m_0 < +\infty$.

Remark 4.3.5 For ARMA processes, we can notice in particular that the discrepancy between the cases [$\alpha \neq 1$ or $(\alpha, \beta) = (1, 0)$] and [$\alpha = 1$, $\beta \neq 0$] vanishes. Also, only the roots of the AR polynomial matter for past-representability, the MA part having no role.

4.3.2 Aggregation of moving averages

As will be seen in the next section, stable moving averages of the form (4.9) generate trajectories bound to feature the same pattern $t \mapsto cd_{\tau-t}$ (up to a scaling c and a time shift τ) recurrently through time. This can be seen as a strong limitation when it comes to time series modelling as argued by Gouriéroux and Zakoian (2017) [63] in the context of explosive bubbles. They suggest to alleviate this restriction by considering processes resulting from the linear combination of different models. These aggregations feature richer dynamics but little results are available to describe them (see for instance [52] for the aggregation of stable anticipative AR(1)). Linear combinations of stable moving averages will fit naturally into our framework and the results will extend.

Definition 4.3.2 Let $(X_{1,t}), \dots, (X_{J,t})$ be $J \geq 1$ stable moving averages, each satisfying (4.9)-(4.11), for some coefficients sequences $(d_{j,k})_k$ and mutually independent error sequences $\varepsilon_{j,t} \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, \beta_j, 1, 0)$, $j =$

$1,\dots,J$. Let also $(\pi_j)_{j=1,\dots,J}$ be positive numbers and define (X_t) as

$$X_t = \sum_{j=1}^{J} \pi_j X_{j,t}, \qquad \text{for} \quad t \in \mathbb{Z}.$$

We will call such process (X_t) a stable *aggregated moving average*, an *aggregated process*, or simply, a *stable aggregate*, and call $(X_{j,t})$, $j = 1,\dots,J$ the *latent* moving averages of (X_t).

We provide the spectral representation of paths of the aggregated process (X_t) on the Euclidean unit sphere in the next lemma.

Lemma 4.3.2 *Let (X_t) an α-stable aggregate with latent moving averages $(X_{1,t}),\dots,(X_{J,t})$ as in Definition 4.3.2, and let $\boldsymbol{X}_t$ as in* (4.12) *for any $m \geq 0$, $h \geq 1$. Then, $\boldsymbol{X}_t$ is α-stable and its spectral representation $(\Gamma, \boldsymbol{\mu}^0)$ on the Euclidean unit sphere S_{m+h+1} writes*

$$\Gamma = \sum_{j=1}^{J} \sum_{\vartheta \in S_1} \sum_{k \in \mathbb{Z}} w_{j,\vartheta} \pi_j^{\alpha} \|\boldsymbol{d}_{j,k}\|_e^{\alpha} \delta_{\left\{ \frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|_e} \right\}}, \tag{4.20}$$

$$\boldsymbol{\mu}^0 = -\mathbb{1}_{\{\alpha=1\}} \frac{2}{\pi} \sum_{j=1}^{J} \sum_{k \in \mathbb{Z}} \pi_j \beta_j \boldsymbol{d}_{j,k} \ln \|\pi_j \boldsymbol{d}_{j,k}\|_e,$$

where $\boldsymbol{d}_{j,k} = (d_{j,k+m},\dots,d_{j,k},d_{j,k-1},\dots,d_{j,k-h})$, $w_{j,\vartheta} = (1+\vartheta\beta_j)/2$, for any $k \in \mathbb{Z}$, $j = 1,\dots,J$, $\vartheta \in S_1$, and if $\boldsymbol{d}_{j,k} = \boldsymbol{0}$, the term vanishes by convention from the sums.

Remark 4.3.6 Notice that $\Gamma = \sum_{j=1}^{J} \pi_j^{\alpha} \Gamma_j$, where Γ_j denotes the spectral measure of the path $\boldsymbol{X}_{j,t}$ from the moving average $(X_{j,t})$, $j = 1,\dots,J$, which is of the form (4.13).

If all the $\boldsymbol{X}_{j,t}$'s are symmetric ($\beta_j = 0$ for all j), then $\boldsymbol{X}_t$ and Γ are symmetric as well, but the reciprocal however does not hold true. The measure Γ will be symmetric if and only if $\sum_{j=1}^{J} \pi_j^{\alpha} \Big(\Gamma_j(A) - \Gamma_j(-A) \Big) = 0$ for any Borel set $A \subset S_{m+h+1}$. The latter condition is necessary and sufficient for $\boldsymbol{X}_t$ to be symmetric in the case where $\alpha \neq 1$, whereas for $\alpha = 1$, it guarantees that $\boldsymbol{X}_t$ will be symmetric up to an additive shifting, as $\boldsymbol{\mu}^0$ may be non-zero. The symmetry of paths intervenes in the representability conditions provided in the following lemma.

Lemma 4.3.3 *Let (X_t) an α-stable aggregate with latent moving averages $(X_{1,t}),\dots,(X_{J,t})$ as in Definition 4.3.2. Let $m \geq 0$, $h \geq 1$, $\boldsymbol{X}_t$ as in* (4.12), *and $\|\cdot\|$ be a semi-norm on $\mathbb{R}^{m+h+1}$ satisfying* (4.8).
When either $\alpha \neq 1$ or $\boldsymbol{X}_t$ S1S, the vector $\boldsymbol{X}_t$ is representable on $C_{m+h+1}^{\|\cdot\|}$ if and only if (4.14) *holds with m for all sequences $(d_{j,k})_k$, $j = 1,\dots,J$.*
For $\alpha = 1$ and $\boldsymbol{X}_t$ asymmetric, the vector $\boldsymbol{X}_t$ is representable on $C_{m+h+1}^{\|\cdot\|}$ if and only if (4.14) *and* (4.15) *hold with m and h for all sequences $(d_{j,k})_k$, $j = 1,\dots,J$.*

The notion of past-representability in Definition 4.3.1 straightforwardly encompasses the case of stable aggregated processes and the next proposition provides a characterisation. In view of Lemma (4.2.1), the condition for the representability of a path $\boldsymbol{X}_t$ on a unit cylinder changes according to whether it is symmetric or not in the case $\alpha = 1$.

Proposition 4.3.2 *Let (X_t) an α-stable aggregate with latent moving averages $(X_{1,t}),\dots,(X_{J,t})$ as in Definition 4.3.2.*

(ι) Define for $j=1,\dots,J$ the sets $\mathcal{M}_j=\{m\geq 1: \ \exists k\in\mathbb{Z}, \ d_{j,k+m}=\dots=d_{j,k+1}=0, \ d_{j,k}\neq 0\}$, and

$$m_{0,j}=\begin{cases}\sup\mathcal{M}_j, & \text{if} \quad \mathcal{M}_j\neq\emptyset,\\ 0, & \text{if} \quad \mathcal{M}_j=\emptyset.\end{cases} \tag{4.21}$$

(a) For $\alpha\neq 1$, the aggregated process (X_t) is past-representable if and only if $(X_{j,t})$ is past-representable for all $j=1,\dots,J$, i.e.,

$$\sup_{j=1,\dots,J} m_{0,j}<+\infty. \tag{4.22}$$

Moreover, letting $m\geq 0$, $h\geq 1$, (X_t) is (m,h)-past-representable if and only if (4.22) holds and $m\geq \max_{j=1,\dots,J} m_{0,j}$.

(b) For $\alpha=1$, the process (X_t) is past-representable if and only if (4.22) holds and there exists a pair (m,h), $m\geq \max_{j=1,\dots,J} m_{0,j}$, $h\geq 1$ such that either

$$\boldsymbol{X}_t \text{ is S1S}, \qquad \text{or}, \qquad \boldsymbol{X}_t \text{ asymmetric and (4.15) holds for all sequences } (d_{j,k})_k,$$

where $\boldsymbol{X}_t$ generically denotes a vector as in (4.12). If such a pair exists, then the process (X_t) is (m,h)-past-representable.

$(\iota\iota)$ Let $\|\cdot\|$ a semi-norm satisfying (4.8) 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 $\boldsymbol{X_t}=(X_{t-m},\dots,X_t,X_{t+1},\dots,X_{t+h})$ on $C^{\|\cdot\|}_{m+h+1}$ is then given by (4.20) with the Euclidean norm $\|\cdot\|_e$ replaced by the semi-norm $\|\cdot\|$.

Remark 4.3.7 The necessary condition (4.22) extends what was noticed in the case of non-aggregated moving averages, namely, that anticipativeness is a minimal requirement for past-representability. A single non-anticipative latent moving average is enough to render the aggregated process not past-representable, regardless of the other latent components.

Remark 4.3.8 For $\alpha\neq 1$, the past-representability of an aggregated process is equivalent to that of its latent moving averages, but this does not seem to hold in general for $\alpha=1$. In the latter case however, if all the latent moving averages are symmetric, that is, $\beta_1=\dots=\beta_J=0$, then the paths $\boldsymbol{X}_t$ are S1S for any $m\geq 0$, $h\geq 1$ and $(\iota)(b)$ collapses to $(\iota)(a)$.

The representability condition also simplifies in the case of aggregated ARMA processes and requires each latent ARMA process to be anticipative.

Corollary 4.3.3 For any $j=1,\dots,J$, let $(X_{j,t})$ be the ARMA strictly stationary solution of $\psi_j(F)\phi_j(B)X_{j,t}=\Theta_j(F)H_j(B)\varepsilon_{j,t}$, as in Corollary 4.3.2, with mutually independent sequences $\varepsilon_{j,t}\overset{i.i.d.}{\sim}$

$\mathcal{S}(\alpha, \beta_j, 1, 0)$. Define $X_t = \sum_{j=1}^J \pi_j X_{j,t}$ for any positive scalings $(\pi_j)_j$. Then, for any $\alpha \in (0,2)$, $(\beta_1, \dots, \beta_J) \in [-1,1]^J$, the following statements are equivalent:

(ι) (X_t) is past-representable,

$(\iota\iota)$ $\inf_j \deg(\psi_j) \geq 1$,

$(\iota\iota\iota)$ $\sup_j m_{0,j} < +\infty$,

with the $m_{0,j}$'s as in (4.21). Moreover, letting $m \geq 0$, $h \geq 1$, the aggregated process (X_t) is (m,h)-past-representable if and only if for any $j = 1, \dots, J$, $m_{0,j} < +\infty$, and $m \geq \max_j m_{0,j}$.

4.4 Conditional tail distribution of stable aggregates

In this section, we will derive the tail conditional distribution of linear stable processes for which Proposition 4.2.2 will be applicable. The case of a general past-representable stable aggregate is considered as well as particular examples.

To be relevant for the prediction framework, the Borel set B appearing in Proposition 4.2.2 has to 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 realisations $X_{t+1}, \dots, X_{t+h}$. For $\|\cdot\|$ a semi-norm on $\mathbb{R}^{m+h+1}$ satisfying (4.8), denote $S_{m+1}^{\|\cdot\|} = \{(s_{-m}, \dots, s_0) \in \mathbb{R}^{m+1} : \|(s_{-m}, \dots, s_0, 0, \dots, 0)\| = 1\}$.[5] Then, for any Borel set $V \subset S_{m+1}^{\|\cdot\|}$, define the Borel set $B(V) \subset C_{m+h+1}^{\|\cdot\|}$ as

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

Notice in particular that for $V = S_{m+1}^{\|\cdot\|}$, we have $B(V) = C_{m+1}^{\|\cdot\|}$. In the following, we will use Borel sets of the above form to condition the distribution of the complete 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 some small neighbourhood on $S_{m+1}^{\|\cdot\|}$. It will be useful in the following to notice that

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

where f the function defined by

$$f : \begin{array}{ccc} \mathbb{R}^{m+h+1} & \longrightarrow & \mathbb{R}^{m+1} \\ (x_{-m}, \dots, x_0, x_1, \dots, x_h) & \longmapsto & (x_{-m}, \dots, x_0) \end{array}. \tag{4.23}$$

4.4.1 Stable aggregates: general case

Let (X_t) an α-stable aggregate as in Definition 4.3.2 (possibly a moving average if $J = 1$). Assume (X_t) is (m,h)-past-representable, for some $m \geq 0$, $h \geq 1$ and let $\boldsymbol{X_t}$ as in (4.12). Denoting $\Gamma^{\|\cdot\|}$ the spectral measure of $\boldsymbol{X}_t$ on the unit cylinder $C_{m+h+1}^{\|\cdot\|}$ for some semi-norm satisfying (4.8), we know by Proposition 4.3.2 $(\iota\iota)$, that $\Gamma^{\|\cdot\|}$ is of the form

$$\Gamma^{\|\cdot\|} = \sum_{j=1}^J \sum_{\vartheta \in S_1} \sum_{k \in \mathbb{Z}} w_{j,\vartheta} \pi_j^\alpha \|\boldsymbol{d}_{j,k}\|^\alpha \delta_{\left\{ \frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|} \right\}}. \tag{4.24}$$

Proposition 4.4.1

[5] 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.

Under the above assumptions, we have

$$\mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t}, A\Big|B(V)\Big) \underset{x\to+\infty}{\longrightarrow} \frac{\Gamma^{\|\cdot\|}\left(\left\{\dfrac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|} \in A : \ \dfrac{\vartheta f(\boldsymbol{d}_{j,k})}{\|\boldsymbol{d}_{j,k}\|} \in V\right\}\right)}{\Gamma^{\|\cdot\|}\left(\left\{\dfrac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|} \in C^{\|\cdot\|}_{m+h+1} : \ \dfrac{\vartheta f(\boldsymbol{d}_{j,k})}{\|\boldsymbol{d}_{j,k}\|} \in V\right\}\right)}, \tag{4.25}$$

for any Borel sets $A \subset C^{\|\cdot\|}_{m+h+1}$, $V \subset S^{\|\cdot\|}_{m+1}$ *such that* $\left\{\dfrac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|} \in C^{\|\cdot\|}_{m+h+1} : \ \dfrac{\vartheta f(\boldsymbol{d}_{j,k})}{\|\boldsymbol{d}_{j,k}\|} \in V\right\} \neq \emptyset$, $\Gamma^{\|\cdot\|}\Big(\partial(A \cap B(V))\Big) = \Gamma^{\|\cdot\|}(\partial B(V)) = 0$, *where* $B(V) = V \times \mathbb{R}^h$ *and* f *is as in* (4.23).

Remark 4.4.1 (ι) Setting $V = S^{\|\cdot\|}_{m+1}$, and A an arbitrarily small closed neighbourhood of all the points $(\vartheta \boldsymbol{d}_{j,k}/\|\boldsymbol{d}_{j,k}\|)_{\vartheta,j,k}$, 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 terms, when far from central values, the trajectory of process (X_t) necessarily features patterns of the same shape as some $\vartheta \boldsymbol{d}_{j,k}/\|\boldsymbol{d}_{j,k}\|$, which is a finite piece of a moving average's coefficient sequence. The index j indicates from which of the J underlying moving averages the pattern stems from, the index k points to which piece $(d_{j,k+m}, \dots, d_{j,k}, d_{j,k-1}, \dots, d_{j,k-h})$ of this moving average it corresponds, 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 $(\varepsilon_{j,\tau})$). The likelihood of a pattern $\vartheta \boldsymbol{d}_{j,k}/\|\boldsymbol{d}_{j,k}\|$ can be evaluated by setting A to be a small neighbourhood of that point. In particular, only one pattern $\boldsymbol{d}_k/\|\boldsymbol{d}_k\|$ can appear through time for $J = 1$ (up to a time shift and sign flipping). This is no longer the case in general for $J \geq 2$, where the shape of each extreme event appears as if being drawn from a collection of patterns.
($\iota\iota$) In view of point (ι), the observed path $(X_{t-m}, \dots, X_{t-1}, X_t)/\|\boldsymbol{X_t}\|$ will *a fortiori* be of the same shape as some $\vartheta(d_{j,k+m}, \dots, d_{j,k+1}, d_{j,k})/\|\boldsymbol{d}_{j,k}\|$ when an extreme event will approach in time. Observing the initial part of the pattern can give information about the remaining unobserved piece: the conditional likelihood of the latter can be assessed by setting V to be a small neighbourhood of the observed pattern.

Remark 4.4.2 The tail conditional distribution given in (4.25) highlights three types of uncertainty/approximation for prediction:[6]
(ι) In practice, events of the type $\{(X_{t-m}, \dots, X_{t-1}, X_t)/\|\boldsymbol{X_t}\| = \vartheta(d_{j,k+m}, \dots, d_{j,k+1}, d_{j,k})/\|\boldsymbol{d}_{j,k}\|\}$ have probability zero of occurring, and only noisy observations such as $(X_{t-m}, \dots, X_{t-1}, X_t)/\|\boldsymbol{X_t}\| \approx \vartheta(d_{j,k+m}, \dots, d_{j,k+1}, d_{j,k})/\|\boldsymbol{d}_{j,k}\|$ are available on a realised trajectory. The choice of an adequate conditioning neighbourhood V in (4.25) given a piece of trajectory will thus have to rely on a statistical approach. One could envision tests of hypotheses to determine whether a piece of realised (noisy) trajectory «is more similar» to a certain pattern 1 or to an other pattern 2, or whether it «is more similar» to a certain pattern 1 rather than any patterns in a certain collection.
($\iota\iota$) Even for an arbitrarily small neighbourhood V –that is, even if the observed path can be confidently identified with a particular pattern– uncertainty regarding the future trajectory may remain. It could indeed

[6] 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 $(d_{j,k})$ has to be inferred from data.

be that several patterns $\vartheta \boldsymbol{d}_{j,k}/\|\boldsymbol{d}_{j,k}\|$ coincide on their first $m+1$ components, but differ by the last h. The stable anticipative AR(1) and its aggregated version are typical examples of this phenomenon that will be studied in the next section.

$(\iota\iota\iota)$ The tail conditional distribution (4.25) is an asymptotic behaviour as the (semi-)norm of $\boldsymbol{X_t}$ grows infinitely large. It is thus only an approximation of the true dynamics during extreme events. It would be interesting to obtain a finer asymptotic development in x of the above convergence to gauge the approximation error of the true conditional distribution. It would be especially useful to quantify how far from/how variable around the predicted patterns the future path can be.

4.4.2 Aggregation of AR(1)

We now consider (X_t) the aggregation of stable anticipative AR(1) processes introduced in [63] defined by

$$X_t = \sum_{j=1}^{J} \pi_j X_{j,t}, \quad X_{j,t} = \rho_j X_{j,t+1} + \varepsilon_{j,t}, \quad 0 < |\rho_j| < 1, \quad j = 1, \ldots, J \tag{4.26}$$

where $\pi_j > 0$ for any j, and $(\varepsilon_{j,t})_{t\in\mathbb{Z}} \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, \beta_j, 1, 0)$ are mutually independent i.i.d. sequences. We assume without loss of generality that the ρ_j's are distinct. For each anticipative AR(1) with parameter ρ_j, the moving average coefficients are of the form $(\rho_j^k \mathbb{1}_{\{k\geq 0\}})_k$, and thus, $m_{0,j} = 0$ for all j, where the $m_{0,j}$'s are given in (4.21). By Corollary (4.3.3), we know for any $m \geq 0$, $h \geq 1$, the aggregated process (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 next lemma.

Lemma 4.4.1 *Let (X_t) be an aggregation of α-stable anticipative AR(1) processes as in* (4.26). *Letting $\boldsymbol{X_t}$ as in* (4.12) *for $m \geq 0$, $h \geq 1$, its spectral measure on $C_{m+h+1}^{\|\cdot\|}$ for a semi-norm satisfying* (4.8) *is given by*

$$\Gamma^{\|\cdot\|} = \sum_{\vartheta\in S_1} \left[w_\vartheta \delta_{\{(\vartheta,0,\ldots,0)\}} + \sum_{j=1}^{J} \pi_j^\alpha \left(w_{j,\vartheta} \sum_{k=-m+1}^{h-1} \|\boldsymbol{d}_{j,k}\|^\alpha \delta_{\left\{\frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|}\right\}} + \frac{\bar{w}_{j,\vartheta}}{1-|\rho_j|^\alpha} \|\boldsymbol{d}_{j,h}\|^\alpha \delta_{\left\{\frac{\vartheta\, \boldsymbol{d}_{j,h}}{\|\boldsymbol{d}_{j,h}\|}\right\}} \right) \right], \tag{4.27}$$

where for all $\vartheta \in S_1$, $j \in \{1, \ldots, J\}$ and $-m+1 \leq k \leq h$,

$$\begin{aligned}
\boldsymbol{d}_{j,k} &= (\rho_j^{k+m} \mathbb{1}_{\{k\geq -m\}}, \ldots, \rho_j^k \mathbb{1}_{\{k\geq 0\}}, \rho_j^{k-1} \mathbb{1}_{\{k\geq 1\}}, \ldots, \rho_j^{k-h} \mathbb{1}_{\{k\geq h\}}),\\
w_{j,\vartheta} &= (1+\vartheta\beta_j)/2,\\
w_\vartheta &= \sum_{j=1}^{J} \pi_j^\alpha w_{j,\vartheta},\\
\bar{w}_{j,\vartheta} &= (1+\vartheta\bar{\beta}_j)/2,\\
\bar{\beta}_j &= \beta_j \frac{1-\rho_j^{<\alpha>}}{1-|\rho_j|^\alpha},
\end{aligned}$$

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 the ρ_j's are positive. Let us first introduce useful neighbourhoods of the distinct charged points of $\Gamma^{\|\cdot\|}$. Denote $\boldsymbol{d}_{0,-m} = (\overbrace{1,0,\ldots,0}^{m+h+1})$ so that the charged points of $\Gamma^{\|\cdot\|}$ are all of the form $\vartheta \boldsymbol{d}_{j,k}/\|\boldsymbol{d}_{j,k}\|$ with indexes (ϑ, j, k) in the set $\mathcal{I} := S_1 \times \Big(\{1,\ldots,J\} \times \{-m, h\} \cup \{(0,-m)\}\Big)$. With f as in (4.23), define for any $(\vartheta_0, j_0, k_0) \in \mathcal{I}$, the set V_0 as any closed neighbourhood of $\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})/\|\boldsymbol{d}_{j_0,k_0}\|$ such that

$$\forall(\vartheta', j', k') \in \mathcal{I}, \qquad \frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} \in V_0 \implies \frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})}{\|\boldsymbol{d}_{j_0,k_0}\|}, \tag{4.28}$$

In other terms, $V_0 \times \mathbb{R}^d$ is a subset of $C^{\|\cdot\|}_{m+h+1}$ in which the only points charged by $\Gamma^{\|\cdot\|}$ all have the first $(m+1)^{\text{th}}$ coinciding with $\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})/\|\boldsymbol{d}_{j_0,k_0}\|$. Define also $A_{\vartheta,j,k}$ for any (ϑ, j, k) as any closed neighbourhood of $\vartheta \boldsymbol{d}_{j,k}/\|\boldsymbol{d}_{j,k}\|$ which does not contain any other charged point of $\Gamma^{\|\cdot\|}$, that is,

$$\forall(\vartheta', j', k') \in \mathcal{I}, \qquad \frac{\vartheta' \boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|} \in A_{\vartheta,j,k} \implies (\vartheta', j', k') = (\vartheta, j, k). \tag{4.29}$$

Proposition 4.4.2 *Let (X_t) be an aggregation of α-stable anticipative AR(1) processes as in (4.26) with $\rho_j \in (0,1)$ for all j's. Let $\boldsymbol{X_t}$, the $\boldsymbol{d}_{j,k}$'s and the spectral measure of $\boldsymbol{X_t}$ be as given in Lemma 4.4.1, for any $m \geq 0$, $h \geq 1$. Let V_0 be any small closed neighbourhood of $\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})/\|\boldsymbol{d}_{j_0,k_0}\|$ in the sense of (4.28) for some $(\vartheta_0, j_0, k_0) \in \mathcal{I}$ and let $B(V_0) = V_0 \times \mathbb{R}^h$. Then, with $A_{\vartheta,j,k}$ an arbitrarily small neighbourhood of some $\vartheta \boldsymbol{d}_{j,k}/\|\boldsymbol{d}_{j,k}\|$ as in (4.29), the following hold.*

(ι) Case $m \geq 1$.

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

$$\mathbb{P}^{\|\cdot\|}_x\Big(\boldsymbol{X_t}, A_{\vartheta,j,k}\Big| B(V_0)\Big) \underset{x\to\infty}{\longrightarrow} \begin{cases} |\rho_{j_0}|^{\alpha k}(1-|\rho_{j_0}|^{\alpha})\delta_{\vartheta_0}(\vartheta)\delta_{j_0}(j), & 0 \leq k \leq h-1, \\[2ex] |\rho_{j_0}|^{\alpha h}\delta_{\vartheta_0}(\vartheta)\delta_{j_0}(j), & k = h. \end{cases}$$

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

$$\mathbb{P}^{\|\cdot\|}_x\Big(\boldsymbol{X_t}, A_{\vartheta,j,k}\Big| B(V_0)\Big) \underset{x\to\infty}{\longrightarrow} \delta_{\vartheta_0}(\vartheta)\delta_{j_0}(j)\delta_{k_0}(k).$$

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

$$\mathbb{P}^{\|\cdot\|}_x\Big(\boldsymbol{X_t}, A_{\vartheta,j,k}\Big| B(V_0)\Big) \underset{x\to\infty}{\longrightarrow} \begin{cases} \dfrac{\sum_{i=1}^J \pi_i^{\alpha} w_{i,\vartheta_0}}{\sum_{i=1}^J p_{i,\vartheta_0}}\delta_{\{\vartheta_0\}}(\vartheta), & k = 0 \\[3ex] \dfrac{p_{j,\vartheta_0}}{\sum_{i=1}^J p_{i,\vartheta_0}}|\rho_j|^{\alpha k}(1-|\rho_j|^{\alpha})\delta_{\{\vartheta_0\}}(\vartheta), & 1 \leq k \leq h-1, \\[3ex] \dfrac{p_{j,\vartheta_0}}{\sum_{i=1}^J p_{i,\vartheta_0}}|\rho_j|^{\alpha h}\delta_{\{\vartheta_0\}}(\vartheta), & k = h, \end{cases}$$

with $p_{j,\vartheta_0} = \pi_j^{\alpha} w_{j,\vartheta_0}/(1-|\rho_j|^{\alpha})$.

Remark 4.4.3 For $m \geq 1$, that is, if the observed path is assumed to be of length at least 2, there is a significant difference between whether $k_0 \in \{0, \dots, h\}$ or $k_0 \in \{-m, \dots, -1\}$. For the latter, the asymptotic probability of the whole path $\boldsymbol{X_t}/\|\boldsymbol{X_t}\|$ being in an arbitrarily small neighbourhood of $\vartheta \boldsymbol{d}_{j,k}/\|\boldsymbol{d}_{j,k}\|$ is 1 if and only if $\vartheta = \vartheta_0$, $j = j_0$, $k = k_0$: given the observed path, the shape of the future trajectory is fully determined. For the former, this probability is strictly positive if and only if $\vartheta = \vartheta_0$ and $j = j_0$, but the observed pattern is compatible with several distinct future paths. One can see why this is the case from the form of the sequences $\boldsymbol{d}_{j,k}/\|\boldsymbol{d}_{j,k}\|$ and of their restrictions to the first $m+1$ components $f(\boldsymbol{d}_{j,k})/\|\boldsymbol{d}_{j,k}\|$. On the one hand (omitting ϑ),

$$\frac{\boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|} = \begin{cases} \dfrac{(\overbrace{\rho_j^{k+m}, \dots, \rho_j^k}^{m+1}, \overbrace{\rho_j^{k-1}, \dots, \rho_j, 1, 0, \dots, 0}^{h})}{\|(\rho_j^{k+m}, \dots, \rho_j^k, \rho_j^{k-1}, \dots, \rho_j, 1, 0, \dots, 0)\|}, & \text{for} \qquad k \in \{0, \dots, h\}, \\[2em] \dfrac{(\rho_j^{k+m}, \dots, \rho_j, 1, 0, \dots, 0, 0, \dots, 0)}{\|(\underbrace{\rho_j^{k+m}, \dots, \rho_j, 1, 0, \dots, 0}_{m+1}, \underbrace{0, \dots, 0}_{h})\|}, & \text{for} \qquad k \in \{-m, \dots, -1\}. \end{cases}$$

We can notice that all the above sequences are pieces of explosive exponentials, terminated at some coordinate. For $k \in \{0, \dots, h\}$, the first zero component –the «crash of the bubble»–, is situated at or after the $(m+2)^{\text{th}}$ component, whereas for $k \in \{-m, \dots, -1\}$, it is situated at or before the $(m+1)^{\text{th}}$. Using the homogeneity of the semi-norm and (4.2), we have on the other hand that

$$\frac{f(\boldsymbol{d}_{j,k})}{\|\boldsymbol{d}_{j,k}\|} = \begin{cases} \dfrac{(\overbrace{\rho_j^{m}, \dots, \rho_j, 1}^{m+1})}{\|(\underbrace{\rho_j^{m}, \dots, \rho_j, 1}_{m+1}, \underbrace{0, \dots, 0, 0, \dots, 0}_{h})\|}, & \text{for} \qquad k \in \{0, \dots, h\}, \\[2em] \dfrac{(\overbrace{\rho_j^{k+m}, \dots, \rho_j, 1, 0, \dots, 0}^{m+1})}{\|(\underbrace{\rho_j^{k+m}, \dots, \rho_j, 1, 0, \dots, 0}_{m+1}, \underbrace{0, \dots, 0}_{h})\|}, & \text{for} \qquad k \in \{-m, \dots, -1\}. \end{cases}$$

Thus, conditioning the trajectory on the event $\{f(\boldsymbol{X_t})/\|\boldsymbol{X_t}\| \approx f(\boldsymbol{d}_{j_0,k_0})/\|\boldsymbol{d}_{j_0,k_0}\|\}$ for some $k_0 \in \{-m, \dots, -1\}$ amounts to condition on the burst of a bubble being observed in the past trajectory with no new bubble forming yet, which allows to identify exactly the position of the pattern on the j^{th} moving average's coefficient sequence.

When conditioning with $k_0 \in \{0, \dots, h\}$ however, the crash date is not observed and can happen either in the next $h-1$ periods, or after the h^{th}. However, the shape of the observed path is that of a piece of exponential with growth rate ρ_j^{-1} regardless of the remaining time before the burst, which leaves several future paths possible. One can quantify the likelihood of each potential scenario: the quantity $|\rho_j|^{\alpha k}(1 - |\rho_j|^{\alpha})$ corresponds to the probability that the bubble will peak in exactly k periods ($0 \leq k < h$), and $|\rho_j|^{\alpha h}$ corresponds to the probability that the bubble will last at least h more periods.

Remark 4.4.4 (ι) The previous remark confirms the interpretation of the conditional moments proposed in [52] for the stable anticipative AR(1) case ($J = 1$). It also extends it in two ways: (ι) by accounting for paths rather than point prediction, ($\iota\iota$) by showing that the aggregation of AR(1) processes also features killed exponential explosive episodes but with various growth rates and crash probabilities. Proposition 4.4.2 furthermore shows that asymptotically, as few as two observations are sufficient to identify the growth rate ρ_j^{-1} of an ongoing extreme episode,[7] and the conditional dynamics within this given event will be similar to that of a simple AR(1) with corresponding parameter. An identification of the growth rate in the early developments of the bubble appears possible, allowing to infer in advance the odds of crashes.
($\iota\iota$) Notice that for $m = 0$ (only the present value is assumed to be observed), no pattern can be observed but only the sign of the shock. Hence, the growth rate $\rho_{j_0}^{-1}$ of the ongoing event is unidentifiable, which is reflected in the fact that the asymptotic probabilities of paths with growth rates ρ_j^{-1}, $j \neq j_0$, are positive (case ($\iota\iota$) of Proposition 4.4.2).

4.4.3 Two examples: the anticipative AR(2) and fractionally integrated AR

We focus here on two processes which both share the peculiar property of having a 0-1 tail conditional distribution whenever the observed path is of length at least 2 (i.e., $m \geq 1$): the anticipative AR(2) and the anticipative fractionally integrated AR. For an adequate choice of the parameters, the former can generate bubble-like trajectories with accelerating or decelerating growth rate and the latter can accommodate hyperbolic bubbles. In contrast with the anticipative AR(1), these bubbles do not display an exponential profile but still feature an inflation-peak-collapse behaviour. The two processes are defined as follows.

Anticipative AR(2)

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

$$(1 - \lambda_1 F)(1 - \lambda_2 F)X_t = \varepsilon_t, \qquad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, \beta, \sigma, 0), \tag{4.30}$$

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 (4.30) admits the moving average representation $X_t = \sum_{k \in \mathbb{Z}} d_k \varepsilon_{t+k}$ with

$$d_k = \begin{cases} \dfrac{\lambda_1^{k+1} - \lambda_2^{k+1}}{\lambda_1 - \lambda_2} \, \mathbb{1}_{\{k \geq 0\}}, & \text{if} \quad \lambda_1 \neq \lambda_2, \\ (k+1)\lambda^k \, \mathbb{1}_{\{k \geq 0\}}, & \text{if} \quad \lambda_1 = \lambda_2 = \lambda. \end{cases} \tag{4.31}$$

Anticipative fractionally integrated AR

[7] This holds asymptotically in the (semi-)norm of the observed path, but in practice it can be expected that the noise surrounding the trajectory will make this identification difficult with only two observations. Longer path lengths (higher m) may provide robustness to the identification, but could also incorporate some bias by taking into account past extreme events, such as now-collapsed bubbles. One can suspect a bias-variance trade-off when searching for an optimal choice of m.

The anticipative fractionally integrated AR process can be defined as the stationary solution of

$$(1-F)^d X_t = \varepsilon_t, \qquad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha,\beta,\sigma,0), \tag{4.32}$$

with $\alpha(d-1) < -1$. The solution of (4.32) admits the moving average representation $X_t = \sum_{k=0}^{+\infty} d_k \varepsilon_{t+k}$ with

$$d_0 = 1, \qquad \text{and} \qquad d_k = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)} \mathbb{1}_{\{k\geq 0\}}, \quad \text{for} \quad k \neq 0, \tag{4.33}$$

where $\Gamma(\,\cdot\,)$ denotes –here only– the Gamma function.

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

Proposition 4.4.3 *Let (X_t) be the α-stable anticipative AR(2) (resp. fractionally integrated AR) as in (4.30)-(4.31) (resp. (4.32)-(4.33)). For any $m \geq 1$ and $h \geq 1$, let $\boldsymbol{X_t}$ as in (4.12) and $\boldsymbol{d}_k = (d_{k+m}, \dots, d_k, d_{k-1}, \dots, d_{k-h})$ where (d_k) is as in (4.31) (resp. (4.33)). Let V_0 a small neighbourhood of $\vartheta_0 \boldsymbol{d}_{k_0}/\|\boldsymbol{d}_{k_0}\|$ as in (4.28) –where we drop the indexes j– for some $\vartheta_0 \in S_1$, $k_0 \geq -m$, and let $B(V_0) = V_0 \times \mathbb{R}^h$. 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 neighbourhood $A \subset C_{m+h+1}^{\|\cdot\|}$ such that $\partial A \cap \{\vartheta \boldsymbol{d}_k/\|\boldsymbol{d}_k\| : \ \vartheta \in S_1, \ k \geq -m\} = \emptyset$.

Remark 4.4.5 Contrary to the anticipative AR(1), the trajectories of the anticipative AR(2) and fractionally integrated processes do not leave room for undeterminancy of the future path. Asymptotically, given any observed path of length at least 2, the shape of the future trajectory can be deduced deterministically. This holds even if the peak/collapse of a bubble is not yet present in the observed piece of trajectory. Therefore, provided the current pattern is properly identified,[8] it appears possible in the framework of these models to infer in advance the peak and crash dates of bubbles with very high confidence –in principle, with certainty.

4.5 A step towards multivariate processes

A simple bi-dimensional process is considered in this section to highlight that the approach developed in this paper can be brought to the multivariate framework and that new properties can also emerge. In essence, the process considered is a vector where each univariate component consists respectively of a stable anticipative AR(1) and a stable non-anticipative AR(1), and dependence between both is allowed. Surprisingly, the

[8] See point (ι) of Remark 4.4.2.

presence of a non-anticipative component will not be pathological here contrary to the univariate case studied above, and Proposition 4.2.2 will be applicable. Formally, define $(\boldsymbol{X}_t)$ for all $t \in \mathbb{Z}$ as

$$\begin{cases} \boldsymbol{X}_t &= (X_{1,t}, X_{2,t})', \\ X_{1,t} &= \rho_1 X_{1,t+1} + \varepsilon_{1,t}, \\ X_{2,t} &= \rho_2 X_{2,t-1} + \varepsilon_{2,t}, \\ \boldsymbol{\varepsilon}_t &= (\varepsilon_{1,t}, \varepsilon_{2,t})' \text{ i.i.d. } S\alpha S \text{ with spectral measure } \Gamma_2 \text{ on } S_2 \text{ and zero shift vector,} \end{cases} \tag{4.34}$$

where $0 < |\rho_i| < 1$, $i = 1, 2$.[9] We again have in mind applying Proposition 4.2.2 to a vector composed of past and future realisations of $(\boldsymbol{X}_t)$. Limiting ourselves to the simplest $m = 0$ and $h = 1$ case, we will consider a vector of the form $\underline{\boldsymbol{X}}_t := (\boldsymbol{X}_t', \boldsymbol{X}_{t+1}')'$, where $\boldsymbol{X}_t$ is the present observation and $\boldsymbol{X}_{t+1}$ the one-step ahead realisation to predict. The next result shows that $\underline{\boldsymbol{X}}_t$ is α-stable, in fact $S\alpha S$, and it provides a necessary and sufficient condition on Γ_2 for its representability on an appropriate unit cylinder.[10]

Proposition 4.5.1 *Let $(\boldsymbol{X}_t)$ as in (4.34), the semi-norm $\|\cdot\|$ on $\mathbb{R}^4$ such that $\|(x_1, x_2, x_3, x_4)\| = \sqrt{x_1^2 + x_2^2}$ for any $(x_1, x_2, x_3, x_4) \in \mathbb{R}^4$, and denote $C_4^{\|\cdot\|}$ its corresponding unit cylinder. The vector $\underline{\boldsymbol{X}}_t$ is then $S\alpha S$, and it is representable on $C_4^{\|\cdot\|}$ if and only if*

$$\Gamma_2\Big(\{(0, -1), (0, +1)\}\Big) = 0. \tag{4.35}$$

The representability condition (4.35) appears in sharp contrast with Remark 4.3.4 and is also reminiscent of Remark 4.2.3. It intuitively means that the joint vector $\underline{\boldsymbol{X}}_t$ will admit a representation on the unit cylinder provided realisations $(\varepsilon_{1,t}, \varepsilon_{2,t})$ where $\varepsilon_{2,t}$ is extreme and $\varepsilon_{1,t}$ is not occur with probability zero. If this holds, then, intuitively, every jump in the trajectory of $(X_{2,t})$ necessarily coincides with a bubble peak in the trajectory of $(X_{1,t})$, and each incoming jump in the former is thus betrayed by the early build-up of a bubble in the latter.[11] When considered univariately, $(X_{2,t})$ features sudden, unpredictable bursts –and is thus not-past-representable–, but this unpredictability appears to fade away when $(X_{2,t})$ is considered jointly with the «informative» process $(X_{1,t})$. The next proposition provides the tail conditional distribution of $\boldsymbol{X}_{t+1}$ given (a large in norm) observation $\boldsymbol{X}_t$, and shows that these heuristics are essentially correct. The anticipative component does inform about incoming jumps in the other component, and, quite surprisingly, the non-anticipative component also brings information about the anticipative one. For expository purposes, we distinguish several cases according to the conditioning event. The proposition is followed by a detailed interpretation of each case.

[9] The $S\alpha S$ assumption on the i.i.d. sequence $(\boldsymbol{\varepsilon}_t)$ is made for the sake of simplicity and implies that Γ_2 is itself symmetric.

[10] For expository purposes, the form of the spectral representations on the Euclidean unit sphere and on the unit cylinder are relegated in the proofs in Appendix.

[11] Note that extreme realisations of $\varepsilon_{1,t}$ may nevertheless occur alongside non-extreme realisations of $\varepsilon_{2,t}$, as $\Gamma_2\Big(\{(-1, 0), (+1, 0)\}\Big)$ can a priori be positive. Thus, intuitively, a bubble peak may be reached in $(X_{1,t})$ with no jump occurring in $(X_{2,t})$.

Proposition 4.5.2 *Let $(\boldsymbol{X}_t)$ as in (4.34) and assume that (4.35) holds. For $\eta_0 > 0$ and $\theta_0 \in]-\pi, \pi]$, define $V_0 = \{(\cos u, \sin u) \in S_2 : \ \ u \in [\theta_0 - \eta_0, \theta_0 + \eta_0]\}$ and let $B(V_0) = V_0 \times \mathbb{R}^2$. Define also, for $\theta \in] - \pi, \pi]$, $\eta > 0$, and P any closed set of $\mathbb{R}^2$,*[12]

$$A_{\theta,\eta,P} = \Big\{(\cos u, \sin u, 0, \rho_2 \sin u) + (0,0,x,y) \in C_4^{\|\cdot\|} : \ \ u \in [\theta-\eta, \theta+\eta] \text{ and } (x,y) \in P\Big\},$$

and $V_{\theta,\eta} = \{(\cos u, \sin u) \in S_2 : \ \ u \in [\theta - \eta, \theta + \eta]\}$.

(ι) Assume that $V_0 \cap \{(\pm 1, 0), (0, \pm 1)\} = \emptyset$. Then,

$$\mathbb{P}_x^{\|\cdot\|}\Big(\underline{\boldsymbol{X}}_t, A_{\theta,\eta,P}\Big|B(V_0)\Big) \underset{x\to+\infty}{\longrightarrow} \frac{\Gamma_2\Big(V_{\theta,\eta} \cap V_0\Big)}{\Gamma_2\Big(V_0\Big)} \delta_{\{(0,0)\}}(P).$$

$(\iota\iota)$ Assume $(0, \vartheta) \in V_0$, for some $\vartheta \in \{-1, +1\}$, and $V_0 \cap \{\pm(1,0), (0, -\vartheta)\} = \emptyset$. Then,

$$\mathbb{P}_x^{\|\cdot\|}\Big(\underline{\boldsymbol{X}}_t, A_{\theta,\eta,P}\Big|B(V_0)\Big) \underset{x\to+\infty}{\longrightarrow} \frac{\dfrac{\sigma_2^\alpha}{2}\dfrac{|\rho_2|^\alpha}{1-|\rho_2|^\alpha}\delta_{\{(0,\vartheta)\}}(V_{\theta,\eta}) + \Gamma_2\Big(V_{\theta,\eta}\cap V_0\Big)}{\dfrac{\sigma_2^\alpha}{2}\dfrac{|\rho_2|^\alpha}{1-|\rho_2|^\alpha} + \Gamma_2\Big(V_0\Big)} \delta_{\{(0,0)\}}(P).$$

$(\iota\iota\iota)$ Assume $(\vartheta, 0) \in V_0$, for some $\vartheta \in \{-1, +1\}$, and $V_0 \cap \{\pm(0,1), (-\vartheta, 0)\} = \emptyset$. Then,

$$\mathbb{P}_x^{\|\cdot\|}\Big(\underline{\boldsymbol{X}}_t, A_{\theta,\eta,P}\Big|B(V_0)\Big)$$

$$\underset{x\to+\infty}{\longrightarrow} \frac{\left(\dfrac{\sigma_1^\alpha}{2}\dfrac{|\rho_1|^{2\alpha}}{1-|\rho_1|^\alpha}\delta_{\{0\}}(P_2) + \dfrac{|\rho_1|^\alpha}{2}\sigma_{1_{|P_2}}^{\ \alpha}\right)\delta_{\{(\vartheta,0)\}}(V_{\theta,\eta})\delta_{\{\vartheta\rho_1^{-1}\}}(P_1) + \Gamma_2\Big(V_{\theta,\eta}\cap V_0\Big)\delta_{\{(0,0)\}}(P)}{\dfrac{\sigma_1^\alpha}{2}\dfrac{|\rho_1|^\alpha}{1-|\rho_1|^\alpha} + \Gamma_2(V_0)},$$

where $P_1 = \{x : \ (x,y) \in P\}$, $P_2 = \{y : \ (x,y) \in P\}$ and $\sigma_{1_{|P_2}} := \left(\displaystyle\int_{S(P_2)} |s_1|^\alpha \Gamma_2(d\boldsymbol{s})\right)^{1/\alpha}$ with $S(P_2) := \left\{\frac{(\rho_1^{-1}, y)}{\sqrt{\rho_1^{-2}+y^2}} \in S_2 : \ \ y \in P_2\right\}$.

Interpretation of Proposition 4.5.2

In the spirit of this proposition, V_0 is typically a small neighbourhood on the unit sphere S_2 accounting for the observed realisation of $(X_{1,t}, X_{2,t})/\sqrt{X_{1,t}^2 + X_{2,t}^2}$, that is, the relative magnitudes of $X_{1,t}$ and $X_{2,t}$.[13] The smaller the neighoubourhood V_0, the more «accurately» we assume to observe these relative magnitudes. The proposition considers three main scenarii: case (ι) $X_{1,t}$ and $X_{2,t}$ are of comparable magnitudes, case $(\iota\iota)$ $X_{2,t}$ is much larger -possibly infinitely larger- than $X_{1,t}$, and case $(\iota\iota\iota)$ $X_{1,t}$ is much larger -possibly infinitely larger- than $X_{2,t}$. Each of these three conditioning leads to different odds regarding

[12] This ensures that $A_{\theta,\eta,P} = \{(0,0)\} \times P \quad + \quad \{(\cos u, \sin u, 0, \rho_2 \sin u) : \ \ u \in [\theta-\eta, \theta+\eta]\}$ defines a proper Borel set, which could fail if P was a general Borel set [36]. One could assume more generally that P is an F_σ set, but a closed set will be enough for our purpose here.

[13] Recall that the results are always conditional on $\sqrt{X_{1,t}^2 + X_{2,t}^2}$ being large: either $X_{1,t}$ is extreme, either $X_{2,t}$ is extreme, or both are.

the potential outcomes at $t+1$.

Case (ι) To fix ideas, let us assume that $X_{1,t}$ and $X_{2,t}$ are observed to be of same signs and approximately of equal magnitudes, that is, V_0 is a small neighbourhood of $c(1,1) \in S_2$, with $c = 2/\sqrt{2}$ (i.e., $\theta_0 = \pi/4$ and $\eta_0 > 0$ small). Now, evaluating the tail conditional probability at $A_{\theta_0,\eta_0,P}$ for P an arbitrarily small closed neighbourhood of $(0,0)$, for instance $P = [-\epsilon_1, \epsilon_1] \times [-\epsilon_2, \epsilon_2]$ for $\epsilon_1, \epsilon_2 > 0$ small, we obtain that

$$\mathbb{P}_x^{\|\cdot\|}\Big(\underline{\boldsymbol{X}}_t, A_{\theta_0,\eta_0,P}\Big|B(V_0)\Big) \underset{x\to+\infty}{\longrightarrow} 1.$$

Intuitively during an extreme event, if $X_{1,t}$ and $X_{2,t}$ are observed to be of approximately equal magnitudes, then the vector $(X_{1,t}, X_{2,t}, X_{1,t+1}, X_{2,t+1})/\sqrt{X_{1,t}^2 + X_{2,t}^2}$ will belong with certainty to a small neighbourhood of $c(1,1,0,\rho_2)$.[14] This straightforwardly extends to the case when $X_{1,t}$ and $X_{2,t}$ are of comparable magnitudes but not necessarily equal ones, i.e., when V_0 is instead a small neighbourhood of $(\cos\theta_0, \sin\theta_0)$. Then, $(X_{1,t}, X_{2,t}, X_{1,t+1}, X_{2,t+1})/\sqrt{X_{1,t}^2 + X_{2,t}^2}$ will belong with certainty to a small neighbourhood of $(\cos\theta_0, \sin\theta_0, 0, \rho_2\sin\theta_0)$.
This reveals that if at any date in time both series are simultaneously extreme, then, with certainty, the anticipative component will collapse at the immediately following date, and the non-anticipative component will decay by ρ_2.

Case ($\iota\iota$) Let us assume here that V_0 is an arbitrarily small neighbourhood of $(0,1)$, i.e., $X_{2,t}$ is observed positive and much larger in magnitude than $X_{1,t}$. Evaluating the conditional probability at $A_{\theta_0,\eta_0,P}$ with $P = [-\epsilon_1, \epsilon_1] \times [-\epsilon_2, \epsilon_2]$ an arbitrarily small neighbourhoods of $(0,0)$, we have that

$$\mathbb{P}_x^{\|\cdot\|}\Big(\underline{\boldsymbol{X}}_t, A_{\theta_0,\eta_0,P}\Big|B(V_0)\Big) \underset{x\to+\infty}{\longrightarrow} 1.$$

Thus, if during an extreme event, $X_{2,t}$ is observed to be much larger than $X_{1,t}$, then the vector $\underline{\boldsymbol{X}}_t/\|\underline{\boldsymbol{X}}_t\|$ will belong with certainty to a small neighbourhood of $(0,1,0,\rho_2)$. Observing at date t an extreme in the non-anticipative component alongside a much smaller, possibly non-extreme value on the anticipative series indicates that at $t+1$, with certainty, the non-anticipative component will decay by ρ_2 whereas the anticipative component will remain small.

Case ($\iota\iota\iota$) Again to fix ideas, assume that V_0 is an arbitrarily small neighbourhood of $(1,0)$, i.e., $X_{1,t}$ is observed positive and much larger in magnitude than $X_{2,t}$. Contrary to (ι) and ($\iota\iota$) where, practically, a single outcome captures all the probability mass, several clearly distinct potential outcomes share the

[14] The size of this neighbourhood will be commensurate to the accuracy of the observed relative magnitudes: for smaller V_0 (higher observation accuracy), smaller neighbourhoods around $c(1,1,0,\rho_2)$ will provide the same level of certainty. For a fixed V_0, one can also evaluate the conditional probability over smaller neighbourhoods within V_0 by considering sets $A_{\theta,\eta,P}$ with $[\theta-\eta, \theta+\eta] \subset [\theta_0-\eta_0, \theta_0+\eta_0]$ for instance, which leads to the ratio in terms of Γ_2 as in the proposition. Note that P can be taken arbitrarily small regardless of V_0 without affecting the conditional probability, provided it contains $(0,0)$.

likelihood in this case.

Letting $A_{\theta_0,\eta_0,P}$ with $P = \Big([\rho_1^{-1} - \epsilon_1, \rho_1^{-1} + \epsilon_1] \cup [-\epsilon_1, \epsilon_1]\Big) \times \mathbb{R}$ for $\epsilon_1 > 0$ arbitrary small, we obtain after elementary computations that

$$\mathbb{P}_x^{\|\cdot\|}\Big(\underline{\boldsymbol{X}}_t, A_{\theta,\eta,P}\Big|B(V_0)\Big) \underset{x\to\infty}{\longrightarrow} 1.$$

Thus, the probability mass appears to be localised in a main region, which is a neighbourhood of the points $(1,0,0,z)$, $(1,0,\rho_1^{-1},z)$, $z \in \mathbb{R}$. Within this main region, the probability mass can be further localised into two distinct areas:

$$\mathbb{P}_x^{\|\cdot\|}\Big(\underline{\boldsymbol{X}}_t, A_{\theta,\eta,P}\Big|B(V_0)\Big) \underset{x\to\infty}{\longrightarrow} \begin{cases} \dfrac{\Gamma_2(V_0)}{\dfrac{\sigma_1^\alpha}{2}\dfrac{|\rho_1|^\alpha}{1-|\rho_1|^\alpha} + \Gamma_2(V_0)}, & \text{for} \quad P = [-\epsilon_1, \epsilon_1] \times [-\epsilon_2, \epsilon_2], \\ \\ \dfrac{\dfrac{\sigma_1^\alpha}{2}\dfrac{|\rho_1|^\alpha}{1-|\rho_1|^\alpha}}{\dfrac{\sigma_1^\alpha}{2}\dfrac{|\rho_1|^\alpha}{1-|\rho_1|^\alpha} + \Gamma_2(V_0)}, & \text{for} \quad P = [\rho_1^{-1} - \epsilon_1, \rho_1^{-1} + \epsilon_1] \times \mathbb{R}, \end{cases}$$

for $\epsilon_1, \epsilon_2 > 0$ small. Given that the two areas have complementary probability masses, it appears that with certainty: $\underline{\boldsymbol{X}}_t/\|\underline{\boldsymbol{X}}_t\|$ will either (1) belong to a small neighbourhood of $(1,0,0,0)$, or (2) belong to a small neighbourhood of the points $(1,0,\rho_1^{-1},z)$, $z \in \mathbb{R}$. The area corresponding to (1) yields a straightforward interpretation:

(1) The outcome $\Big\{\underline{\boldsymbol{X}}_t/\|\underline{\boldsymbol{X}}_t\|$ belongs to a small neighbourhood of $(1,0,0,0)\Big\}$ corresponds to an event in which the anticipative component is extreme at date t and collapses at $t+1$ while the non-anticipative series is small both at t and $t+1$. The conditional likelihood of this outcome can be arbitrarily large or small according to how much weight Γ_2 charges on the neighbourhood V_0 of $(1,0)$.

(2) Contrary to the previous case, the probability mass on the area $(1,0,\rho_1^{-1},z)$, $z \in \mathbb{R}$ does not appear to be localised in an arbitrarily small neighbourhood but can in general be dispersed over all z on the real line. This family of events describes outcomes for which, from date t to date $t+1$, the anticipative component increases by a factor ρ_1^{-1} while the non-anticipative series remains either non-extreme or jumps to some extreme value.

One can evaluate the probability mass of events corresponding to specific jumps sizes of the non-anticipative components. For any closed set $P_2 \subset \mathbb{R}$, one has

$$\mathbb{P}_x^{\|\cdot\|}\Big(\underline{\boldsymbol{X}}_t, A_{\theta,\eta,P}\Big|B(V_0)\Big) \underset{x\to\infty}{\longrightarrow} \frac{\dfrac{\sigma_1^\alpha}{2}\dfrac{|\rho_1|^{2\alpha}}{1-|\rho_1|^\alpha}\delta_{\{0\}}(P_2) + \dfrac{|\rho_1|^\alpha}{2}{\sigma_1}^\alpha_{|P_2}}{\dfrac{\sigma_1^\alpha}{2}\dfrac{|\rho_1|^\alpha}{1-|\rho_1|^\alpha} + \Gamma_2(V_0)}, \quad \text{for} \;\; P = [\rho_1^{-1} - \epsilon_1, \rho_1^{-1} + \epsilon_1] \times P_2.$$

- Taking for instance $P_2 = [M, +\infty[$ (resp. $P_2 =]-\infty, -M] \cup [M, +\infty[$), for some $M > 0$, one gets the conditional likelihood of events $(1,0,\rho^{-1},z)$, for $|z| \geq M$, i.e., outcomes for which the anticipative component increases by a factor ρ_1^{-1} and the non-anticipative jumps above the positive threshold M

(resp. outside the interval $]-M, M[$):

$$\mathbb{P}_x^{\|\cdot\|}\Big(\underline{\boldsymbol{X}}_t, A_{\theta,\eta,P}\Big|B(V_0)\Big) \underset{x\to\infty}{\longrightarrow} \frac{\dfrac{|\rho_1|^\alpha}{2}\displaystyle\int_{S(P_2)}|s_1|^\alpha\Gamma_2(d\boldsymbol{s})}{\dfrac{\sigma_1^\alpha}{2}\dfrac{|\rho_1|^\alpha}{1-|\rho_1|^\alpha}+\Gamma_2(V_0)}, \quad \text{for } P=[\rho^{-1}-\epsilon_1,\rho^{-1}+\epsilon_1]\times P_2,$$

with $S(P_2) := \left\{\frac{(\rho_1^{-1},y)}{\sqrt{\rho_1^{-2}+y^2}} \in S_2 : \ y \in P_2\right\}$.

- For $P_2 = [-\epsilon_2, \epsilon_2]$ with $\epsilon_2 > 0$ small, one can gauge the conditional likelihood of the anticipative component increasing by a factor ρ_1^{-1} from t to $t+1$ while the non-anticipative component remains close to non-extreme.

$$\mathbb{P}_x^{\|\cdot\|}\Big(\underline{\boldsymbol{X}}_t, A_{\theta,\eta,P}\Big|B(V_0)\Big) \underset{x\to\infty}{\longrightarrow} \frac{\dfrac{\sigma_1^\alpha}{2}\dfrac{|\rho_1|^{2\alpha}}{1-|\rho_1|^\alpha}+\dfrac{|\rho_1|^\alpha}{2}\displaystyle\int_{S(P_2)}|s_1|^\alpha\Gamma_2(d\boldsymbol{s})}{\dfrac{\sigma_1^\alpha}{2}\dfrac{|\rho_1|^\alpha}{1-|\rho_1|^\alpha}+\Gamma_2((1,0))}$$

for $P = [\rho^{-1}-\epsilon_1, \rho^{-1}+\epsilon_1]\times[-\epsilon_2,\epsilon_2]$.

Remark 4.5.1 Further insights can be drawn from case $(\iota\iota\iota)$ if $\Gamma_2((\pm1,0)) = 0$, i.e., if realisations $(\varepsilon_{1,t}, \varepsilon_{2,t})$ where $\varepsilon_{1,t}$ is extreme and $\varepsilon_{2,t}$ is not almost never occur. Then, $\Gamma_2(V_0)$ becomes arbitrarily close to 0 for V_0 an arbitrarily small neighbourhood of $(\vartheta, 0)$.[15] In that case, neglecting the difference and assuming $\Gamma_2(V_0) = \Gamma_2((\vartheta,0)) = 0$, we have

$$\mathbb{P}_x^{\|\cdot\|}\Big(\underline{\boldsymbol{X}}_t, A_{\theta,\eta,P}\Big|B(V_0)\Big) \underset{x\to\infty}{\longrightarrow} \begin{cases} 0, & \text{for} \quad P=[-\epsilon_1,\epsilon_1]\times[-\epsilon_2,\epsilon_2], \\ \\ 1, & \text{for} \quad P=[\rho_1^{-1}-\epsilon_1,\rho_1^{-1}+\epsilon_1]\times\mathbb{R}, \end{cases}$$

which indicates that a bubble in $(X_{1,t})$ necessarily reaches its peak at a jump date in $(X_{2,t})$. The bubble peak is always «signaled». Observing $X_{1,t}$ extreme and $X_{2,t}$ non-extreme thus implies that the bubble will last at least one more period.

Taking now $P_2 =]-\infty, M]\cup[M,+\infty[$ for $M > 0$ (resp. $P_2 = [-\epsilon_2,\epsilon_2]$ for $\epsilon_2 > 0$) arbitrarily small, notice that the integral $\int_{S(P_2)}|\sigma_1|^\alpha\Gamma_2(d\boldsymbol{s})$ can be made arbitrarily close to $\sigma_1^\alpha - \Gamma_2((\pm1,0))$ (resp. $\Gamma_2((\pm1,0))$). Again, assuming that $\Gamma_2((\pm1,0)) = 0$ and neglecting the difference, this yields that

$$\mathbb{P}_x^{\|\cdot\|}\Big(\underline{\boldsymbol{X}}_t, A_{\theta,\eta,P}\Big|B(V_0)\Big) \underset{x\to\infty}{\longrightarrow} \begin{cases} 1-|\rho_1|^\alpha, & \text{for} \quad P=[\rho_1^{-1}-\epsilon_1,\rho_1^{-1}+\epsilon_1]\times(]-\infty,M]\cup[M,+\infty[), \\ \\ |\rho_1|^\alpha, & \text{for} \quad P=[\rho_1^{-1}-\epsilon_1,\rho_1^{-1}+\epsilon_1]\times[-\epsilon_2,\epsilon_2], \end{cases}$$

for $M > 0$ and $\epsilon_2 > 0$ arbitrarily small. One recognises the probability of a bubble surviving one or more period, and its complementary, in the univariate anticipative AR(1) model.[16]

Table 4.1 summarises the potential outcomes of each specific conditioning event and illustrates the typical profile of the trajectory of $(\boldsymbol{X}_t)$ in each case.

[15] This holds because Γ_2 is a finite measure.

[16] Here, it would be more accurate to speak about the probability of a bubble in $(X_{1,t})$ surviving at least two more periods, given that it will survive at least one more.

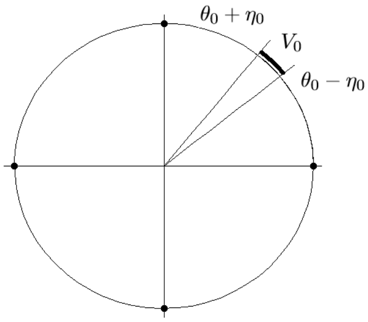
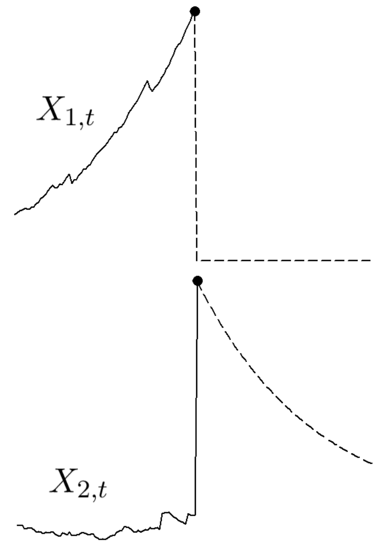
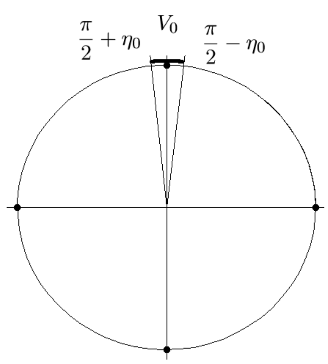
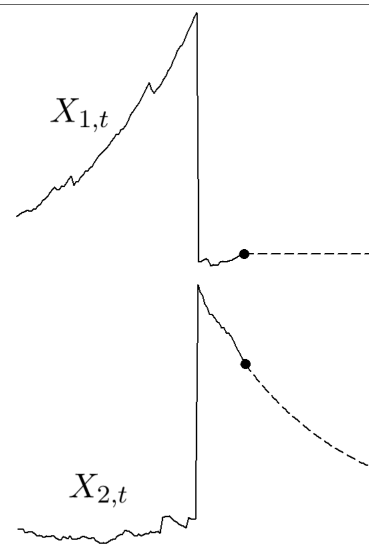
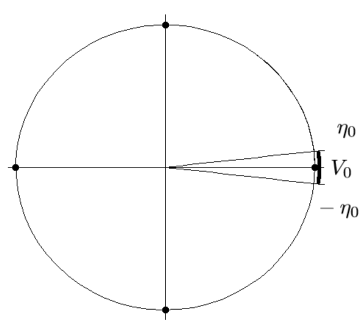
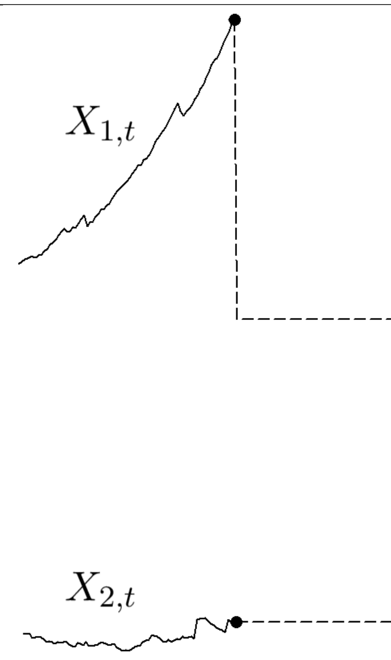
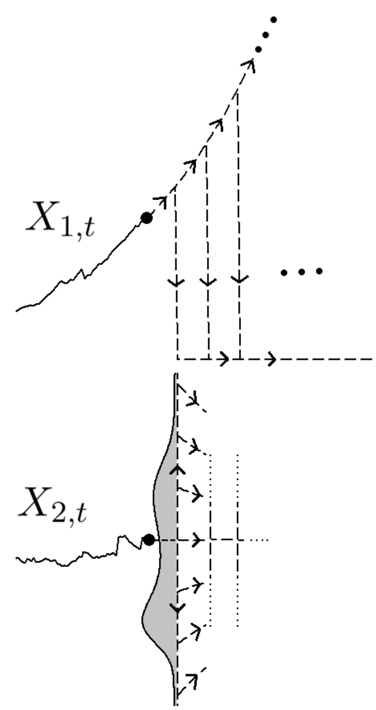
Observation $\frac{(X_{1,t}, X_{2,t})}{\sqrt{X_{1,t}^2 + X_{2,t}^2}} \in V_0$	Potential outcomes (neighbourhood of)	Conditional probability	Trajectorial interpretation
 $X_{1,t}$ and $X_{2,t}$ both extreme	$(\cos u, \sin u, 0, \rho_2 \sin u)$ $u \in [\theta_0 - \eta_0, \theta_0 + \eta_0]$	1	 Bubble peak signaled by jump
 $\lvert X_{1,t}\rvert << \lvert X_{2,t}\rvert$	$(0, 1, 0, \rho_2)$	1	 Post crash, jump decaying

Table 4.1: For each case considered in Proposition 4.5.2 (first column), the potential outcomes for $\underline{\boldsymbol{X}}_t/\|\underline{\boldsymbol{X}}_t\|$ is provided (second column), alongside the asymptotic conditional probability mass over the corresponding outcomes (third column). Each case can be related to specific events in the trajectory of $(\boldsymbol{X}_t)$, which are illustrated and labeled in the last column. The solid lines represent past trajectories and the present dates are symbolised by points. In the outcome $(1, 0, \rho_1^{-1}, z)$, $z \in \mathbb{R}$, the bubble survives at least one more period, but could survive more. Also, when the peak will be reached, a jump of *a priori* any size (including zero) may occur and then decay. Multiple potential paths are thus represented in dashed lines oriented by arrows, and the grey shaded area symbolises the jump size distribution. Here: $\omega_1 = \Gamma_2(V_0)$ and $\omega_2 = \sigma_1^\alpha |\rho_1|^\alpha / \Big(2(1 - |\rho_1|^\alpha)\Big)$.

<table>
<tr><th>Observation $\frac{(X_{1,t}, X_{2,t})}{\sqrt{X_{1,t}^2 + X_{2,t}^2}} \in V_0$</th><th>Potential outcomes (neighbourhood of)</th><th>Conditional probability</th><th>Trajectorial interpretation</th></tr>
<tr><td rowspan="2">

$|X_{1,t}| >> |X_{2,t}|$</td><td>$(1, 0, 0, 0)$</td><td>$\frac{\omega_1}{\omega_1 + \omega_2}$</td><td>

Bubble peak, no jump signal</td></tr>
<tr><td>$(1, 0, \rho_1^{-1}, z)$
for $z \in \mathbb{R}$</td><td>$\frac{\omega_2}{\omega_1 + \omega_2}$</td><td>

Pre-peak bubble inflation,
potential peak/jump at $t + 1$</td></tr>
</table>

4.6 Concluding remarks

By reformulating the path prediction problem of α-stable processes into a semi-norm representation problem of α-stable random vectors, we obtain the conditional distribution of future paths during extreme events. In this framework, our approach reveals that instead of their attractive «causal» interpretation, non-anticipative processes appear to rather presume, by construction, the unpredictability of extreme events. Anticipative processes however, instead of «depending on the future», rather assume that future events feature early visible signs betraying their incoming occurrences. These early signs take the form of emerging trends and patterns that an observer can identify and use to infer about future potential outcomes. Whether extreme events in some time series data feature early visible signs or not is arguably an intrinsic property of the natural phenomenon being measured rather than one of the modelling. One can nevertheless see that enforcing a non-anticipative process on any given time series data mechanically leads to a model which assumes that extremes are not inferrable beyond their unconditional likelihood of occurrence. It appears in addition that modelling a time series by a single, say, AR recursive equation, entails assuming that the considered series is determined by a single pattern appearing recurrently through time.[17]

In the univariate framework, processes resulting from linearly aggregating anticipative processes thus circumvent two implicitly «built-in» limitations of classical time series modelling, at least from a probabilistic standpoint. Numerous questions and perspectives remain nevertheless open, especially on statistical aspects:

- How variable can the future trajectory be around its predicted deterministic paths ?
- How can such processes be estimated/learnt from time series data ?
- What conditioning neighbourhood V in (4.25) to select given an observed piece of trajectory ?
- How many past observations $m+1$ to include in the prediction exercise ?
- What can be said when such processes are not far from their central values ?

The multivariate framework, which we only illustrated on a simple example, already displays the richness of interactions that can exist between several time series. Non-anticipative components sharing adequate dependence with anticipative ones become more predictable when both are considered jointly and can even bring information about future outcomes of the latter. The general case remains open as well and additional properties could emerge.

Last, it is often argued that linear processes suffer intrinsic limitations with regards to their dynamics and the type of patterns they can capture or reproduce. Proposition 4.4.1 and its subsequent remarks however show that not only are linear processes actually able to generate trajectories featuring any number

[17] At least in the heavy-tailed framework. In lighter-tailed frameworks, patterns are more weakly observed, if at all, and the dynamics is dominated by the persistence of the past trajectory [110].

of any kind of patterns through time, by the tuning of J and of the sequences $(d_{j,k})$ upon which only very mild assumptions are imposed, but that their conditional dynamics is moreover tractable. Future developments could even extend the notion of stable aggregates from the linear combination of a finite number of moving averages to a countable or a continuum of moving averages, as was done in [63] in the case of the anticipative AR(1). If the linearity assumption surely entails certain dynamical restrictions, the pattern-complexity of trajectories cannot be counted among these weaknesses. We conclude with an illustration of a linear process exhibiting strophoidal –looping-like– patterns.

Consider for a, b positive real numbers the horizontal strophoid $\mathcal{S} = \{(x(t), y(t)) \in \mathbb{R}^2 : \quad t \in \mathbb{R}\}$, where for any $t \in \mathbb{R}$,

$$x(t) = -at\frac{b-t^2}{1+t^2}, \qquad\qquad y(t) = \frac{a(b+1)}{1+t^2}.$$

Figure 4.1 provides an illustration of the horizontal strophoid for $a = 100$, $b = 5$. Letting for any $(x, y) \in \mathbb{R}^2$

$$\Pi_x(y) := y^3 - a(b+3)y^2 + (x^2 + a^2(2b+3))y - a^3(b+1),$$

a Cartesian equation of the locus of the strophoid is given by $\Pi_x(y) = 0$. Construct now a *non-random* sequence (d_k) in the following way: for a given $k \in \mathbb{Z}$, draw an element uniformly at random in the set $\{y \in \mathbb{R} : \ \Pi_k(y) = 0\}$ –which may contain either one, two or three elements– and assign it to d_k. Define then the process $X_t = \sum_{k\in\mathbb{Z}} d_k\varepsilon_{t+k}$ for (ε_t) an i.i.d. α-stable sequence, $1/2 < \alpha < 2$.[18] It can be checked that the process (X_t) is (m, h)-past-representable for any $m \geq 0$, $h \geq 1$. Proposition 4.4.1 applied to $\boldsymbol{X_t} = (X_{t-m}, \dots, X_{t+h})$ with $V = S_{m+1}$ shows that $\boldsymbol{X_t}/\|\boldsymbol{X_t}\|$ is asymptotically of the form $\pm\boldsymbol{d}_{k_0}/\|\boldsymbol{d}_{k_0}\|$, for some $k_0 \in \mathbb{Z}$. Given the construction of (d_k), we deduce that the linear process (X_t) features looping-like patterns in its trajectories, as depicted on Figure 4.2 for the choice of parameters $a = 100$ and $b = 5$. Its thorough analysis is left for further research.

[18] An elementary analysis shows that $\lim_{t\to t_1} |x(t)| = +\infty$, $\lim_{t\to t_2} y(t) = 0$ if and only if $t_1 = \pm\infty$ and $t_2 = \pm\infty$, and that $y(t)/x^2(t) \to (b+1)/a$ for $t \to \pm\infty$. Thus, $d_k \underset{|k|\to\infty}{\sim} \text{const } k^{-2}$ and (X_t) is well defined for $1/2 < \alpha < 2$.

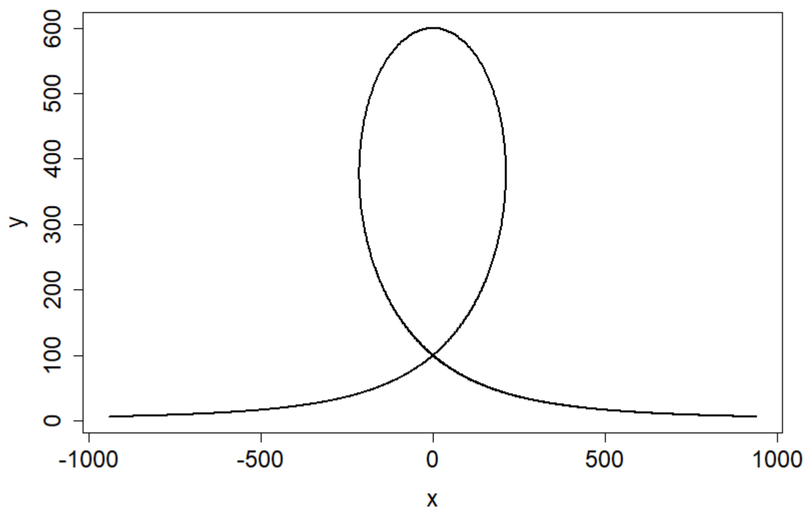


Figure 4.1: Horizontal strophoid for $a = 100$, $b = 5$.

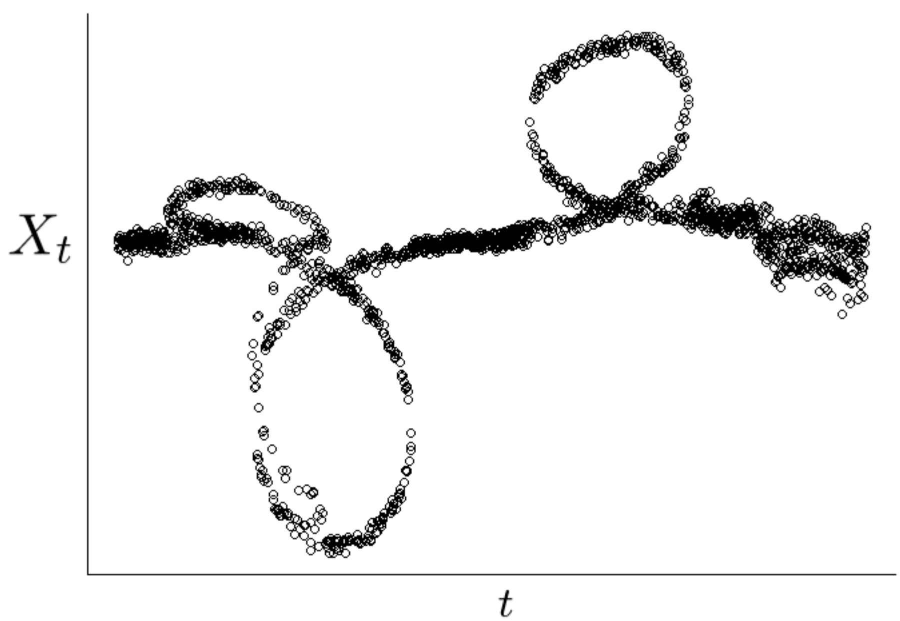


Figure 4.2: Sample path of a linear 1-stable process featuring strophoidal patterns ($a = 100$, $b = 5$).

4.7 Postponed proofs

4.7.1 Proof of Proposition 4.2.1

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Reciprocally, assume there exists a measure $\gamma^{\|\cdot\|}$ on $C_d^{\|\cdot\|}$ satisfying (4.6) and a non-random vector $\boldsymbol{m}^0_{\|\cdot\|} \in \mathbb{R}^d$ such that

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

First, we can see that

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

We will later show the following result:

Lemma 4.7.1 *Let $\gamma^{\|\cdot\|}$ a Borel measure on $C_d^{\|\cdot\|}$ satisfying* (4.6). *Then,*

$$\int_{C_d^{\|\cdot\|}} \|\boldsymbol{s}\|_e \Big| \ln \|\boldsymbol{s}\|_e \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty. \tag{4.36}$$

Assuming Lemma 4.7.1 holds, then by the Cauchy-Schwarz inequality, we have $\int_{C_d^{\|\cdot\|}} |\langle \boldsymbol{u}, \boldsymbol{s}\rangle| \Big| \ln \|\boldsymbol{s}\|_e \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty$, and thus

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

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

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

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

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

Last, as $\int_{C_d^{\|\cdot\|}} \|\boldsymbol{s}\|_e \Big| \ln \|\boldsymbol{s}\|_e \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty$ (Lemma 4.7.1) and $\Gamma(K^{\|\cdot\|}) = 0$, we have by a change of variable

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

which concludes the proof of Proposition 4.2.1.

Proof of Lemma 4.7.1

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

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

We will show that both I_1 and I_2 are finite. Focus first on I_2. From (4.6), we know that for all $\boldsymbol{u} \in \mathbb{R}^d$

$$\int_{C_d^{\|\cdot\|}} |\langle \boldsymbol{u}, \boldsymbol{s}\rangle| \Big| \ln |\langle \boldsymbol{u}, \boldsymbol{s}\rangle| \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) = \int_{C_d^{\|\cdot\|} \cap \{b \leq \|\boldsymbol{s}\|_e \leq M\}} + \int_{C_d^{\|\cdot\|} \cap \{\|\boldsymbol{s}\|_e > M\}} < +\infty. \tag{4.37}$$

and thus, in particular

$$\begin{aligned}\int_{\{\boldsymbol{s}' \in C_d^{\|\cdot\|}:\ \|\boldsymbol{s}'\|_e > M\}} & |\langle \boldsymbol{u}, \boldsymbol{s}\rangle| \Big| \ln |\langle \boldsymbol{u}, \boldsymbol{s}\rangle| \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) \\ &= \int_{\{\boldsymbol{s}' \in C_d^{\|\cdot\|}:\ \|\boldsymbol{s}'\|_e > M\}} |\langle \boldsymbol{u}, \boldsymbol{s}\rangle| \Big| \ln \|\boldsymbol{s}\|_e + \ln |\langle \boldsymbol{u}, \frac{\boldsymbol{s}}{\|\boldsymbol{s}\|_e}\rangle| \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty. \end{aligned} \tag{4.38}$$

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

$$\begin{aligned}\int_{\{\boldsymbol{s}' \in C_d^{\|\cdot\|}:\ \|\boldsymbol{s}'\|_e > M\}} & |\langle \boldsymbol{u}, \boldsymbol{s}\rangle| \Big| \ln \|\boldsymbol{s}\|_e + \ln |\langle \boldsymbol{u}, \frac{\boldsymbol{s}}{\|\boldsymbol{s}\|_e}\rangle| \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) \\ &= \int_{\{\boldsymbol{s}' \in C_d^{\|\cdot\|}:\ \|\boldsymbol{s}'\|_e > M\}} |\langle \boldsymbol{u}, \boldsymbol{s}\rangle| \Big| \ln \|\boldsymbol{s}\|_e \Big| \Big| 1 + \frac{\ln |\langle \boldsymbol{u}, \boldsymbol{s}/\|\boldsymbol{s}\|_e\rangle|}{\ln \|\boldsymbol{s}\|_e} \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) \\ &\geq \int_{\{\boldsymbol{s}' \in C_d^{\|\cdot\|}:\ \|\boldsymbol{s}'\|_e > M\}} |\langle \boldsymbol{u}, \boldsymbol{s}\rangle| \Big| \ln \|\boldsymbol{s}\|_e \Big| \Big| 1 - \Big| \frac{\ln |\langle \boldsymbol{u}, \boldsymbol{s}/\|\boldsymbol{s}\|_e\rangle|}{\ln \|\boldsymbol{s}\|_e} \Big| \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) \end{aligned} \tag{4.39}$$

Let us now partition the space $\mathbb{R}^d$ into subsets $R_1, \dots, R_d$ such that, for any $i = 1, \dots, d$ and any $\boldsymbol{s} = (s_1, \dots, s_d) \in R_i$, $\sup_j |s_j| = |s_i|$.[19] We have by (4.38)-(4.39) that for any $i = 1, \dots, d$, any $\boldsymbol{u} \in \mathbb{R}^d$,

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

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

$$\int_{\{\boldsymbol{s}' \in C_d^{\|\cdot\|}:\ \|\boldsymbol{s}'\|_e > M\} \cap R_i} |\langle \boldsymbol{e}_i, \boldsymbol{s}\rangle| \Big| \ln \|\boldsymbol{s}\|_e \Big| \Big| 1 - \Big| \frac{\ln |\langle \boldsymbol{e}_i, \boldsymbol{s}/\|\boldsymbol{s}\|_e\rangle|}{\ln \|\boldsymbol{s}\|_e} \Big| \Big| \gamma^{\|\cdot\|}(d\boldsymbol{s}) < +\infty. \tag{4.40}$$

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

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

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

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

Thus, for all $n \geq 1$

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

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

and

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

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

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

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

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

Hence, $\boldsymbol{s}_n/\|\boldsymbol{s}_n\|_e \underset{n\to+\infty}{\longrightarrow} 0$, which is impossible since $\Big\|\boldsymbol{s}_n/\|\boldsymbol{s}_n\|_e\Big\|_e = 1$ for all $n \geq 1$. The function $\boldsymbol{s} \longmapsto \ln|\langle \boldsymbol{e}_i, \boldsymbol{s}/\|\boldsymbol{s}\|_e\rangle|$ is thus bounded on $\{\boldsymbol{s} \in C_d^{\|\cdot\|} : \|\boldsymbol{s}\|_e > M\} \cap R_i$, say $\Big|\ln|\langle \boldsymbol{e}_i, \boldsymbol{s}/\|\boldsymbol{s}\|_e\rangle|\Big| \leq A$ for some $A > 0$. Provided M is taken large enough (e.g., $M > 2A$), we will have in (4.40)

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

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

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

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

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

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

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

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

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

we get

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

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

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

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

for any $i = 1, \dots, d$, $\boldsymbol{s} \in \{\boldsymbol{s'} \in C_d^{\|\cdot\|} :\ b \leq \|\boldsymbol{s'}\|_e \leq M\} \cap R_i$. From (4.37), we have for any $\boldsymbol{u} \in \mathbb{R}^d$

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

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

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

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

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

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

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

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

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

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

4.7.2 Proof of Proposition 4.2.2

The proposition is an immediate consequence of Bayes formula and of the following result, which is an adaptation of Theorem 4.4.8 by Samorodnitsky and Taqqu (1994) [117] to semi-norms.

Proposition 4.7.1 *Let $\boldsymbol{X} = (X_1, \dots, X_d)$ be an α-stable random vector and let $\|\cdot\|$ be a semi-norm on $\mathbb{R}^d$ such that $\boldsymbol{X}$ is representable on $C_d^{\|\cdot\|}$. Then, for every Borel set $A \subseteq C_d^{\|\cdot\|}$ with $\Gamma^{\|\cdot\|}(\partial A) = 0$,*

$$\lim_{x \to +\infty} x^\alpha \mathbb{P}\Big(\|\boldsymbol{X}\| > x, \frac{\boldsymbol{X}}{\|\boldsymbol{X}\|} \in A\Big) = C_\alpha \Gamma^{\|\cdot\|}(A), \tag{4.41}$$

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

Proof.

We follow the proof of Theorem 4.4.8 by Samorodnitsky and Taqqu (1994) [117]. The main hurdle is to show that, with $\|\cdot\|$ a semi-norm, $K^{\|\cdot\|} = \{\boldsymbol{s} \in S_d : \|\boldsymbol{s}\| = 0\}$, and $\Gamma^{\|\cdot\|}(K^{\|\cdot\|}) = 0$, we have the series representation of $\boldsymbol{X}$, $(X_1, \ldots, X_d) \stackrel{d}{=} (Z_1, \ldots, Z_d)$ where

$$Z_k = (C_\alpha \Gamma^{\|\cdot\|}(C_d^{\|\cdot\|}))^{1/\alpha} \sum_{i=1}^{\infty} [\Gamma_i^{-1/\alpha} S_i^{(k)} - b_{i,k}(\alpha)], \quad k = 1, \ldots, d, \tag{4.42}$$

with $\boldsymbol{S}_i = (S_i^{(1)}, \ldots, S_i^{(d)})$, $i \geq 1$, are i.i.d. $C_d^{\|\cdot\|}$-valued random vectors with common law $\Gamma^{\|\cdot\|}/\Gamma^{\|\cdot\|}(C_d^{\|\cdot\|})$ and the $b_{i,k(\alpha)}$'s are constants.

By Proposition 4.2.1, we know that $\boldsymbol{X}$ admits a characteristic function of the form (4.3). This allows to restate the integral representation Theorem 3.5.6 in [117] on the semi-norm unit cylinder as follows: with the measurable space $(E, \mathcal{E}) = (C_d^{\|\cdot\|}, \text{Borel}\,\sigma\text{-algebra on}\, C_d^{\|\cdot\|})$, let M be an α-stable random measure on $(E, \mathcal{E})$ with control measure $m = \Gamma^{\|\cdot\|}$, skewness intensity $\beta(\,\cdot\,) \equiv 1$ (see Definition 3.3.1 in [117] for details). Letting also $f_j : C_d^{\|\cdot\|} \longrightarrow \mathbb{R}$ defined by $f_j\Big((s_1, \ldots, s_d)\Big) = s_j$, $j = 1, \ldots, d$, then

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

This representation can be checked directly by comparing the characteristic functions of the left-hand and right-hand sides. We can now apply Theorem 3.10.1 in [117] to the above integral representation with $(E, \mathcal{E}, m)$ the measure space as described before, and $\hat{m} = \Gamma^{\|\cdot\|}/\Gamma^{\|\cdot\|}(C_d^{\|\cdot\|})$. This establishes (4.42). The rest of the proof is similar to that of Theorem 4.4.8 in [117]. We rely on the triangle inequality property of semi-norms and the fact that any norm is finer than any semi-norm in finite dimension.[20] □

4.7.3 Proof of Lemma 4.3.1

From Proposition 4.2.1, we know that a necessary condition for the representability of $\boldsymbol{X_t}$ on $C_{m+h+1}^{\|\cdot\|}$ is $\Gamma(K^{\|\cdot\|}) = 0$, where $K^{\|\cdot\|} = \{\boldsymbol{s} \in S_{m+h+1} : \|\boldsymbol{s}\| = 0\}$. This condition is also sufficient when either $\alpha \neq 1$ or $\alpha = 1$, $\beta = 0$. Using the fact that Γ only charges discrete atoms on $C_{m+h+1}^{\|\cdot\|}$,

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

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

by (4.8). Now assume that the following holds:

$$\forall k \in \mathbb{Z}, \quad \Big[(d_{k+m}, \dots, d_k) = \mathbf{0} \Longrightarrow (d_{k+m}, \dots, d_{k-h}) = \mathbf{0}\Big]. \tag{4.43}$$

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

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

It implies that

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

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

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

Invoking (4.43), we deduce by recurrence that for any $n \geq 0$,

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

Therefore, (4.43) implies

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

The reciprocal is clearly true. This establishes that (4.14) is a necessary and sufficient condition for $\boldsymbol{X}_t$ to be representable on $C_d^{\|\cdot\|}$ in the cases where either $\alpha \neq 1$, or $\alpha = 1$, $\beta = 0$.

In the case $\alpha = 1$, $\beta \neq 0$, Proposition 4.2.1 states that the necessary and sufficient condition for representability reads $\int_{S_d} \Big| \ln \|\boldsymbol{s}\| \Big| \Gamma(d\boldsymbol{s}) < +\infty$. That is

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

Substituting Γ by its expression in (4.13), the above condition holds if and only if (4.14) is true and

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

the latter being equivalent to

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

4.7.4 Proof of Proposition 4.3.1

By Definition 4.3.1, (X_t) is past-representable if and only if there exists $m \geq 0$, $h \geq 1$ such that the vector $(X_{t-m}, \dots, X_t, X_{t+1}, \dots, X_{t+h})$ is representable on $C_{m+h+1}^{\|\cdot\|}$. Consider first point $(\iota)(a)$, that is, the case $\alpha \neq 1$, $(\alpha, \beta) = (1, 0)$. By Lemma 4.3.1,

$$\begin{aligned}(X_t) \text{ is past-representable } &\iff \text{There exist } m \geq 0,\ h \geq 1, \text{ such that (4.14) holds} \\ &\iff \exists m \geq 0, \forall k \in \mathbb{Z}, \Big[d_{k+m} = \ldots = d_k = 0 \implies \forall \ell \leq k-1,\ \ d_\ell = 0\Big].\end{aligned}$$

Thus,

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

hence (4.17).

Regarding the last statement of point $(\iota)(a)$, assume first that $m_0 < +\infty$ and $m \geq m_0$. Property (4.14) necessarily holds with m_0. Indeed, if it did not, there would exist $k \in \mathbb{Z}$ such that

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

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

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

As (4.14) holds with m_0, it holds *a fortiori* for any $m' \geq m_0$. Thus, $\boldsymbol{X}_t = (X_{t-m}, \ldots, X_t, X_{t+1}, \ldots, X_{t+h})$ is representable for any $m' \geq m_0$, $h \geq 1$ by Lemma 4.3.1, and (X_t) is in particular (m,h)-past-representable. Reciprocally let $m \geq 0$, $h \geq 1$ and assume that (X_t) is (m,h)-past-representable. The process (X_t) is thus in particular past-representable, which as we have shown previously, implies that $m_0 < +\infty$. *Ad absurdum*, suppose now that $0 \leq m < m_0 < +\infty$. If $m_0 = 0$, there is nothing to do. Otherwise if $m_0 \geq 1$, by definition, there exists a $k \in \mathbb{Z}$ such that

$$d_{k+m_0} = \ldots = d_{k+1} = 0, \quad \text{and} \quad d_k \neq 0. \tag{4.44}$$

Because (X_t) is (m,h)-past-representable, we have by Lemma 4.3.1 that (4.14) holds with m. As $m < m_0$ and $d_{k+m_0} = \ldots = d_{k+1} = 0$, we thus have that $d_\ell = 0$ for all $\ell \leq k+1$, and in particular $d_k = 0$, hence the contradiction. We conclude that $m \geq m_0$.

Consider now point $(\iota)(b)$, i.e., the case $\alpha = 1$ and $\beta \neq 0$. From Lemma 4.3.1,

$$(X_t) \text{ is past-representable } \iff \text{ There exist } m \geq 0, \ h \geq 1, \text{ such that (4.14) and (4.15) hold}$$

From the previous proof, we moreover have that

$$\exists\, m \geq 0, \text{ such that (4.14) holds } \iff m_0 < +\infty \iff \begin{cases} m_0 < +\infty \\ \forall\, m' \geq m_0, \text{ (4.14) holds} \\ \forall\, m' < m_0, \text{ (4.14) does not hold} \end{cases}$$

Hence

$$
\begin{aligned}
&\exists\, m \geq 0,\ h \geq 1, \text{ such that (4.14) and (4.15) hold} \\
&\qquad\qquad\qquad \iff \begin{cases} m_0 < +\infty \\ \forall\, m' \geq m_0,\ \text{(4.14) holds} \\ \forall\, m' < m_0,\ \text{(4.14) does not hold} \\ \exists\, m \geq 0,\ h \geq 1, \text{ such that (4.14) and (4.15) hold.} \end{cases}
\end{aligned}
$$

The latter in particular implies $m_0 < +\infty$ and the existence of $m \geq m_0$, $h \geq 1$ such that (4.15) holds. Reciprocally,

$$
\begin{aligned}
&\begin{cases} m_0 < +\infty \\ \exists\, m \geq m_0,\ h \geq 1, \text{ such that (4.15) holds} \end{cases} \\
&\qquad\qquad\qquad \Longrightarrow \begin{cases} m_0 < +\infty \\ \forall\, m' \geq m_0,\ \text{(4.14) holds} \\ \exists\, m \geq m_0,\ h \geq 1, \text{ such that (4.15) holds,} \end{cases}
\end{aligned}
$$

which in particular implies that there exists $m \geq m_0$, $h \geq 1$ such that both (4.14) and (4.15) hold. Hence the past-representability of (X_t).

In view of Definition 4.3.1, point $(\iota\iota)$ is a direct consequence of the second part of Proposition 4.2.1.

4.7.5 Proof of Corollary 4.3.1

Letting k_0 be the greatest integer such that $d_{k_0} \neq 0$ (such an index exists by (4.10)), then immediately, for any $m \geq 1$, $d_{k_0+m} = \ldots = d_{k_0+1} = 0$ and therefore $m_0 = +\infty$.

4.7.6 Proof of Corollary 4.3.2

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

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

Reciprocally, assume $\deg(\psi) = p \geq 1$. Let us first show that (4.17) holds.
Denote $\psi(F)\phi(B) = \sum_{i=-q}^{p} \varphi_i F^i$ and $\Theta(F)H(B) = \sum_{k=-r}^{s} \theta_i F^i$, for any non-negative degrees $q = \deg(\phi)$,

$r = \deg(H)$, $s = \deg(\Theta)$. From the recursive equation satisfied by (X_t), we have that

$$
\begin{aligned}
& & \sum_{i=-q}^{p} \varphi_i X_{t+i} &= \sum_{k=-r}^{s} \theta_k \varepsilon_{t+k} \\
\Longleftrightarrow & & \sum_{i=-q}^{p} \varphi_i \sum_{k\in\mathbb{Z}} d_k \varepsilon_{t+k+i} &= \sum_{k=-r}^{s} \theta_k \varepsilon_{t+k} \\
\Longleftrightarrow & & \sum_{k\in\mathbb{Z}} \left(\sum_{i=-q}^{p} \varphi_i d_{k-i} \right) \varepsilon_{t+k} &= \sum_{k=-r}^{s} \theta_k \varepsilon_{t+k}. \qquad (4.45)
\end{aligned}
$$

Proceeding by identification using the uniqueness of representation of heavy-tailed moving averages (see [62]), we get that for $|k| > \max(r, s)$,

$$\sum_{i=-q}^{p} \varphi_i d_{k-i} = 0. \qquad (4.46)$$

Ad absurdum, if (X_t) is not past-representable, then by Proposition 4.3.1

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

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

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

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

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

There are two subsequences such that $m_{g(n)} \longrightarrow +\infty$ and $|k_{g(n)}| \longrightarrow +\infty$. For some n large enough such that (4.46) holds and $m_{g(n)} \geq p + q$, we have both

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

and

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

Hence,

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

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

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

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

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

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

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

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

$$X_t = \sum_{k=-\infty}^{\bar{k}} d_k \varepsilon_{t+k}, \qquad t \in \mathbb{Z}. \tag{4.47}$$

However, we also have by partial fraction decomposition

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

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

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

for some sequences of coefficients (c_k), (e_k), and where ℓ_1 is the degree of the largest order monomial in B of $\Theta(F)H(B)b_1(B)$ (recall that $F = B^{-1}$) and ℓ_2 is the degree of the largest monomial in F of $B^p\Theta(F)H(B)b_2(B)$. By (4.47), we deduce by identification that there is some $\bar{\ell} \in \mathbb{Z}$ such that $c_k = 0$ for all $k \geq \bar{\ell} + 1$ and

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

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

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

which yields

$$\Theta(z^{-1})H(z)b_1(z) = \psi(z^{-1})(P(z^{-1}) + Q(z)), \qquad z \in \mathbb{C}. \tag{4.48}$$

As $\deg(\psi) = p$ and $\psi(z) = 0$ if and only if $|z| > 1$, we know that there are p complex numbers $z_1, \ldots, z_p$ such that $0 < |z_i| < 1$ and $\psi(z_i^{-1}) = 0$ for $i = 1, \ldots, p$. Evaluating (4.48) at the z_i's, we get that

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

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

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

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

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

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

Let us now show that whenever $m_0 < +\infty$, then (4.15) holds for any $m \geq m_0$.

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

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

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

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

and

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

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

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

The equivalence between (ι) and ($\iota\iota\iota$) is now clear: on the one hand, if $m_0 < +\infty$, then (4.15) holds for all $m \geq m_0$, $h \geq 1$, which yields the (m,h)-past-representability of $(X_{t-m}, \ldots, X_t, X_{t+1}, \ldots, X_{t+h})$ for any $m \geq m_0$, $h \geq 1$, by Lemma 4.3.1. In particular, (X_t) is past-representable. On the other hand, assuming (X_t) is past-representable, then necessarily $m_0 < +\infty$.

Regarding the last statement, it follows from the above proof that the condition $m_0 < +\infty$ and $m \geq m_0$ is sufficient for (m,h)-past-representability. It is also necessary, as (4.14) never holds with $m < m_0$ (*a fortiori*, with $m < m_0 = +\infty$), concluding the proof.

4.7.7 Proof of Lemma 4.3.2

Denote $\boldsymbol{X}_{j,t} = (X_{j,t-m}, \ldots, X_{j,t}, X_{j,t+1}, \ldots, X_{j,t+h})$ the paths of the moving averages $(X_{j,t})$, for $j = 1, \ldots, J$. The $\boldsymbol{X}_{j,t}$'s are independent α-stable random vectors with spectral representations $(\Gamma_j, \boldsymbol{\mu}_j^0)$ of the form (4.13). We consider only the more delicate case $\alpha = 1$ and $\beta_j \in [-1, 1]$ for $j = 1, \ldots, J$. Because of the independence between $\boldsymbol{X}_{1,t}, \ldots, \boldsymbol{X}_{J,t}$, we have with $a = 2/\pi$

$$\begin{aligned}
\mathbb{E}\Big[e^{i\langle \boldsymbol{u}, \boldsymbol{X_t}\rangle}\Big] &= \mathbb{E}\Big[e^{i\langle \boldsymbol{u}, \sum_{j=1}^J \pi_j \boldsymbol{X}_{j,t}\rangle}\Big] = \prod_{j=1}^J \mathbb{E}\Big[e^{i\langle \pi_j \boldsymbol{u}, \boldsymbol{X}_{j,t}\rangle}\Big] \\
&= \prod_{j=1}^J \exp\left\{ - \int_{S_{m+h+1}} \Big(|\langle \pi_j \boldsymbol{u}, \boldsymbol{s}\rangle| + ia\langle \pi_j \boldsymbol{u}, \boldsymbol{s}\rangle \ln|\langle \pi_j \boldsymbol{u}, \boldsymbol{s}\rangle|\Big)\Gamma_j(d\boldsymbol{s}) + i\langle \pi_j \boldsymbol{u}, \boldsymbol{\mu_j}^0\rangle \right\} \\
&= \exp\left\{ - \int_{S_{m+h+1}} \Big(|\langle \boldsymbol{u}, \boldsymbol{s}\rangle| + ia\langle \boldsymbol{u}, \boldsymbol{s}\rangle \ln|\langle \boldsymbol{u}, \boldsymbol{s}\rangle|\Big) \sum_{j=1}^J \pi_j \Gamma_j(d\boldsymbol{s}) \right. \\
&\qquad\qquad \left. + i \sum_{j=1}^J \left(\langle \boldsymbol{u}, \pi_j \boldsymbol{\mu_j}^0\rangle - a\pi_j \ln \pi_j \int_{S_{m+h+1}} \langle \boldsymbol{u}, \boldsymbol{s}\rangle \Gamma_j(d\boldsymbol{s}) \right) \right\}.
\end{aligned}$$

Focusing on the shift vector, we have

$$\sum_{j=1}^J \left(\langle \boldsymbol{u}, \pi_j \boldsymbol{\mu_j}^0\rangle - a\pi_j \ln \pi_j \int_{S_{m+h+1}} \langle \boldsymbol{u}, \boldsymbol{s}\rangle \Gamma_j(d\boldsymbol{s}) \right) = \langle \boldsymbol{u}, \sum_{j=1}^J \pi_j(\boldsymbol{\mu}_j^0 - a \ln \pi_j \tilde{\boldsymbol{\mu}}_j)\rangle,$$

with $\tilde{\boldsymbol{\mu}}_j = (\tilde{\mu}_{j,\ell})$ and $\tilde{\mu}_{j,\ell} = \int_{S_{m+h+1}} s_\ell \Gamma_j(d\boldsymbol{s})$, $\ell = -m, \ldots, 0, 1, \ldots, h$. Using the form of Γ_j in (4.13), i.e., $\Gamma_j = \sum_{\vartheta \in S_1} \sum_{k \in \mathbb{Z}} w_{j,\vartheta} \|\boldsymbol{d}_{j,k}\|_e \delta_{\left\{\frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|_e}\right\}}$, we get

$$\tilde{\mu}_{j,\ell} = \int_{S_{m+h+1}} s_\ell \Gamma_j(d\boldsymbol{s}) = \sum_{\vartheta \in S_1} \sum_{k \in \mathbb{Z}} w_{j,\vartheta} \|\boldsymbol{d}_{j,k}\|_e \frac{\vartheta d_{j,k+\ell}}{\|\boldsymbol{d}_{j,k}\|_e} = \beta_j \sum_{k \in \mathbb{Z}} d_{j,k+\ell}, \quad \ell = -m, \ldots, h.$$

Hence, $\tilde{\boldsymbol{\mu}}_j = \beta_j \sum_{k \in \mathbb{Z}} \boldsymbol{d}_{j,k}$, and using the form of $\boldsymbol{\mu}_j^0$ as given in (4.13),

$$\begin{aligned}
\sum_{j=1}^J \pi_j(\boldsymbol{\mu}_j^0 - a \ln \pi_j \tilde{\boldsymbol{\mu}}_j) &= \sum_{j=1}^J \pi_j \left(\beta_j \sum_{k \in \mathbb{Z}} \boldsymbol{d}_{j,k} \ln \|\boldsymbol{d}_{j,k}\|_e - a \ln \pi_j \beta_j \sum_{k \in \mathbb{Z}} \boldsymbol{d}_{j,k} \right) \\
&= -a \sum_{j=1}^J \sum_{k \in \mathbb{Z}} \pi_j \beta_j \boldsymbol{d}_{j,k} \ln \|\pi_j \boldsymbol{d}_{j,k}\|_e \\
&:= \boldsymbol{\mu}^0.
\end{aligned}$$

Therefore,

$$\mathbb{E}\Big[e^{i\langle \boldsymbol{u},\boldsymbol{X_t}\rangle}\Big] = \exp\left\{-\int_{S_{m+h+1}}\Big(|\langle \boldsymbol{u},\boldsymbol{s}\rangle| + ia\langle \boldsymbol{u},\boldsymbol{s}\rangle \ln|\langle \boldsymbol{u},\boldsymbol{s}\rangle|\Big)\sum_{j=1}^{J}\pi_j\Gamma_j(d\boldsymbol{s}) + i\langle \boldsymbol{u},\boldsymbol{\mu}^0\rangle\right\},$$

and the random vector $\boldsymbol{X}_t$ is 1-stable with spectral measure

$$\sum_{j=1}^{J}\pi_j\Gamma_j = \sum_{j=1}^{J}\sum_{\vartheta\in S_1}\sum_{k\in\mathbb{Z}} w_{j,\vartheta}\pi_j^{\alpha}\|\boldsymbol{d}_{j,k}\|_e^{\alpha}\delta_{\left\{\frac{\vartheta\boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|_e}\right\}},$$

by (4.13), and shift vector as announced in the lemma.

4.7.8 Proof of Lemma 4.3.3

With the usual notations, let the $\boldsymbol{X}_{j,t}$'s be the paths of the moving averages $(X_{j,t})$'s and let Γ_j, $j=1,\dots,J$, their spectral measures on the Euclidean unit sphere. Let Γ the spectral measure of $\boldsymbol{X}_t$. By Lemma (4.3.2), $\Gamma=\sum_{j=1}^{J}\pi_j^{\alpha}\Gamma_j$. Thus, by Proposition 4.2.1, in the cases where either $\alpha\neq 1$ or $\boldsymbol{X}_t$ is symmetric, the vector $\boldsymbol{X}_t$ is representable on $C_{m+h+1}^{\|\cdot\|}$ if and only if

$$\begin{aligned}\Gamma(K^{\|\cdot\|})=0 \quad &\iff \quad \sum_{j=1}^{J}\pi_j^{\alpha}\Gamma_j(K^{\|\cdot\|})=0\\ &\iff \quad \Gamma_j(K^{\|\cdot\|})=0,\ \ \forall\, j=1,\dots,J.\end{aligned}$$

Given that the Γ_j's are the spectral measures of paths of non-aggregated moving averages, it has been shown in the proof of Lemma 4.3.1 $\Gamma_j(K^{\|\cdot\|})$ if and only if (4.14) holds for m and the sequence $(d_{j,k})_k$. The conclusion in that case follows. The case $\alpha\neq 1$ and $\boldsymbol{X}_t$ asymmetric is similar.

4.7.9 Proof of Proposition 4.3.2

If $\alpha\neq 1$, we have by Lemma 4.3.1 and the proof of Proposition 4.3.1,

$$\begin{aligned}(X_t)\text{ past-representable} \quad &\iff \quad \exists\, m\geq 0,\ (4.14)\text{ holds with } m \text{ for all sequences } (d_{j,k})_k\\ &\iff \quad \forall\, j=1,\dots,J,\ m_{0,j}<+\infty\\ &\iff \quad \forall\, j=1,\dots,J,\ (X_{j,t})\text{ past-representable.}\end{aligned}$$

For a given series $(d_{j,k})_k$, (4.14) holds with $m\geq m_{0,j}$ and does not hold with $m<m_{0,j}$. Regarding the last statement, we know that for (X_t) (m,h)-past-representable, (4.14) holds with the same m for all the sequences $(d_{j,k})_k$, $j=1,\dots,J$. This holds if $m\geq \max_j m_{0,j}$ and cannot hold for if $m<\max_j m_{0,j}$.

In the case where $\alpha = 1$, again by Lemma 4.3.1 and denoting generically by $\boldsymbol{X}_t$ a vector

$(X_{t-m}, \dots, X_t, X_{t+1}, \dots, X_{t+h})$ of size $m+h+1$,

$$
\begin{aligned}
&(X_t) \text{ past-representable} \\
&\iff \exists\, m \geq 0, h \geq 1, \begin{cases} \boldsymbol{X}_t \ S1S \text{ and (4.14) holds with } m \text{ for all sequences } (d_{j,k})_k \\ \text{or} \\ \boldsymbol{X}_t \text{ asymmetric and (4.14)-(4.15) hold with } m, h \text{ for all sequences } (d_{j,k})_k \end{cases} \\
&\iff \forall\, j = 1, \dots, J,\ m_{0,j} < +\infty,\ \text{and } \exists\, m \geq 0, h \geq 1, \begin{cases} \boldsymbol{X}_t \ S1S \\ \text{or} \\ \boldsymbol{X}_t \text{ asymmetric and (4.15) hold} \\ \text{with } m, h \text{ for all sequences } (d_{j,k})_k \end{cases}
\end{aligned}
$$

We conclude again by noting that the necessary condition (4.14) holds for $m \geq \max_j m_{0,j}$ and is violated for $m < \max_j m_{0,j}$.

4.7.10 Proof of Corollary 4.3.3

The equivalence between (ι) and $(\iota\iota)$ follows from Corollary (4.3.2). From the proof of Corollary (4.3.2), we also know that, for any j, if $m_{0,j} < +\infty$, then (4.15) holds for the sequence $(d_{j,k})_k$ for any $m \geq m_{0,j}$. Hence,

$$
\begin{aligned}
\sup_j m_{0,j} < +\infty &\implies \text{(4.15) holds for any sequence } (d_{j,k})_k \text{ for any } m \geq m_{0,j} \\
&\implies \text{(4.15) holds for any sequence } (d_{j,k})_k \text{ for any } m \geq \max_j m_{0,j} m_{0,j}
\end{aligned}
$$

Thus, $(\iota\iota)$ implies (ι). The reciprocal is clear.

Regarding the last statement, notice that (X_t) if (m,h)-past-representable for some $m < \max_j m_{0,j}$, there would then exists some j such that $m < m_{0,j}$. Hence, (4.14) does not holds with m for some particular sequence $(d_{j,k})_k$, which is impossible by Lemma 4.3.3.

4.7.11 Proof of Proposition 4.4.1

By Proposition 4.2.2

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

The conclusion follows by considering the points of $B(V)$ and $A \cap B(V)$ that are charged by the spectral measure $\Gamma^{\|\cdot\|}$ in (4.24).

4.7.12 Proof of Lemma 4.4.1

By Proposition 4.3.2, we have

$$\Gamma^{\|\cdot\|} = \sum_{j=1}^{J} \sum_{\vartheta \in S_1} \sum_{k \in \mathbb{Z}} w_{j,\vartheta} \pi_j^{\alpha} \|\boldsymbol{d}_{j,k}\|^{\alpha} \delta_{\left\{\frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|}\right\}},$$

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

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

Therefore,

$$\Gamma^{\|\cdot\|} = \sum_{j=1}^{J} \sum_{\vartheta \in S_1} w_{j,\vartheta} \pi_j^{\alpha} \left[\sum_{k=-m}^{h-1} \|\boldsymbol{d}_{j,k}\|^{\alpha} \delta_{\left\{\frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|}\right\}} + \sum_{k=h}^{+\infty} |\rho_j|^{\alpha(k-h)} \|\boldsymbol{d}_{j,h}\|^{\alpha} \delta_{\left\{\frac{\vartheta \rho_j^{k-h} \boldsymbol{d}_{j,h}}{|\rho_j|^{k-h} \|\boldsymbol{d}_{j,h}\|}\right\}} \right].$$

Moreover,

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

Finally, noticing that for $k = -m$ and any $j \in \{1, \ldots, J\}$, $\boldsymbol{d}_{j,k} = (1, 0, \ldots, 0)$

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

4.7.13 Proof of Proposition 4.4.2

Lemma 4.7.2 *Let $\Gamma^{\|\cdot\|}$ be the spectral measure given in Lemma 4.4.1 and assume that the ρ_j's are all positive.*

Letting $(\vartheta_0, j_0, k_0) \in \mathcal{I}$, consider

$$I_0 := \left\{ \frac{\vartheta' \boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|} : \quad \frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})}{\|\boldsymbol{d}_{j_0,k_0}\|} \;\; \textit{for} \;\; (\vartheta', j', k') \in \mathcal{I} \right\}.$$

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

$$I_0 = \left\{ \frac{\vartheta_0 \boldsymbol{d}_{j_0,k'}}{\|\boldsymbol{d}_{j_0,k'}\|} : \quad 0 \leq k' \leq h \right\}.$$

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

$$I_0 = \begin{cases} \left\{ \dfrac{\vartheta_0 \boldsymbol{d}_{j_0,k_0}}{\|\boldsymbol{d}_{j_0,k_0}\|} \right\}, & \text{if} \quad -m+1 \leq k_0 \leq -1 \\ \\ \left\{ \dfrac{\vartheta_0 \boldsymbol{d}_{0,k_0}}{\|\boldsymbol{d}_{0,k_0}\|} \right\} = \{(\vartheta_0, 0, \ldots, 0)\}, & \text{if} \quad k_0 = -m. \end{cases}$$

For $m = 0$, then

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

Proof.

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

If $k' \in \{-m, \ldots, -1\}$, the $(m+1)^{\text{th}}$ component of $f(\boldsymbol{d}_{j',k'})$ is zero, whereas the $(m+1)^{\text{th}}$ component of $f(\boldsymbol{d}_{j_0,k_0})$ is $\rho_{j_0}^{k_0} \neq 0$. Necessarily, $\vartheta' f(\boldsymbol{d}_{j',k'})/\|\boldsymbol{d}_{j',k'}\| \neq \vartheta_0 f(\boldsymbol{d}_{j_0,k_0})/\|\boldsymbol{d}_{j_0,k_0}\|$ and

$$I_0 = \left\{ \frac{\vartheta' \boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|} : \quad \frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})}{\|\boldsymbol{d}_{j_0,k_0}\|} \ \text{ for } \ (\vartheta', j', k') \in \{-1, +1\} \times \{1, \ldots, J\} \times \{0, \ldots, h\} \right\}.$$

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

$$\begin{aligned} f(\boldsymbol{d}_{j',k'}) &= (\rho_{j'}^{k'+m}, \ldots, \rho_{j'}^{k'+1}, \rho_{j'}^{k'}), \\ f(\boldsymbol{d}_{j_0,k_0}) &= (\rho_{j_0}^{k_0+m}, \ldots, \rho_{j_0}^{k_0+1}, \rho_{j_0}^{k_0}), \end{aligned}$$

and by (4.8) we also have that

$$\begin{aligned} \|\boldsymbol{d}_{j',k'}\| &= \|(\rho_{j'}^{k'+m}, \ldots, \rho_{j'}^{k'+1}, \rho_{j'}^{k'}, \overbrace{0, \ldots, 0}^{h})\|, \\ \|\boldsymbol{d}_{j_0,k_0}\| &= \|(\rho_{j_0}^{k_0+m}, \ldots, \rho_{j_0}^{k_0+1}, \rho_{j_0}^{k_0}, \underbrace{0, \ldots, 0}_{h})\|. \end{aligned}$$

Thus,

$$\begin{aligned} &\frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})}{\|\boldsymbol{d}_{j_0,k_0}\|} \\ &\qquad \iff \frac{\vartheta' \rho_{j'}^{k'} f\left(\boldsymbol{d}_{j',0}\right)}{|\rho_{j'}|^{k'} \|\boldsymbol{d}_{j',0}\|} = \frac{\vartheta_0 \rho_{j_0}^{k_0} f\left(\boldsymbol{d}_{j_0,0}\right)}{|\rho_{j_0}|^{k_0} \|\boldsymbol{d}_{j_0,0}\|} \\ &\qquad \iff \frac{\vartheta' \rho_{j'}^{\ell}}{\|\boldsymbol{d}_{j',0}\|} = \frac{\vartheta_0 \rho_{j_0}^{\ell}}{\|\boldsymbol{d}_{j_0,0}\|}, \qquad \ell = 0, \ldots, m \\ &\qquad \iff \vartheta' \vartheta_0 \frac{\|\boldsymbol{d}_{j_0,0}\|}{\|\boldsymbol{d}_{j',0}\|} = \left(\frac{\rho_{j_0}}{\rho_{j'}} \right)^{\ell}, \quad \ell = 0, \ldots, m \\ &\qquad \iff \rho_{j'} = \rho_{j_0} \quad \text{and} \quad \vartheta' \vartheta_0 = 1 \\ &\qquad \iff j' = j_0 \quad \text{and} \quad \vartheta' = \vartheta_0, \end{aligned}$$

because the ρ_j's are assumed to be non-zero and distinct.

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

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

$$\frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})}{\|\boldsymbol{d}_{j_0,k_0}\|} \Longrightarrow k' = k_0.$$

$$\begin{aligned}
f(\boldsymbol{d}_{j',k'}) &= (\overbrace{\rho_{j'}^{k'+m}, \dots, \rho_{j'}, 1, 0, \dots, 0}^{m+1}, \overbrace{0, \dots, 0}^{h}),\\
f(\boldsymbol{d}_{j_0,k_0}) &= (\underbrace{\rho_{j_0}^{k_0+m}, \dots, \rho_{j_0}, 1, 0, \dots, 0}_{m+1}, \underbrace{0, \dots, 0}_{h}),
\end{aligned}$$

and we also have that

$$\begin{aligned}
\|\boldsymbol{d}_{j',k'}\| &= \|(\overbrace{\rho_{j'}^{k'+m}, \dots, \rho_{j'}, 1, 0, \dots, 0}^{m+1}, \overbrace{0, \dots, 0}^{h})\|,\\
\|\boldsymbol{d}_{j_0,k_0}\| &= \|(\underbrace{\rho_{j_0}^{k_0+m}, \dots, \rho_{j_0}, 1, 0, \dots, 0}_{m+1}, \underbrace{0, \dots, 0}_{h})\|.
\end{aligned}$$

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

$$\begin{aligned}
\frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} &= \frac{\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})}{\|\boldsymbol{d}_{j_0,k_0}\|}\\
&\Longleftrightarrow \ \frac{\vartheta' \rho_{j'}^{\ell}}{\|\boldsymbol{d}_{j',k_0}\|} = \frac{\vartheta_0 \rho_{j_0}^{\ell}}{\|\boldsymbol{d}_{j_0,k_0}\|}, \quad \ell = 0, \dots, m + k_0, \ \text{ and } \ k' = k_0\\
&\Longleftrightarrow \ \vartheta' \vartheta_0 \frac{\|\boldsymbol{d}_{j_0,k_0}\|}{\|\boldsymbol{d}_{j',k_0}\|} = \left(\frac{\rho_{j_0}}{\rho_{j'}}\right)^{\ell}, \quad \ell = 0, \dots, m + k_0, \ \text{ and } \ k' = k_0.
\end{aligned}$$

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

$$\begin{aligned}
\vartheta' \vartheta_0 \frac{\|\boldsymbol{d}_{j_0,k_0}\|}{\|\boldsymbol{d}_{j',k_0}\|} = \left(\frac{\rho_{j_0}}{\rho_{j'}}\right)^{\ell}, \quad \ell = 0, 1, \dots, m + k_0, \ \text{ and } \ k' = k_0\\
\Longleftrightarrow \ \vartheta' = \vartheta_0 \quad \text{and} \quad j' = j_0 \quad \text{and} \quad k' = k_0.
\end{aligned}$$

If $k_0 = -m$, given that $(\vartheta_0, j_0, k_0) \in \mathcal{I} = S_1 \times \Big(\{1, \dots, J\} \times \{-m, \dots, -1, 0, 1, \dots, h\} \cup \{(0, -m)\}\Big)$, then necessarily $j_0 = 0$. Furthermore, as $k' = k_0 = -m$, we similarly have that $j' = j_0 = 0$ and thus $\boldsymbol{d}_{j',k_0} = \boldsymbol{d}_{j_0,k_0} = \boldsymbol{d}_{0,-m} = (1, 0, \dots, 0)$. Hence

$$\begin{aligned}
\vartheta' \vartheta_0 \frac{\|\boldsymbol{d}_{j_0,k_0}\|}{\|\boldsymbol{d}_{j',k_0}\|} = \left(\frac{\rho_{j_0}}{\rho_{j'}}\right)^{\ell}, \quad \ell = 0, \ \text{ and } \ k' = k_0 = -m \ \text{ and } \ j' = j_0 = 0,\\
\Longleftrightarrow \ \vartheta' = \vartheta_0 \ \text{ and } \ k' = k_0 = -m \ \text{ and } \ j' = j_0 = 0
\end{aligned}$$

Case $\boldsymbol{m = 0}$

If $k_0 \in \{1, \dots, h\}$ then $f(\boldsymbol{d}_{j_0,k_0}) = \rho_{j_0}^{k_0}$ and by (4.8), $\|\boldsymbol{d}_{j_0,k_0}\| = |\rho_{j_0}|^{k_0}$. Thus, $\vartheta_0 f(\boldsymbol{d}_{j_0,k_0}) / \|\boldsymbol{d}_{j_0,k_0}\| = \vartheta_0$. If $k_0 = -m = 0$, then $j_0 = 0$ and $f(\boldsymbol{d}_{j_0,k_0}) = 1$ and $\vartheta_0 f(\boldsymbol{d}_{j_0,k_0}) / \|\boldsymbol{d}_{j_0,k_0}\| = \vartheta_0$. The same holds for $(\vartheta', j', k') \in \mathcal{I}$ and we obtain that

$$\frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})}{\|\boldsymbol{d}_{j_0,k_0}\|} \ \Longleftrightarrow \ \vartheta' = \vartheta_0.$$

$\square$

Let us now prove Proposition 4.4.2. By Proposition 4.4.1,

$$\mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t}, A_{\vartheta,j,k}\Big|B(V_0)\Big) \underset{x\to\infty}{\longrightarrow} \frac{\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta'\boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|} \in A_{\vartheta,j,k}: \quad \frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} \in V_0\right\}\right)}{\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta'\boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|} \in C^{\|\cdot\|}_{m+h+1}: \quad \frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} \in V_0\right\}\right)}. \tag{4.49}$$

Focusing on the denominator, we have by (4.28)

$$\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta'\boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|} \in C^{\|\cdot\|}_{m+h+1}: \quad \frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} \in V_0\right\}\right) = \Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta'\boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|} \in C^{\|\cdot\|}_{m+h+1}: \quad \frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})}{\|\boldsymbol{d}_{j_0,k_0}\|}\right\}\right)$$

We will now distinguish the cases arising from the application of Lemma 4.7.2. Recall that we assume for this proposition that the ρ_j's are positive. Thus, $\operatorname{sign}(\rho_j) = 1$ and $\bar{\beta}_j = \beta_j \frac{1-|\rho_j|^\alpha}{1-\rho_j^{<\alpha>}} = \beta_j$ and $\bar{w}_{j,\vartheta} = w_{j,\vartheta}$ in (4.27) for all j's and $\vartheta \in \{-1,+1\}$.

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

By Lemma 4.7.2,

$$\begin{aligned}
\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta'\boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|} \in C^{\|\cdot\|}_{m+h+1}: \quad \frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})}{\|\boldsymbol{d}_{j_0,k_0}\|}\right\}\right)& \\
= \Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta_0\boldsymbol{d}_{j_0,k'}}{\|\boldsymbol{d}_{j_0,k'}\|}: \quad 0 \leq k' \leq h\right\}\right)& \\
= \pi_{j_0}^\alpha\left[w_{j_0,\vartheta_0}\sum_{k'=0}^{h-1}\|\boldsymbol{d}_{j_0,k'}\|^\alpha + \frac{\bar{w}_{j_0,\vartheta_0}}{1-|\rho_j|^\alpha}\|\boldsymbol{d}_{j_0,h}\|^\alpha\right]&
\end{aligned}$$

By (4.8), for $k' \in \{0,1,\ldots,h\}$

$$\begin{aligned}
\|\boldsymbol{d}_{j_0,k'}\| &= \|(\rho_{j_0}^{k'+m},\ldots,\rho_{j_0}^{k'+1},\rho_{j_0}^{k'},\underbrace{0,\ldots,0}_{h})\| \\
&= |\rho_{j_0}|^{k'-h}\|(\rho_{j_0}^{m+h},\ldots,\rho_{j_0}^{h+1},\rho_{j_0}^{h},\underbrace{0,\ldots,0}_{h})\| \\
&= |\rho_{j_0}|^{k'-h}\|\boldsymbol{d}_{j_0,h}\|.
\end{aligned}$$

Thus,

$$\begin{aligned}
\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta'\boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|} \in C^{\|\cdot\|}_{m+h+1}: \quad \frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})}{\|\boldsymbol{d}_{j_0,k_0}\|}\right\}\right) &= \pi_{j_0}^\alpha w_{j_0,\vartheta_0}\|\boldsymbol{d}_{j_0,h}\|^\alpha\left[\sum_{k'=0}^{h-1}\rho_{j_0}^{\alpha(k'-h)} + \frac{1}{1-|\rho_j|^\alpha}\right] \\
&= \pi_{j_0}^\alpha w_{j_0,\vartheta_0}\|\boldsymbol{d}_{j_0,h}\|^\alpha\frac{|\rho_j|^{-\alpha h}}{1-|\rho_j|^\alpha}.
\end{aligned}$$

Similarly for the numerator in (4.49), by (4.29),

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

$$= \Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta_0 \boldsymbol{d}_{j_0,k'}}{\|\boldsymbol{d}_{j_0,k'}\|} \in A_{\vartheta,j,k}: \quad 0 \le k' \le h\right\}\right)$$

$$= \begin{cases} \Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta_0 \boldsymbol{d}_{j_0,k}}{\|\boldsymbol{d}_{j_0,k}\|}\right\}\right), & \text{if} \quad j = j_0 \quad \text{and} \quad \vartheta = \vartheta_0, \\ \Gamma^{\|\cdot\|}(\emptyset), & \text{if} \quad j \neq j_0 \quad \text{or} \quad \vartheta \neq \vartheta_0, \end{cases}$$

$$= \begin{cases} \pi_{j_0}^{\alpha} w_{j_0,\vartheta_0} \|\boldsymbol{d}_{j_0,h}\|^{\alpha} |\rho_{j_0}|^{\alpha(k-h)} \delta_{\{\vartheta_0\}}(\vartheta)\delta_{\{j_0\}}(j), & \text{if} \quad 0 \le k \le h-1, \\ \pi_{j_0}^{\alpha} w_{j_0,\vartheta_0} \|\boldsymbol{d}_{j_0,h}\|^{\alpha} \frac{1}{1-|\rho_{j_0}|^{\alpha}} \delta_{\{\vartheta_0\}}(\vartheta)\delta_{\{j_0\}}(j), & \text{if} \quad k = h. \end{cases}$$

The conclusion follows.

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

We have by Lemma 4.7.2

$$\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta' \boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|} \in C_{m+h+1}^{\|\cdot\|}: \quad \frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})}{\|\boldsymbol{d}_{j_0,k_0}\|}\right\}\right) = \Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta_0 \boldsymbol{d}_{j_0,k_0}}{\|\boldsymbol{d}_{j_0,k_0}\|}\right\}\right). \qquad =$$

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

$$\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta_0 \boldsymbol{d}_{j_0,k_0}}{\|\boldsymbol{d}_{j_0,k_0}\|}\right\}\right) = \pi_{j_0}^{\alpha} w_{j_0,\vartheta_0} \|\boldsymbol{d}_{j_0,k_0}\|^{\alpha},$$

and

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

$$= \Gamma^{\|\cdot\|}\left(A_{\vartheta,j,k} \cap \left\{\frac{\vartheta_0 \boldsymbol{d}_{j_0,k_0}}{\|\boldsymbol{d}_{j_0,k_0}\|}\right\}\right)$$

$$= \begin{cases} \Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta_0 \boldsymbol{d}_{j_0,k_0}}{\|\boldsymbol{d}_{j_0,k_0}\|}\right\}\right), & \text{if} \quad j = j_0 \quad \text{and} \quad \vartheta = \vartheta_0, \quad \text{and} \quad k = k_0, \\ \Gamma^{\|\cdot\|}(\emptyset), & \text{if} \quad j \neq j_0 \quad \text{or} \quad \vartheta \neq \vartheta_0 \quad \text{or} \quad k \neq k_0, \end{cases}$$

$$= \pi_{j_0}^{\alpha} w_{j_0,\vartheta_0} \|\boldsymbol{d}_{j_0,k_0}\|^{\alpha} \delta_{\{\vartheta_0\}}(\vartheta)\delta_{\{j_0\}}(j)\delta_{\{k_0\}}(k).$$

If $k_0 = -m$, then $\boldsymbol{d}_{j_0,k_0} = \boldsymbol{d}_{0,-m} = (1, 0, \ldots, 0)$, and

$$\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta_0 \boldsymbol{d}_{j_0,k_0}}{\|\boldsymbol{d}_{j_0,k_0}\|}\right\}\right) = \Gamma^{\|\cdot\|}\Big(\{\vartheta_0(1, 0, \ldots, 0)\}\Big) = w_{\vartheta_0},$$

and

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

$$= \Gamma^{\|\cdot\|}\left(A_{\vartheta,j,k} \cap \left\{\frac{\vartheta_0 \boldsymbol{d}_{j_0,k_0}}{\|\boldsymbol{d}_{j_0,k_0}\|}\right\}\right)$$

$$= \begin{cases} \Gamma^{\|\cdot\|}\Big(A_{\vartheta,j,k} \cap \{\vartheta_0(1,0,\ldots,0)\}\Big), & \text{if } \vartheta = \vartheta_0, \text{ and } k = k_0 = -m, \text{ and } j = j_0 = 0 \\ \Gamma^{\|\cdot\|}(\emptyset), & \text{if } \vartheta \neq \vartheta_0 \text{ or } k \neq k_0, \text{ or } j \neq j_0 \end{cases}$$

$$= w_{\vartheta_0}\delta_{\{\vartheta_0\}}(\vartheta)\delta_{\{j_0\}}(j)\delta_{\{k_0\}}(k).$$

The conclusion follows as previously.

Case $m = 0$

By Lemma 4.7.2, as the ρ_j's are positive

$$\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta' \boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|} \in C^{\|\cdot\|}_{m+h+1} : \quad \frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})}{\|\boldsymbol{d}_{j_0,k_0}\|}\right\}\right)$$

$$= \Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta_0 \boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|} \in C^{\|\cdot\|}_{m+h+1} : \quad (j',k') \in \{1,\ldots,J\} \times \{0,\ldots,h\} \cup \{(0,0)\}\right\}\right)$$

Given that $w_{\vartheta_0} = \sum_{j'=1}^{J} \pi_j^\alpha w_{j',\vartheta_0}$ and $\|\boldsymbol{d}_{j',k'}\| = |\rho_{j'}|^{k'}$, for any $1 \leq j' \leq J$, $1 \leq k' \leq h$,

$$\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta' \boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|} \in C^{\|\cdot\|}_{m+h+1} : \quad \frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})}{\|\boldsymbol{d}_{j_0,k_0}\|}\right\}\right)$$

$$\begin{aligned}
&= w_{\vartheta_0} + \sum_{j'=1}^{J} \pi_j^\alpha w_{j',\vartheta_0}\left[\sum_{k'=1}^{h-1} \|\boldsymbol{d}_{j',k'}\|^\alpha + \frac{\|\boldsymbol{d}_{j',h}\|^\alpha}{1-|\rho_{j'}|^\alpha}\right] \\
&= \sum_{j'=1}^{J} \pi_j^\alpha w_{j',\vartheta_0}\left[1 + \sum_{k'=1}^{h-1} |\rho_{j'}|^{\alpha k'} + \frac{|\rho_{j'}|^{\alpha h}}{1-|\rho_{j'}|^\alpha}\right] \\
&= \sum_{j'=1}^{J} \pi_{j'}^\alpha w_{j',\vartheta_0}\left[\frac{1-|\rho_{j'}|^{\alpha h}}{1-|\rho_{j'}|^\alpha} + \frac{|\rho_{j'}|^{\alpha h}}{1-|\rho_{j'}|^\alpha}\right] \\
&= \sum_{j'=1}^{J} \pi_{j'}^\alpha w_{j',\vartheta_0}\frac{1}{1-|\rho_{j'}|^\alpha}.
\end{aligned}$$

Similarly, by (4.29),

$$
\begin{aligned}
\Gamma^{\|\cdot\|}&\left(\left\{\frac{\vartheta' \boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|}\in A_{\vartheta,j,k}:\quad \frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|}\in V_0\right\}\right)\\
&\qquad\qquad = \Gamma^{\|\cdot\|}\left(A_{\vartheta,j,k}\cap\left\{\frac{\vartheta_0 \boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|}\in C^{\|\cdot\|}_{m+h+1}:\quad (j',k')\in\{1,\dots,J\}\times\{0,\dots,h\}\cup\{(0,0)\}\right\}\right)\\
&\qquad\qquad = \begin{cases} \Gamma^{\|\cdot\|}\left(\left\{\dfrac{\vartheta_0 \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|}\right\}\right), & \text{if } \vartheta=\vartheta_0,\\ \Gamma^{\|\cdot\|}(\emptyset), & \text{if } \vartheta\neq\vartheta_0,\end{cases}\\
&\qquad\qquad = \begin{cases} \sum_{j'=1}^{J}\pi_{j'}^{\alpha} w_{j',\vartheta_0}\delta_{\{\vartheta_0\}}(\vartheta), & \text{if}\quad k=0,\\ \pi_j^{\alpha} w_{j,\vartheta_0}|\rho_j|^{\alpha k}\delta_{\{\vartheta_0\}}(\vartheta), & \text{if}\quad 1\le k\le h-1,\\ \pi_j^{\alpha} w_{j,\vartheta_0}\dfrac{|\rho_j|^{\alpha h}}{1-|\rho_j|^{\alpha}}\delta_{\{\vartheta_0\}}(\vartheta), & \text{if}\quad k=h.\end{cases}
\end{aligned}
$$

The conclusion follows.

4.7.14 Proof of Proposition 4.4.3

Lemma 4.7.3 *Let X_t be the α-stable anticipative AR(2) (resp. fractionally integrated AR) as in* (4.30) *(resp.* (4.32)*). With f as in* (4.23)*, and for any $m\ge 1$, $h\ge 0$,*

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

Proof.

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

$$
\begin{aligned}
\frac{f(\vartheta_1\boldsymbol{d}_k)}{\|\boldsymbol{d}_k\|}=\frac{f(\vartheta_2\boldsymbol{d}_\ell)}{\|\boldsymbol{d}_\ell\|} &\iff \left[\forall\, i=0,\dots,m,\quad \frac{\vartheta_1 d_{k+i}}{\|\boldsymbol{d}_k\|}=\frac{\vartheta_2 d_{\ell+i}}{\|\boldsymbol{d}_\ell\|}\right]\\
&\iff \frac{d_k}{d_\ell}=\frac{d_{k+1}}{d_{\ell+1}}=\ldots=\vartheta_1\vartheta_2\frac{\|\boldsymbol{d}_k\|}{\|\boldsymbol{d}_\ell\|}.
\end{aligned} \tag{4.50}
$$

The last statement in particular implies that $\dfrac{d_k}{d_\ell}=\dfrac{d_{k+1}}{d_{\ell+1}}$.
For the anticipative AR(2), if $\lambda_1\neq\lambda_2$, we then have

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

This case $\lambda_1=\lambda_2=\lambda$ is similar. For the anticipative fractionally integrated AR, given that $\Gamma(z+1)=z\Gamma(z)$

for any $z \in \mathbb{C}$, we have

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

Therefore, in all cases,

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

$\square$

Let us now prove Proposition 4.4.3. The spectral measure of $\boldsymbol{X}_t$ writes

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

where the sequences (d_k) are given respectively by (4.31) and (4.33) for the anticipative AR(2) and fractionally integrated processes. By Proposition 4.2.2,

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

On the one hand, we have by definition of $B(V_0)$, V_0 and Lemma 4.7.3,

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

Similarly, it is easily shown that

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

The conclusion follows.

4.7.15 Proof of Proposition 4.5.1

We start with a lemma showing that $\underline{\boldsymbol{X}}_t$ is indeed $S\alpha S$ and providing the form of its spectral measure on the Euclidean unit sphere. The representability condition on the unit cylinder $C_4^{\|\cdot\|}$ will follow.

Lemma 4.7.4 *Let $(\boldsymbol{X}_t)$ as in (4.34). Then, the vector $\underline{\boldsymbol{X}}_t = (X_{1,t}, X_{2,t}, X_{1,t+1}, X_{2,t+1})'$ is $S\alpha S$ with zero shift vector and spectral measure given by*

$$\Gamma_4 = \Delta + \Gamma_{4,1} + \Gamma_{4,2}.$$

Here,

$$\Delta = \sum_{i=1,2} \frac{\sigma_i^\alpha}{2}(1+\rho_i^2)^{\alpha/2}\frac{|\rho_i|^\alpha}{1-|\rho_i|^\alpha}(\delta_{\{\boldsymbol{x}_i/\|\boldsymbol{x}_i\|_e\}} + \delta_{\{-\boldsymbol{x}_i/\|\boldsymbol{x}_i\|_e\}}),$$

with $\sigma_i^\alpha := \int_{S_2} |s_i|^\alpha \Gamma_2(d\boldsymbol{s})$, points $\boldsymbol{x}_1 = (1, 0, \rho_1^{-1}, 0)$, $\boldsymbol{x}_2 = (0, 1, 0, \rho_2)$,

$$\Gamma_{4,1}(d\boldsymbol{s}) = \|B\boldsymbol{s}\|_e^{-\alpha}\tilde{\Gamma}_{4,1} \circ T_B(d\boldsymbol{s}),$$

$$\Gamma_{4,2}(d\boldsymbol{s}) = \|C\boldsymbol{s}\|_e^{-\alpha}\tilde{\Gamma}_{4,2} \circ T_C(d\boldsymbol{s}),$$

with $T_A : S_4 \to S_4$ is defined by $T_A(\underline{\boldsymbol{s}}) = A\underline{\boldsymbol{s}}/\|A\underline{\boldsymbol{s}}\|_e$, for any invertible matrix A of dimension 4,

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -\rho_2 & 0 & 1 \end{pmatrix}, \quad \textit{and} \quad C = \begin{pmatrix} 1 & 0 & -\rho_1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

and $\tilde{\Gamma}_{4,i}(\,\cdot\,) = \Gamma_2 \circ h_i(S_{4,i} \cap \,\cdot\,)$, $i = 1, 2$, where $S_{4,1} = \{(s_1, s_2, 0, 0) \in S_4 : \ (s_1, s_2) \in S_2\}$, $S_{4,2} = \{(0, 0, s_3, s_4) \in S_4 : \ (s_3, s_4) \in S_2\}$, and $h_1, h_2 : S_4 \to S_2$ are the functions defined by $h_1((s_1, s_2, s_3, s_4)) = (s_1, s_2)$ and $h_2((s_1, s_2, s_3, s_4)) = (s_3, s_4)$.

Proof.

Let $\underline{\boldsymbol{u}} = (\boldsymbol{u}_0', \boldsymbol{u}_1')' \in \mathbb{R}^4$ with $\boldsymbol{u}_i = (u_{1,i}, u_{2,i})'$, $i = 0, 1$. The characteristic function of $\underline{\boldsymbol{X}}_t$ reads

$$\begin{aligned} \varphi(\underline{\boldsymbol{u}}) &:= \mathbb{E}[\exp\{i\langle \underline{\boldsymbol{u}}, \underline{\boldsymbol{X}}_t\rangle\}] = \mathbb{E}[\exp\{i\sum_{j=0}^{1}\langle \boldsymbol{u}_j, \boldsymbol{X}_{t+j}\rangle\}] = \mathbb{E}[\exp\{i\sum_{k\in\mathbb{Z}}\sum_{j=0}^{1}\langle \boldsymbol{u}_j, \boldsymbol{A}_k\boldsymbol{\varepsilon}_{t+k+j}\rangle\}] \\ &= \prod_{k\in\mathbb{Z}} \mathbb{E}[\exp\{i\langle\sum_{j=0}^{1} \boldsymbol{A}_{k-j}'\boldsymbol{u}_j, \boldsymbol{\varepsilon}_{t+k}\rangle\}], \end{aligned}$$

where for all $k \in \mathbb{Z}$

$$A_k = \begin{pmatrix} \rho_1^k \mathbb{1}_{\{k\geq 0\}} & 0 \\ 0 & \rho_2^{-k}\mathbb{1}_{\{k\leq 0\}} \end{pmatrix}.$$

Thus,

$$\begin{aligned} -\ln\,\varphi(\underline{\boldsymbol{u}}) &= \sum_{k\in\mathbb{Z}}\int_{S_2} |\langle A_k\boldsymbol{u}_0 + A_{k-1}\boldsymbol{u}_1, \boldsymbol{s}\rangle|^\alpha \Gamma_2(d\boldsymbol{s}) \\ &= \sum_{k\leq -1}\int_{S_2} |\rho_2^{-k}(u_{2,0} + \rho_2 u_{2,1})s_2|^\alpha\Gamma_2(d\boldsymbol{s}) + \sum_{k\geq 2}\int_{S_2} |\rho_1^{k-1}(\rho_1 u_{1,0} + u_{1,1})s_1|^\alpha\Gamma_2(d\boldsymbol{s}) \\ &\quad + \int_{S_2} |u_{1,0}s_1 + (u_{2,0} + \rho_2 u_{2,1})s_2|^\alpha\Gamma_2(d\boldsymbol{s}) + \int_{S_2} |(\rho_1 u_{1,0} + u_{1,1})s_1 + u_{2,1}s_2|^\alpha\Gamma_2(d\boldsymbol{s}) \\ &= \sigma_2^\alpha\frac{|\rho_2|^\alpha}{1-|\rho_2|^\alpha}|u_{2,0} + \rho_2 u_{2,1}|^\alpha + \sigma_1^\alpha\frac{|\rho_1|^\alpha}{1-|\rho_1|^\alpha}|\rho_1 u_{1,0} + u_{1,1}|^\alpha \\ &\quad + \int_{S_2} |u_{1,0}s_1 + (u_{2,0} + \rho_2 u_{2,1})s_2|^\alpha\Gamma_2(d\boldsymbol{s}) + \int_{S_2} |(\rho_1 u_{1,0} + u_{1,1})s_1 + u_{2,1}s_2|^\alpha\Gamma_2(d\boldsymbol{s}), \end{aligned} \tag{4.51}$$

where $\sigma_i^\alpha := \int_{S_2} |s_i|^\alpha \Gamma_2(d\boldsymbol{s})$, $i = 1, 2$. We notice that the characteristic function of $\underline{\boldsymbol{X}}_t$ is real. Hence, $\underline{\boldsymbol{X}}_t$ being α-stable is equivalent to $\underline{\boldsymbol{X}}_t$ being symmetric α-stable, and therefore, by Theorem 2.4.3 in [117], $\underline{\boldsymbol{X}}_t$ will be α-stable if and only if there exists a unique symmetric finite measure Γ_4 on the Euclidean unit sphere such that

$$-\ln \varphi(\underline{\boldsymbol{u}}) = \int_{S_4} |\langle \underline{\boldsymbol{u}}, \underline{\boldsymbol{s}} \rangle| \Gamma_4(d\boldsymbol{s}). \tag{4.52}$$

We will thus rewrite (4.51) to exhibit such a symmetric measure. The two first terms are easily rewritten with charged atoms on S_4: for all $\underline{\boldsymbol{u}} \in \mathbb{R}^4$,

$$\sigma_2^\alpha \frac{|\rho_2|^\alpha}{1-|\rho_2|^\alpha} |u_{2,0} + \rho_2 u_{2,1}|^\alpha + \sigma_1^\alpha \frac{|\rho_1|^\alpha}{1-|\rho_1|^\alpha} |\rho_1 u_{1,0} + u_{1,1}|^\alpha = \int_{S_4} |\langle \underline{\boldsymbol{u}}, \underline{\boldsymbol{s}} \rangle|^\alpha \Delta(d\underline{\boldsymbol{s}}), \tag{4.53}$$

where $\Delta = \sum_{i=1,2} \frac{\sigma_i^\alpha}{2} (1+\rho_i^2)^{\alpha/2} \frac{|\rho_i|^\alpha}{1-|\rho_i|^\alpha} (\delta_{\{\boldsymbol{x}_i/\|\boldsymbol{x}_i\|_e\}} + \delta_{\{-\boldsymbol{x}_i/\|\boldsymbol{x}_i\|_e\}})$. The third and fourth terms in (4.51) can also be rewritten as integral over S_4. Starting with the third term, notice that the integral over S_2 can be seen as an integral over S_4 with a spectral measure $\tilde{\Gamma}_{4,1}$ coinciding with Γ_2 on $S_{4,1} = \{(s_1, s_2, 0, 0) \in S_4 : (s_1, s_2) \in S_2\}$ and having zero mass outside:

$$\int_{S_2} |u_{1,0}s_1 + (u_{2,0} + \rho_2 u_{2,1})s_2|^\alpha \Gamma_2(d\boldsymbol{s}) = \int_{S_4} |u_{1,0}s_1 + (u_{2,0} + \rho_2 u_{2,1})s_2 + u_{1,1}s_3 + u_{2,1}s_4|^\alpha \tilde{\Gamma}_{4,1}(d\underline{\boldsymbol{s}}),$$

with $\tilde{\Gamma}_{4,1}(\,\cdot\,) = \Gamma_2 \circ h_1(S_{4,1} \cap \,\cdot\,)$, where $h_1 : S_4 \to S_2$ is the function defined by $h_1((s_1, s_2, s_3, s_4)) = (s_1, s_2)$. Thus,

$$\begin{aligned}\int_{S_2} |u_{1,0}s_1 + (u_{2,0} + \rho_2 u_{2,1})s_2|^\alpha \Gamma_2(d\boldsymbol{s}) &= \int_{S_4} |\langle \boldsymbol{b}\underline{\boldsymbol{u}}, \underline{\boldsymbol{s}} \rangle|^\alpha \tilde{\Gamma}_{4,1}(d\underline{\boldsymbol{s}}) \\ &= \int_{S_4} |\langle \underline{\boldsymbol{u}}, \boldsymbol{b}'\underline{\boldsymbol{s}} \rangle|^\alpha \tilde{\Gamma}_{4,1}(d\underline{\boldsymbol{s}}),\end{aligned} \tag{4.54}$$

with

$$\boldsymbol{b} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \rho_2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

As the matrix $\boldsymbol{b}'$ is invertible and $B = \boldsymbol{b}'^{-1}$, where B is as stated in the lemma, we notice that $T_{\boldsymbol{b}'}^{-1} = T_{\boldsymbol{b}'^{-1}} = T_B$, where $T_{\boldsymbol{b}'} : S_4 \to S_4$ is the transformation such that $T_{\boldsymbol{b}'}(\underline{\boldsymbol{s}}) = \boldsymbol{b}'\underline{\boldsymbol{s}}/\|\boldsymbol{b}'\underline{\boldsymbol{s}}\|_e$. Performing a change of variable in (4.54) using $T_{\boldsymbol{b}'}$, we get

$$\begin{aligned}\int_{S_2} |u_{1,0}s_1 + (u_{2,0} + \rho_2 u_{2,1})s_2|^\alpha \Gamma_2(d\boldsymbol{s}) &= \int_{S_4} |\langle \underline{\boldsymbol{u}}, \underline{\boldsymbol{s}} \rangle|^\alpha \|\boldsymbol{b}'^{-1}\underline{\boldsymbol{s}}\|^{-\alpha} \tilde{\Gamma}_{4,1} \circ T_{\boldsymbol{b}'^{-1}}(d\underline{\boldsymbol{s}}) \\ &= \int_{S_4} |\langle \underline{\boldsymbol{u}}, \underline{\boldsymbol{s}} \rangle|^\alpha \Gamma_{4,1}(d\underline{\boldsymbol{s}}).\end{aligned} \tag{4.55}$$

Similarly for the fourth term in (4.51), we have

$$\begin{aligned}\int_{S_2} |(\rho_1 u_{1,0} + u_{1,1})s_1 + u_{2,1}s_2|^\alpha \Gamma_2(d\boldsymbol{s}) &= \int_{S_4} |u_{1,0}s_1 + u_{2,0}s_2 + (\rho_1 u_{1,0} + u_{1,1})s_3 + u_{2,1}s_4|^\alpha \tilde{\Gamma}_{4,2}(d\boldsymbol{s}) \\ &= \int_{S_4} |\langle \boldsymbol{c}\underline{\boldsymbol{u}}, \underline{\boldsymbol{s}} \rangle|^\alpha \tilde{\Gamma}_{4,2}(d\underline{\boldsymbol{s}}) \\ &= \int_{S_4} |\langle \underline{\boldsymbol{u}}, \boldsymbol{c}'\underline{\boldsymbol{s}} \rangle|^\alpha \tilde{\Gamma}_{4,2}(d\underline{\boldsymbol{s}}),\end{aligned}$$

with $\tilde{\Gamma}_{4,2}(\,\cdot\,) = \Gamma_2 \circ h_2(S_{4,2} \cap \,\cdot\,)$, where $h_2 : S_4 \to S_2$ is the function defined by $h_2((s_1, s_2, s_3, s_4)) = (s_3, s_4)$, $S_{4,2} = \{(0, 0, s_3, s_4) \in S_4 : \ (s_3, s_4) \in S_2\}$, and

$$\boldsymbol{c} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \rho_1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

With a change of variable using $T_{\boldsymbol{c}'}$, and since $\boldsymbol{c}'^{-1} = C$,

$$\begin{aligned} \int_{S_2} |(\rho_1 u_{1,0} + u_{1,1}) s_1 + u_{2,1} s_2|^\alpha \Gamma_2(d\boldsymbol{s}) &= \int_{S_4} |\langle \underline{\boldsymbol{u}}, \underline{\boldsymbol{s}} \rangle|^\alpha \|\boldsymbol{c}'^{-1} \underline{\boldsymbol{s}}\|^{-\alpha} \tilde{\Gamma}_{4,2} \circ T_{\boldsymbol{c}'^{-1}}(d\underline{\boldsymbol{s}}) \\ &= \int_{S_4} |\langle \underline{\boldsymbol{u}}, \underline{\boldsymbol{s}} \rangle|^\alpha \Gamma_{4,2}(d\underline{\boldsymbol{s}}). \end{aligned} \tag{4.56}$$

Finally, by (4.51), (4.53), (4.55) and (4.56), we have that (4.52) holds with $\Gamma_4 = \Delta + \Gamma_{4,1} + \Gamma_{4,2}$. One can check that Γ_4 is indeed symmetric: for any transformation g among $\{T_B, T_C, h_1, h_2\}$, we have $g(-\boldsymbol{s}) = -g(\boldsymbol{s})$ for any $\boldsymbol{s} \in S_4$, and as Γ_2 is symmetric by assumption, it is easy to check that the measures $\Gamma_{4,1}$ and $\Gamma_{4,2}$ are also symmetric. The case of Δ is obvious. □

Return to the proof of Proposition 4.5.1

By Lemma 4.7.4 and Proposition 4.2.1, we know that $\underline{\boldsymbol{X}}_t$ will be representable on $C_4^{\|\cdot\|}$ if and only if

$$\Gamma_4(K^{\|\cdot\|}) = 0,$$

where $K^{\|\cdot\|} = \{\underline{\boldsymbol{s}} \in S^4 : \ \|\underline{\boldsymbol{s}}\| = 0\} = \{\underline{\boldsymbol{s}} \in S^4 : \ \underline{s}_1 = \underline{s}_2 = 0\} = S_{4,2}$. We have

$$\Gamma_4(K^{\|\cdot\|}) = \Delta(K^{\|\cdot\|}) + \Gamma_{4,1}(K^{\|\cdot\|}) + \Gamma_{4,2}(K^{\|\cdot\|}),$$

with Δ, $\Gamma_{4,1}$ and $\Gamma_{4,2}$ are as in Lemma 4.7.4. Given the points charged by Δ, it is easily seen that $\Delta(K^{\|\cdot\|}) = 0$. Turning to $\Gamma_{4,1}(K^{\|\cdot\|})$, we have

$$\Gamma_{4,1}(K^{\|\cdot\|}) = \int_{K^{\|\cdot\|}} \|B\boldsymbol{s}\|_e^{-\alpha} \tilde{\Gamma}_{4,1} \circ T_B(d\boldsymbol{s}) = \tilde{\Gamma}_{4,1} \circ T_B(K^{\|\cdot\|}).$$

Given that $T_B(K^{\|\cdot\|}) = K^{\|\cdot\|}$ and $S_{4,1} \cap K^{\|\cdot\|} = \emptyset$, we have $\tilde{\Gamma}_{4,1} = \Gamma_2 \circ h_1(S_{4,1} \cap K^{\|\cdot\|}) = 0$. Last, we have

$$\Gamma_{4,2}(K^{\|\cdot\|}) = \int_{K^{\|\cdot\|}} \|C\boldsymbol{s}\|_e^{-\alpha} \tilde{\Gamma}_{4,2} \circ T_C(d\boldsymbol{s}) = \tilde{\Gamma}_{4,2} \circ T_C(K^{\|\cdot\|}),$$

with

$$\begin{aligned} T_C(K^{\|\cdot\|}) &= \left\{ \frac{(-\rho_1 s_3, 0, s_3, s_4)}{(1+\rho_1^2) s_3^2 + s_4^2} : \ \forall (s_3, s_4) \in \mathbb{R}^2, \ s_3^2 + s_4^2 = 1 \right\} \\ &= \left\{ \frac{(-\rho_1 s_3, 0, s_3, s_4)}{(1+\rho_1^2) s_3^2 + s_4^2} : \ \forall (s_3, s_4) \in \mathbb{R}^2, \ s_3^2 + s_4^2 = 1 \ \text{ and } \ s_3 \neq 0 \right\} \cup \{(0,0,0,\pm 1)\} \\ &:= K' \cup \{(0,0,0,-1), (0,0,0,+1)\} \end{aligned}$$

Given that $K' \cap S_{4,2} = \emptyset$, we have by the σ-additivity of $\tilde{\Gamma}_{4,2}$,

$$\begin{aligned}\tilde{\Gamma}_{4,2}\Big(T_C(K^{\|\cdot\|})\Big) &= \tilde{\Gamma}_{4,2}(K') + \tilde{\Gamma}_{4,2}\Big(\{(0,0,0,-1),(0,0,0,+1)\}\Big)\\ &= \Gamma_2 \circ h_2(S_{4,2} \cap K') + \Gamma_2 \circ h_2(\{(0,0,0,-1),(0,0,0,+1)\})\\ &= \Gamma_2(\{(0,-1),(0,+1)\})\end{aligned}$$

Hence, $\Gamma_4(K^{\|\cdot\|}) = \Gamma_2\Big(\{(0,-1),(0,+1)\}\Big)$ and the representability condition for $\underline{\boldsymbol{X}}_t$ follows.

4.7.16 Proof of Proposition 4.5.2

To prove Proposition 4.5.2, will make use of Proposition 4.2.2. We first provide the form of the spectral representation of $\underline{\boldsymbol{X}}_t$ on $C_4^{\|\cdot\|}$ in the next lemma.

Lemma 4.7.5 *Under the assumptions of Proposition* (4.5.1) *and assuming in addition that* (4.35) *holds, then the characteristic function of the random vector $\underline{\boldsymbol{X}}_t$ can be written as*

$$\mathbb{E}[e^{i\langle \underline{\boldsymbol{u}}, \underline{\boldsymbol{X}}_t\rangle}] = \exp\left\{-\int_{C_4^{\|\cdot\|}} |\langle \underline{\boldsymbol{u}}, \underline{\boldsymbol{s}}\rangle|^\alpha \Gamma_4^{\|\cdot\|}(d\underline{\boldsymbol{s}})\right\}, \qquad \textit{for all} \quad \underline{\boldsymbol{u}} \in \mathbb{R}^4,$$

where

$$\Gamma_4^{\|\cdot\|} = \Delta^{\|\cdot\|} + \Gamma_{4,1}^{\|\cdot\|} + \Gamma_{4,2}^{\|\cdot\|}. \tag{4.57}$$

Here,

$$\Delta^{\|\cdot\|} = \frac{\sigma_1^\alpha}{2}\frac{|\rho_1|^{2\alpha}}{1-|\rho_1|^\alpha}(\delta_{\{\boldsymbol{x}_1\}} + \delta_{\{-\boldsymbol{x}_1\}}) + \frac{\sigma_2^\alpha}{2}\frac{|\rho_2|^\alpha}{1-|\rho_2|^\alpha}(\delta_{\{\boldsymbol{x}_2\}} + \delta_{\{-\boldsymbol{x}_2\}}), \tag{4.58}$$

with points $\boldsymbol{x}_1, \boldsymbol{x}_2$ as in Lemma 4.7.4, and

$$\Gamma_{4,1}^{\|\cdot\|}(d\boldsymbol{s}) = \|B\boldsymbol{s}\|_e^{-\alpha}\tilde{\Gamma}_{4,1} \circ T_B \circ T_{\|\cdot\|}^{-1}(d\boldsymbol{s}), \tag{4.59}$$

$$\Gamma_{4,2}^{\|\cdot\|}(d\boldsymbol{s}) = \|C\boldsymbol{s}\|_e^{-\alpha}\tilde{\Gamma}_{4,2} \circ T_C \circ T_{\|\cdot\|}^{-1}(d\boldsymbol{s}). \tag{4.60}$$

Moreover, for any Borel set $A \subset C_4^{\|\cdot\|}$,

$$\Gamma_{4,1}^{\|\cdot\|}(A) = \Gamma_2 \circ h_1\Big(T_B \circ T_{\|\cdot\|}^{-1}(A \cap C_{4,1}^{\|\cdot\|})\Big), \tag{4.61}$$

$$\Gamma_{4,2}^{\|\cdot\|}(A) = |\rho_1|^\alpha \int_{T_C \circ T_{\|\cdot\|}^{-1}(A \cap C_{4,2}^{\|\cdot\|})} |s_3|^\alpha \Gamma_2 \circ h_2(d\boldsymbol{s}), \tag{4.62}$$

where

$$C_{4,1}^{\|\cdot\|} = \{(s_1, s_2, 0, \rho_2 s_2) \in C_4^{\|\cdot\|} : \quad (s_1, s_2) \in S_2\}, \tag{4.63}$$

$$C_{4,2}^{\|\cdot\|} = \{\vartheta'(1, 0, \rho_1^{-1}, s_4) \in C_4^{\|\cdot\|} : \quad s_4 \in \mathbb{R},\ \vartheta' \in \{-1,+1\}\} \tag{4.64}$$

Proof.

Starting from Γ_4 as given in Lemma 4.7.4 and applying a change of variable using $T_{\|\cdot\|}$ yields (4.57)-(4.60). To show (4.61), consider

$$\Gamma_{4,1}^{\|\cdot\|}(A) = \int_A \|B\boldsymbol{s}\|_e^{-\alpha}\tilde{\Gamma}_{4,1} \circ T_B \circ T_{\|\cdot\|}^{-1}(d\boldsymbol{s}),$$

and perform the first change of variable $\boldsymbol{s}' = T_{\|\cdot\|}^{-1}(\boldsymbol{s}) = \boldsymbol{s}/\|\boldsymbol{s}\|_e$ and get

$$\Gamma_{4,1}^{\|\cdot\|}(A) = \int_{T_{\|\cdot\|}^{-1}(A)} \|B\boldsymbol{s}'\|_e^{-\alpha}\|\boldsymbol{s}'\|^{\alpha}\tilde{\Gamma}_{4,1}\circ T_B(d\boldsymbol{s}').$$

With the second change of variable $\boldsymbol{s} = T_B(\boldsymbol{s}') = B\boldsymbol{s}'/\|B\boldsymbol{s}'\|_e$, we obtain

$$\Gamma_{4,1}^{\|\cdot\|}(A) = \int_{T_B\circ T_{\|\cdot\|}^{-1}(A)} \|B^{-1}\boldsymbol{s}\|^{-\alpha}\tilde{\Gamma}_{4,1}(d\boldsymbol{s}).$$

Given that $\tilde{\Gamma}_{4,1}(\,\cdot\,) = \Gamma_2 \circ h_1(\,\cdot\,\cap S_{4,1})$,

$$\Gamma_{4,1}^{\|\cdot\|}(A) = \int_{T_B\circ T_{\|\cdot\|}^{-1}(A)\cap S_{4,1}} \|B^{-1}\boldsymbol{s}\|^{-\alpha}\tilde{\Gamma}_{4,1}(d\boldsymbol{s}),$$

and noticing that for any $\boldsymbol{s}\in S_{4,1}$, $B^{-1}\boldsymbol{s} = \boldsymbol{s}$ and $\|\boldsymbol{s}\| = 1$, we get

$$\Gamma_{4,1}^{\|\cdot\|}(A) = \Gamma_2\circ h_1\Big(T_B\circ T_{\|\cdot\|}^{-1}(A)\cap S_{4,1}\Big),$$

and using the fact that $T_B\circ T_{\|\cdot\|}^{-1}$ is bijective, we have

$$T_B\circ T_{\|\cdot\|}^{-1}(A)\cap S_{4,1} = T_B\circ T_{\|\cdot\|}^{-1}\Big(A\cap \;\; T_{\|\cdot\|}\circ T_B^{-1}(S_{4,1})\Big) = T_B\circ T_{\|\cdot\|}^{-1}(A\cap C_{4,1}^{\|\cdot\|}),$$

and (4.61) follows. We proceed similarly for (4.62) using in addition the fact that $\Gamma_2\Big(\{(0,-1),(0,+1)\}\Big) = 0$.

□

Return to the proof of Proposition 4.5.2

(ι) As (4.35) holds, we know by Proposition 4.5.1 that $\underline{\boldsymbol{X}}_t$ is representable on $C_4^{\|\cdot\|}$. By Lemma 4.7.5, we further know that its spectral measure $\Gamma_4^{\|\cdot\|}$ satisfies (4.57)-(4.58) and (4.61)-(4.62). Thus, by Proposition 4.2.2,

$$\mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t}, A_{\theta,\eta,P}\Big|B(V_0)\Big) \underset{x\to+\infty}{\longrightarrow} \frac{\Gamma_4^{\|\cdot\|}\Big(A_{\theta,\eta,P}\cap B(V_0)\Big)}{\Gamma_4^{\|\cdot\|}\Big(B(V_0)\Big)},$$

and

$$\frac{\Gamma_4^{\|\cdot\|}\Big(A_{\theta,\eta,P}\cap B(V_0)\Big)}{\Gamma_4^{\|\cdot\|}\Big(B(V_0)\Big)} = \frac{\Big[\Delta^{\|\cdot\|}+\Gamma_{4,1}^{\|\cdot\|}+\Gamma_{4,2}^{\|\cdot\|}\Big]\Big(A_{\theta,\eta,P}\cap B(V_0)\Big)}{\Big[\Delta^{\|\cdot\|}+\Gamma_{4,1}^{\|\cdot\|}+\Gamma_{4,2}^{\|\cdot\|}\Big]\Big(B(V_0)\Big)}. \tag{4.65}$$

From (4.58), (4.61)-(4.62), we can see that for any Borel set $A\subset C_4^{\|\cdot\|}$,

$$\begin{aligned}\Delta^{\|\cdot\|}(A) &= \Delta^{\|\cdot\|}(A\cap\{\pm\boldsymbol{x}_1,\pm\boldsymbol{x}_2\}),\\ \Gamma_{4,1}^{\|\cdot\|}(A) &= \Gamma_{4,1}^{\|\cdot\|}(A\cap C_{4,1}^{\|\cdot\|})\\ \Gamma_{4,2}^{\|\cdot\|}(A) &= \Gamma_{4,2}^{\|\cdot\|}(A\cap C_{4,2}^{\|\cdot\|}),\end{aligned}$$

where $C_{4,1}^{\|\cdot\|}$, $C_{4,2}^{\|\cdot\|}$ are given in (4.63) and (4.64). We thus proceed in three steps: (1) we derive the form of sets in the right-hand side of the above equations in the case $A = B(V_0)$, (2) we then consider $A = A_{\theta,\eta,P}\cap B(V_0)$,

(3) we finally evaluate the mass over the obtained sets to derive the numerator and denominator in (4.65). Let us consider the denominator. Because we assume $V_0 \cap \{(\pm 1, 0), (0, \pm 1)\} = \emptyset$, it is easy to see that

$$
\begin{aligned}
B(V_0) \cap \{\pm \boldsymbol{x}_1, \pm \boldsymbol{x}_2\} &= \emptyset, \\
B(V_0) \cap C_{4,1}^{\|\cdot\|} &= \{(\cos u, \sin u, 0, \rho_2 \sin u) : \quad u \in [\theta_0 - \eta_0, \theta_0 + \eta_0]\}, \\
B(V_0) \cap C_{4,2}^{\|\cdot\|} &= \emptyset.
\end{aligned}
$$

Thus, by (4.61),

$$
\begin{aligned}
\Gamma_4^{\|\cdot\|}\Big(B(V_0)\Big) &= \Gamma_{4,1}^{\|\cdot\|}\Big(B(V_0) \cap C_{4,1}^{\|\cdot\|}\Big) \\
&= \Gamma_2 \circ h_1\Big(T_B \circ T_{\|\cdot\|}^{-1}((B(V_0) \cap C_{4,1}^{\|\cdot\|})\Big) \\
&= \Gamma_2 \circ h_1\Big(\{(\cos u, \sin u, 0, 0) : \quad u \in [\theta_0 - \eta_0, \theta_0 + \eta_0]\}\Big) \\
&= \Gamma_2(V_0).
\end{aligned}
$$

For the numerator, we have

$$
A_{\theta,\eta,P} \cap B(V_0) \cap C_{4,1}^{\|\cdot\|} = \begin{cases} V_{\theta,\eta} \cap V_0, & \text{if} \quad (0,0) \in P, \\ \emptyset, & \text{if} \quad (0,0) \notin P. \end{cases}
$$

The conclusion follows.

$(\iota\iota)$ We proceed as in point (ι). Given the assumptions on V_0, we have

$$
\begin{aligned}
B(V_0) \cap \{\pm \boldsymbol{x}_1, \pm \boldsymbol{x}_2\} &= \{\vartheta \boldsymbol{x}_2\}, \\
B(V_0) \cap C_{4,1}^{\|\cdot\|} &= \{(\cos u, \sin u, 0, \rho_2 \sin u) : \quad u \in [\theta_0 - \eta_0, \theta_0 + \eta_0]\}, \\
B(V_0) \cap C_{4,2}^{\|\cdot\|} &= \emptyset.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\Gamma_4^{\|\cdot\|}\Big(B(V_0)\Big) &= \Delta^{\|\cdot\|}(\{\vartheta \boldsymbol{x}_2\}) + \Gamma_{4,1}^{\|\cdot\|}\Big(B(V_0) \cap C_{4,1}^{\|\cdot\|}\Big) \\
&= \frac{\sigma_2^\alpha}{2} \frac{|\rho_2|^\alpha}{1 - |\rho_2|^\alpha} + \Gamma_2(V_0).
\end{aligned}
$$

Turning to $\Gamma_4^{\|\cdot\|}\Big(A_{\theta,\eta,P} \cap B(V_0)\Big)$, consider

$$
\begin{aligned}
A_{\theta,\eta,P} &\cap B(V_0) \cap \{\pm x_1, \pm x_2\} = A_{\theta,\eta,P} \cap \{\vartheta \boldsymbol{x}_2\} \\
&= \{\vartheta(0,1,0,\rho_2)\} \cap \{(\cos u, \sin u, 0, \rho_2 \sin u) + (0,0,x,y) : \quad u \in [\theta - \eta, \theta + \eta] \text{ and } (x,y) \in P\}.
\end{aligned}
$$

Noticing that for any u such that $(\cos u, \sin u) \neq (0, \vartheta)$, necessarily

$$
\vartheta(0,1,0,\rho_2) \neq (\cos u, \sin u, 0, \rho_2 \sin u) + (0,0,x,y), \quad \text{for all} \quad (x,y) \in P,
$$

we have

$$\{\vartheta \boldsymbol{x}_2\} \cap A_{\theta,\eta,P} = \begin{cases} \{\vartheta \boldsymbol{x}_2\} \cap \{\vartheta \boldsymbol{x}_2 + (0,0,x,y): \ (x,y) \in P\}, & \begin{array}{c}\text{if} \quad (\cos u, \sin u) = (0,\vartheta), \\ \text{for some } u \in [\theta-\eta, \theta+\eta],\end{array} \\ \\ \emptyset, & \text{otherwise.} \end{cases}$$

Hence

$$A_{\theta,\eta,P} \cap B(V_0) \cap \{\pm x_1, \pm x_2\} = \begin{cases} \{\vartheta x_2\}, & \text{if} \quad (0,\vartheta) \in V_{\theta,\eta} \text{ and } (0,0) \in P \\ \emptyset, & \text{otherwise.} \end{cases}$$

Similarly, we have

$$\begin{aligned} A_{\theta,\eta,P} \cap B(V_0) \cap C_{4,1}^{\|\cdot\|} &= \begin{cases} V_{\theta,\eta} \cap V_0, & \text{if} \quad (0,0) \in P, \\ \emptyset, & \text{if} \quad (0,0) \notin P, \end{cases} \\ A_{\theta,\eta} \cap B(V_0) \cap C_{4,2}^{\|\cdot\|} &= \emptyset. \end{aligned}$$

The result follows by evaluating $\Gamma_4^{\|\cdot\|}$ on the above sets.

($\iota\iota\iota$) Proceeding as above, we first have that

$$\begin{aligned} B(V_0) \cap \{\pm x_1, \pm x_2\} &= \{\vartheta \boldsymbol{x}_1\}, \\ B(V_0) \cap C_{4,1}^{\|\cdot\|} &= \{(\cos u, \sin u, 0, \rho_2 \sin u): \quad u \in [\theta_0 - \eta_0, \theta_0 + \eta_0]\}, \\ B(V_0) \cap C_{4,2}^{\|\cdot\|} &= \{\vartheta(1,0,\rho_1^{-1}, s_4): \quad s_4 \in \mathbb{R}\}. \end{aligned}$$

Hence,

$$\Gamma_4^{\|\cdot\|}(B(V_0)) = \frac{\sigma_1^\alpha}{2} \frac{|\rho_1|^{2\alpha}}{1-|\rho_1|^\alpha} + \Gamma_2(V_0) + \Gamma_{4,2}^{\|\cdot\|}(\{\vartheta(1,0,\rho_1^{-1},s_4): \quad s_4 \in \mathbb{R}\}).$$

Given that

$$T_C \circ T_{\|\cdot\|}^{-1}\left(B(V_0) \cap C_{4,2}^{\|\cdot\|}\right) = \left\{\vartheta \frac{(0,0,\rho_1^{-1},s_4)}{\sqrt{\rho_1^{-2}+s_4^2}}: \quad s_4 \in \mathbb{R}\right\} = \{(0,0,s_1,s_2): \quad (s_1,s_2) \in S_2, \quad s_1 \vartheta \rho_1 > 0\},$$

the third term in $\Gamma_4^{\|\cdot\|}$ can be rewritten using (4.62) as

$$\Gamma_{4,2}^{\|\cdot\|}(\{\vartheta(1,0,\rho_1^{-1},s_4): \quad s_4 \in \mathbb{R}\}) = |\rho_1|^\alpha \int_{\{(s_1,s_2)\in S_2: \quad s_1\vartheta\rho_1>0\}} |s_1|^\alpha \Gamma_2(d\boldsymbol{s}).$$

Since, by assumption, Γ_2 is symmetric and does not charge masses at $(0,\pm 1)$,

$$\int_{\{(s_1,s_2)\in S_2: \quad s_1\vartheta\rho_1>0\}} |s_1|^\alpha \Gamma_2(d\boldsymbol{s}) = \frac{1}{2}\int_{S_2} |s_1|^\alpha \Gamma_2(d\boldsymbol{s}) = \sigma_1^\alpha/2,$$

and thus

$$\begin{aligned} \Gamma_4^{\|\cdot\|}(B(V_0)) &= \frac{\sigma_1^\alpha}{2} \frac{|\rho_1|^{2\alpha}}{1-|\rho_1|^\alpha} + \Gamma_2(V_0) + |\rho_1|^\alpha \frac{\sigma_1^\alpha}{2} \\ &= \Gamma_2(V_0) + \frac{\sigma_1^\alpha}{2} \frac{|\rho_1|^\alpha}{1-|\rho_1|^\alpha}. \end{aligned}$$

We last turn to $\Gamma_4\Big(A_{\theta,\eta,P}\cap B(V_0)\Big)$. We have

$$
\begin{aligned}
&A_{\theta,\eta,P}\cap B(V_0)\cap\{\pm x_1,\pm x_2\} = A_{\theta,\eta,P}\cap\{\vartheta\boldsymbol{x}_1\}\\
&\quad = \{\vartheta(1,0,\rho_1^{-1},0)\}\cap\{(\cos u,\sin u,0,\rho_2\sin u)+(0,0,x,y):\ \ u\in[\theta-\eta,\theta+\eta]\text{ and }(x,y)\in P\}\\
&\quad = \begin{cases} \{\vartheta\boldsymbol{x}_1\}\cap\{\vartheta(1,0,0,0)+(0,0,x,y):\ (x,y)\in P\}\,, & \text{if}\quad (\vartheta,0)\in V_{\theta,\eta}\\ \emptyset, & \text{otherwise.}\end{cases}\\
&\quad = \begin{cases} \{\vartheta x_1\}, & \text{if}\quad (\vartheta,0)\in V_{\theta,\eta}\ \text{ and }\ (\vartheta\rho_1^{-1},0)\in P,\\ \emptyset, & \text{otherwise.}\end{cases}
\end{aligned}
$$

As previously, we also have that

$$
A_{\theta,\eta,P}\cap B(V_0)\cap C_{4,1}^{\|\cdot\|} = \begin{cases} V_{\theta,\eta}\cap V_0, & \text{if}\quad (0,0)\in P,\\ \emptyset, & \text{if}\quad (0,0)\notin P,\end{cases}
$$

Finally,

$$
\begin{aligned}
&A_{\theta,\eta,P}\cap B(V_0)\cap C_{4,2}^{\|\cdot\|} = A_{\theta,\eta,P}\cap\{\vartheta(1,0,\rho_1^{-1},s_4):\ \ s_4\in\mathbb{R}\}\\
&\quad = \begin{cases} \{\vartheta(1,0,\rho_1^{-1},s_4):\ \ s_4\in\mathbb{R}\}\cap\{\vartheta(1,0,0,0)+(0,0,x,y):\ (x,y)\in P\}\,, & \text{if}\quad (\vartheta,0)\in V_{\theta,\eta}\\ \emptyset, & \text{otherwise.}\end{cases}\\
&\quad = \begin{cases} \{\vartheta(1,0,\rho_1^{-1},y):\ \ y\in P_2\}, & \text{if}\quad (\vartheta,0)\in V_{\theta,\eta}\ \text{ and }\ \vartheta\rho_1^{-1}\in P_1,\\ \emptyset, & \text{otherwise,}\end{cases}
\end{aligned}
$$

and

$$
T_C\circ T_{\|\cdot\|}^{-1}\Big(\{\vartheta(1,0,\rho_1^{-1},y):\ \ y\in P_2\}\Big) = \left\{\vartheta\frac{(0,0,\rho_1^{-1},y)}{\sqrt{\rho_1^{-2}+y^2}}:\ \ y\in P_2\right\}.
$$

$\square$

Chapter 5

Conclusion and Perspectives

5.1 Conclusion

This thesis focuses on estimation and conditional dynamics aspects of a class of anticipative α-stable time series. In the framework of AR processes, it is shown that conditional moments of higher order than marginal ones exist provided the autoregressive polynomial admits at least one root inside the unit circle. The forms of the first and second order conditional moments and causal representations are exhibited in special cases for AR processes with roots both inside and outside the unit circle. The case of the stable anticipative AR(1) introduced by Gouriéroux and Zakoïan (2017) [63], and for which only partial results were available, is revisited. It is shown that any two realisations at different time points on a trajectory form a bivariate stable random vector with a specific spectral measure charging either two or four atoms on the unit circle. The theory of bivariate stable random vectors applies: a different proof for the existence of higher-order conditional moments is provided and the form of the conditional expectation can be derived for any parameterisation of the process. We contribute to the literature on general bivariate stable vectors by providing the functional forms of the conditional variance, skewness and kurtosis under any admissible parameterisation. These results are applied to pairs of realisations on trajectories of the anticipative AR(1) and yield the functional forms of the first four conditional moments of the process. During extreme events, these moments are shown to be asymptotically equivalent to those of a weighted Bernoulli distribution charging two polarly-opposite future paths: exponential growth or collapse. A similar behaviour is shown to hold for the continuous time counterpart of the AR(1), the anticipative α-stable Ornstein-Uhlenbeck process. From the moments perspective, the conditional dynamics of anticipative stable AR processes differs sharply from that of their non-anticipative counterparts. Yet, on sample trajectories, the classical Least Squares estimation method is unable to identify whether the autoregressive polynomial admits roots inside the unit circle. We establish that the LS procedure provides a consistent estimator of an all-pass causal representation of the process, the validity of which can be tested by a Monte Carlo portmanteau-type test. All the roots of this all-pass representation lie outside the unit circle and each corresponds either to a root of the original representation of the AR process, or to its reciprocal. We justify, based on point-process arguments, that the two types of roots can be discriminated by searching for evidence of extreme clustering in the residuals of other all-pass representations.

For α-stable infinite moving averages, the conditional distribution of future paths given the observed past trajectory during extreme events is obtained on the basis of a new spectral representation of stable random vectors on unit cylinders relative to semi-norms. Contrary to the case of norms, such representations yield a multivariate regularly varying tails property that is appropriate for prediction purposes, however not all stable random vectors can be represented on semi-norm unit cylinders. A characterisation is provided and finite length paths of α-stable moving averages, which are themselves multivariate α-stable, are embedded into this framework. We show that such paths admit semi-norm representations that are appropriate for prediction purposes if and only if the moving average process is «anticipative enough». The gap between anticipative and non-anticipative processes stems from the fact that for the latter, future extreme events are independent of the past trajectory, whereas for the former, future extreme events are not independent of the observed trajectory, their incoming occurrences being signaled by early trends and patterns. The approach extends to processes resulting from the linear combination of α-stable moving averages, coined *stable aggregates*, which can accommodate very rich dynamics. For eligible processes, that is, anticipative ones, the conditional dynamics during extreme events is derived and their trajectories are shown to follow precise patterns whose forms are determined by the moving averages coefficients sequence. In particular, it is noticed that non-aggregated moving averages generate trajectories which can only feature a single pattern, recurrently from one extreme event to another, whereas trajectories of aggregated moving averages can display any number of different patterns. The path prediction problem then translates into a pattern identification task: given a piece of observed trajectory, what piece of pattern does it correspond to and what are the potential future paths compatible with this pattern ? Provided the spectral representation on the corresponding unit cylinder is known, our results can be used to evaluate the respective odds of the potential future paths conditionally on the identified pattern.

5.2 Perspectives

Future lines of research could focus on several opened questions, both on probabilistic and statistical aspects.

1. The conditional distribution of aggregated moving averages holds asymptotically in the (semi-)norm of the observed trajectory being large. Could we evaluate the approximation error made when using the asymptotic distribution in lieu of the finite distance one for prediction ? What could be said about the conditional distribution when the process is close to its central values ? A perhaps dual question would be: could we evaluate how variable the future trajectory may be around the predicted deterministic paths ? One could expect in particular that this variability may increase with the prediction horizon.

2. For practical use in applications, estimation/learning methods need to be developed to infer the structure of «a best approximating» aggregated moving average to some time series data. This requires identifying the coefficients sequences of the moving averages involved in the aggregation (which characterise the shape of patterns appearing during extreme events), the number of moving averages involved (the number of different patterns) and the distribution parameters of the i.i.d. errors driving the process. Recovering the patterns amounts in the general case to estimating an infinite number of parameters and one is likely to seek instead a parsimonious, low-dimensional structure in the coefficients sequence. This structure need not assume that the coefficients of the moving averages satisfy some linear recursive relation, as in the case of ARMA processes. Complex patterns can be achieved even with few parameters, as illustrated by the strophoid-generating process defined at the end of Chapter 4, which coefficients sequence is characterised by only two parameters.

3. The conditional distribution derived in Chapter 4 obviously requires to provide a conditioning Borel set, which represents the information about the shape of the observed trajectory. This piece of observed trajectory can be viewed as a single noisy realisation of a piece of pattern generated by the process, which leaves room for uncertainty in the identification of that pattern. The choice of an adequate conditioning Borel should thus rely on a statistical method. One could envision for instance tests of hypotheses to determine whether the observed trajectory is «more akin» a certain pattern 1 or another pattern 2, or to any other pattern of a certain collection. Moreover, the length of the observed trajectory does not have to remain fixed: shorter observation windows closer to the present date may contain more «up-to-date» information, less influenced by now vanished past extreme events on the one hand, but on the other hand could also be more subject to noise and make the pattern identification more difficult. Conversely, wider observation windows may provide more robust pattern identification but may also incorporate biased information, being influenced by now irrelevant past events. One could envision looking for event-driven optimal window length based on a bias-variance trade-off.

4. A multivariate extension is illustrated on a simple bivariate process with one purely anticipative and one purely non-anticipative component. New properties already emerge, such as the fact that while univariate non-anticipative processes never induce paths representable on unit cylinders, their paths may nonetheless be representable in higher dimensions when considered alongside an «informative» anticipative process. Both components are more predictable when considered jointly rather than univariately. The general multivariate framework can be readily embedded in the approach of Chapter 4 but numerous potential interactions between univariate components render the task challenging.

5. It is known for infinite variance moving averages that empirical autocorrelations tend towards the theoretical autocorrelations that would prevail in the finite variance case as the sample size grows to infinity [40, 41]. In the case of stable aggregated moving averages, to the best of our knowledge, it is both unknown what the «theoretical» autocorrelations would be, as well as what asymptotic behaviour

would hold for the empirical ones. One may moreover aim at inferring the autocorrelation structures of the latent moving averages involved in the aggregation, rather than that of the observed process itself. If the latent moving averages are by definition unobserved, they recurrently reveal themselves during extreme events. A localised autocorrelations estimator may enable to recover the information about the latent components.

6. The conditional dynamics of the continuous time stable anticipative Ornstein-Uhlenbeck was shown to be similar to that of its discrete time counterpart, at least from the moments perspective. The approach followed for discrete time stable aggregates in Chapter 4 could be applied to paths of continuous time processes, viewed as vectors of arbitrarily spaced points on a trajectory.

7. As long as a single tail index α prevails for all moving averages involved in the definition of an aggregated process, or across all the components of a multivariate stable process, one can undoubtedly expect to work with some multivariate α-stable distribution downstream. Allowing for different tail indexes would be appealing in order to alleviate a certainly undue limitation. This may prove however difficult to work with, as the sum-stability property which is very convenient when working with linear processes would be lost.

Bibliography

[1] An, H.Z., and Z.G., Chen. 1982. On convergence of LAD estimates in autoregression with infinite variance. *Journal of Multivariate Analysis* 12, 335-345.

[2] Anderson, T. W. 1959. On asymptotic distributions of estimates of parameters of stochastic difference equations. *The Annals of Mathematical Statistics*, 676-687.

[3] Andrews, B., Calder, M., and R., Davis. 2009. Maximum likelihood estimation for α-stable autoregressive process. *Annals of Statistics*, 37, 1946-1982.

[4] Andrews, B. and R. A., Davis. 2013. Model identification for infinite variance autoregressive processes. *Journal of Econometrics*, 172(2), 222-234.

[5] Andrews, B., Davis, R. A., and F. J., Breidt. 2007. Rank-based estimation for all-pass time series models. *The Annals of Statistics*, 35, 844-869.

[6] Bartlett, M. S. 1946. On the theoretical specification and sampling properties of autocorrelated time-series. *Supplement to the Journal of the Royal Statistical Society*, 8, 27-41.

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

[8] Basrak, B., and J., Segers. 2009. Regularly varying multivariate time series. *Stochastic processes and their applications*, 119, 1055-1080.

[9] Behme, A. D. 2011. Distributional properties of solutions of $dV_t = V_t - dU_t + dL_t$ with Lévy noise. *Advances in Applied Probability*, 43(3), 688-711.

[10] Behme, A., Lindner, A., and R., Maller. 2011. Stationary solutions of the stochastic differential equation with Lévy noise. *Stochastic Processes and their Applications*, 121, 91-108.

[11] Benveniste, A., Goursat, M., and G., Ruget. 1980. Robust identification of a nonminimum phase system: Blind adjustment of a linear equalizer in data communications. *IEEE Transactions on Automatic Control*, 25, 385-399.

[12] Bierens, H. 1982. Consistent model specification tests. *Journal of Econometrics*, 20, 105-134.

[13] Blanchard, O., and M., Watson. 1982. Bubbles, rational expectations and financial markets (No. 0945). National Bureau of Economic Research, Inc.

[14] Box, G. and G., Jenkins. 1970. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.

[15] Breidt, F.J., Davis, R.A. and A.A. Trindade. 2001. Least absolute deviation estimation for all-Pass time series models. *Annals of Statistics*, 29, 919-946.

[16] Breidt, F. J., Davis, R., Lii, K. S., and M., Rosenblatt. 1991. Maximum likelihood estimation for noncausal autoregressive processes. *Journal of Multivariate Analysis*, 36, 175-198.

[17] Breton, A., and D. T., Pham. 1989. On the bias of the least squares estimator for the first order autoregressive process. *Annals of the institute of Statistical Mathematics*, 41, 555-563.

[18] Brock, W. A., Dechert, W. D., LeBaron, B., and J. A. Scheinkman. 1996. A test for independence based on the correlation dimension. *Econometric Reviews*, 15(3), 197-235.

[19] Brockwell, P.J. and R.A., Davis. 1991. *Time series: theory and methods*, Springer-Verlag, New York, 2nd edition.

[20] Cambanis, S., and I, Fakhre-Zakeri. 1995. On prediction of heavy-tailed autoregressive sequences: forward versus reversed time. *Theory of Probability & Its Applications*, 39, 217-233.

[21] Cambanis, S., and S., Fotopoulos. 1995. Conditional variance for stable random vectors. *Probability and Mathematical Statistics*, Wrocław University, 15, 195-214.

[22] Cambanis, S., and G., Miller. 1981. Linear problems in pth order and stable processes. SIAM *Journal on Applied Mathematics*, 41, 43–69.

[23] Cambanis, S., and W., Wu. 1992. Multiple regression on stable vectors. *Journal of multivariate analysis*, 41, 243-272.

[24] Cavaliere, G., Nielsen, H.B, and A. Rahbek. 2018. Bootstrapping non-causal autoregressions: with applications to explosive bubble modelling. Forthcoming in *Journal of Business and Economic Statistics*.

[25] Chavez-Demoulin, V., and A. C., Davison. 2012. Modelling time series extremes. *REVSTAT-Statistical Journal*, 10, 109-133.

[26] Chen, B., Choi, J., and J. C., Escanciano. 2017. Testing for fundamental vector moving average representations. *Quantitative Economics*, 8, 149-180.

[27] Chen, Y., Phillips, P.C.B., and J. Yu. 2017. Inference in continuous systems with mildly explosive regressors. *Journal of Econometrics*, 201, 400-416.

[28] Chien, H. M., Yang, H. L., and C. Y., Chi 1997. Parametric cumulant based phase estimation of 1-D and 2-D nonminimum phase systems by allpass filtering. *IEEE Transactions on Signal Processing*, 45, 1742-1762.

[29] Chi, C. Y., and J. Y., Kung. 1995. A new identification algorithm for allpass systems by higher-order statistics. *Signal Processing*, 41, 239-256.

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

[31] Cioczek-Georges, R., and M. S., Taqqu. 1995. Form of the conditional variance for stable random variables. *Statistica Sinica*, 351-361.

[32] Cioczek-Georges, R., and M. S., Taqqu. 1995. Necessary conditions for the existence of conditional moments of stable random variables. *Stochastic processes and their applications*, 56, 233-246.

[33] Cioczek-Georges, R., and M. S., Taqqu. 1998. Sufficient conditions for the existence of conditional moments of stable random variables. *Stochastic Processes and Related Topics*, Birkhäuser Boston, 35-67.

[34] Cline, D. 1983. Estimation and linear prediction for regression, autoregression and ARMA with infinite variance data. Ph.D. Thesis, Department of Statistics, Colorado State University.

[35] Cubadda, G., Hecq, A., and S., Telg. 2017. Detecting Co-Movements in Noncausal Time Series. Maastricht University, https://mpra.ub.uni-muenchen.de/77254/

[36] Erdős, P. and A. H., Stone. 1970. On the sum of two Borel sets. *Proceedings of the American Mathematical Society*, 25, 304-306.

[37] Daley, D. J. and D., Vere-Jones. 2007. *An Introduction to the theory of point processes*, Springer, New York, vol. 1.

[38] Davis, R., Knight, K. and J., Liu. 1992. M-estimation for autoregressions with infinite variance. *Stochastic Processes and their Applications*, 40, 145-180.

[39] Davis, R., and H, Liu. 2012. Theory and inference for a class of observation driven models with application to time series of counts. ArXiv preprint. arXiv:1204.3915.

[40] Davis, R. and S., Resnick. 1985. Limit theory for moving averages of random variables with regularly varying tail probabilities. *Annals of Probability*, 13, 179-195.

[41] Davis, R. and S., Resnick. 1986. Limit theory for the sample covariance and correlation functions of moving averages. *Annals of Statistics*, 14, 533-558.

[42] Davis, R. and L., Song. 2012. Noncausal vector AR process with application to economic time series. Discussion Paper, Columbia University.

[43] De Haan, L., Mercadier, C., and C. Zhou. 2016. Adapting extreme value statistics to financial time series: dealing with bias and serial dependence. *Finance and Stochastics*, 20, 321-354.

[44] Dombry, C., Hashorva, E., and P., Soulier. 2017. Tail measure and tail spectral process of regularly varying time series. arXiv preprint arXiv:1710.08358.

[45] Donoho, D. 1981. On minimum entropy deconvolution. In *Applied Time Series Analysis II* (D. F. Findley, ed.) 565-608. Academic Press, New York.

[46] Duchesne, P., Ghoudi, K. and B. Rémillard. 2012. On testing for independence between the innovations of several time series. *Canadian Journal of Statistics*, 40, 447-479.

[47] Embrechts, P., Klüppelberg, C. and T. Mikosch. 1997. *Modelling Extremal Events. Applications of Mathematics*, Springer-Verlag, Berlin.

[48] Ferro, C. A., and J., Segers. 2003. Inference for clusters of extreme values. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65, 545-556.

[49] Fotopoulos, S. B. 1998. On the conditional variance-covariance of stable random vectors, II. *Stochastic Processes and Related Topics*, Birkhäuser Boston, 123-140.

[50] Francq, C. and J-M., Zakoian. 2010. *GARCH models: structure, statistical inference and financial applications.* John Wiley.

[51] Fries, S. and J.-M., Zakoian. 2017. Mixed causal-noncausal AR processes and the modelling of explosive bubbles. University Paris-Saclay, https://mpra.ub.uni-muenchen.de/81345/

[52] Fries, S. 2018. Conditional moments of anticipative α-stable Markov processes. ENSAE-ParisTech, arXiv:1805.05397.

[53] Gassiat, E. 1990. Estimation semi-paramétrique d'un modèle autorégressif stationnaire multi-indice non nécessairement causal. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 26, 181-205.

[54] Gassiat, E. 1990. Semi-parametric estimation of a stationary, non-necessary causal AR(p) process with infinite variance. *Journal of multivariate analysis*, 32, 161-170.

[55] Gassiat, E. 1993. Adaptive estimation in noncausal stationary AR processes. *The Annals of Statistics*, 21, 2022-2042.

[56] Giannakis, G. B., and J. M., Mendel. 1989. Identification of nonminimum phase systems using higher order statistics. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37, 360-377.

[57] Giannakis, G. B., and A., Swami. 1990. On estimating noncausal nonminimum phase ARMA models of non-Gaussian processes. *IEEE transactions on acoustics, speech, and signal processing*, 38, 478-495.

[58] Godfrey, R., and F., Rocca. 1981. Zero memory non-linear deconvolution. *Geophysical Prospecting*, 29, 189-228.

[59] Gouriéroux, C., and J., Jasiak. 2016. Filtering, prediction and simulation methods for noncausal processes. *Journal of Time Series Analysis*, 37, 405-430.

[60] Gouriéroux, C., and J., Jasiak. 2017. Noncausal vector autoregressive process: Representation, identification and semi-parametric estimation. *Journal of Econometrics*, 200, 118-134.

[61] Gouriéroux, C., Jasiak, J. and A., Monfort. 2016. Stationary bubble equilibria in rational expectation models. Working Papers from Centre de Recherche en Economie et Statistique (CREST), No 2016-31.

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

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

[64] Hardin Jr, C. D., Samorodnitsky, G., and M. S., Taqqu. 1991. Nonlinear regression of stable random variables. *The Annals of Applied Probability*, 582-612.

[65] Harvey, C. R., Liechty, J. C., Liechty, M. W., and P., Müller. 2010. Portfolio selection with higher moments. *Quantitative Finance*, 10, 469-485.

[66] Hecq, A., Issler, J. V., and S. M., Telg. 2017. Mixed causal-noncausal autoregressions with strictly exogenous regressors. MPRA Paper 80767, Germany.

[67] Hecq, A., Lieb, L., and S. M., Telg. 2016. Identification of mixed causal-noncausal models in finite samples. *Annals of Economics and Statistics*, 123/124, 307-331.

[68] Hecq, A., Telg, S., and L., Lieb. 2017. Do seasonal adjustments induce noncausal dynamics in inflation rates? *Econometrics*, 5, 48.

[69] Hecq, A., Telg, S., and L, Lieb. 2017. Simulation, estimation and selection of mixed causal-noncausal autoregressive models: The MARX Package. Available at SSRN: https://ssrn.com/abstract=3015797.

[70] Hencic, A., and C., Gouriéroux. 2015. Noncausal autoregressive model in application to bitcoin/usd exchange rates. *Econometrics of Risk*, Springer International Publishing, 17-40.

[71] Hill, B.M. 1975. A simple general approach to inference about the tail of a distribution. *Annals of Mathematical Statistics* 3, 1163-1174.

[72] Hill, J.B. 2010. On Tail Index Estimation For Dependent, Heterogeneous Data. *Econometric Theory*, 26, 1398-1436.

[73] Hsing, T., Hüsler, J., and M. R., Leadbetter. 1988. On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields*, 78, 97-112.

[74] Hsueh, A.C., and J. M. Mendel. 1985. Minimum-variance and maximum-likelihood deconvolution for noncausal channel models. *IEEE Trans. Geoscience and Remote Sensing*, 23, 797-808.

[75] Huang, J., and Y., Pawitan. 2000. Quasi-likelihood estimation of non-invertible moving average processes. *Scandinavian Journal of Statistics*, 27, 689-702.

[76] Izenman, A. J. 1983. JR Wolf and HA Wolfer: An historical note on the Zurich sunspot relative numbers. *Journal of the Royal Statistical Society. Series A (General)*, 311-318.

[77] Johansen, S., and T., Lange. 2011. Some econometric results for the Blanchard-Watson bubble model. No. 11-15, University of Copenhagen. Department of Economics.

[78] Janßen, A. (2017). Spectral tail processes and max-stable approximations of multivariate regularly varying time series. arXiv preprint arXiv:1704.06179.

[79] Janßen, A., and J., Segers. 2014. Markov tail chains. *Journal of Applied Probability*, 51, 1133-1153.

[80] Jondeau, E., and M., Rockinger. 2006. Optimal portfolio allocation under higher moments. *European Financial Management*, 12, 29-55.

[81] Karcher, W., Shmileva, E., and E., Spodarev. 2013. Extrapolation of stable random fields. *Journal of Multivariate Analysis*, 115, 516-536.

[82] Kendall, M. G. 1944. On autoregressive time series. *Biometrika*, 33, 105-122.

[83] Knight, K. 1989. Consistency of Akaike's information criterion for infinite variance autoregressive processes. *The Annals of Statistics*, 17, 824-840.

[84] Lai, T. L., and C. Z., Wei. 1983. Asymptotic properties of general autoregressive models and strong consistency of least-squares estimates of their parameters. *Journal of Multivariate Analysis*, 13, 1-23.

[85] Lanne, M., Luoto, J., and P, Saikkonen. 2012. Optimal forecasting of noncausal autoregressive time series. *International Journal of Forecasting*, 28, 623-631.

[86] Lanne, M., Nyberg, H., and E., Saarinen. 2012. Does noncausality help in forecasting economic time series? *Economics Bulletin*, 32, 2849-2859.

[87] Lanne, M., and P., Saikkonen. 2011. Noncausal autogressions for economic time series. *Journal of Time Series Econometrics*, 3.

[88] Lanne, M., and P., Saikkonen. 2013. Noncausal vector autoregression. *Econometric Theory*, 29, 447-481.

[89] Leadbetter, M. R., and S., Nandagopalan. 1989. On exceedance point processes for stationary sequences under mild oscillation restrictions. *Lecture Notes in Statistics*, 51, 69-80.

[90] Lii, K. S., and M., Rosenblatt. 1982. Deconvolution and estimation of transfer function phase and coefficients for nongaussian linear processes. *The annals of statistics*, 10, 1195-1208.

[91] Lii, K. S., and M., Rosenblatt. 1988. Estimation and deconvolution when the transfer function has zeros. *Journal of Theoretical Probability*, 1, 93-113.

[92] Lii, K. S., and M., Rosenblatt. 1992. An approximate maximum likelihood estimation for non-Gaussian non-minimum phase moving average processes. *Journal of Multivariate Analysis*, 43, 272-299.

[93] Lii, K. S., and M., Rosenblatt. 1996. Maximum likelihood estimation for nonGaussian nonminimum phase ARMA sequences. *Statistica Sinica*, 1-22.

[94] Lin, J. W., and A. I., McLeod. 2008. Portmanteau tests for ARMA models with infinite variance. *Journal of Time Series Analysis*, 29, 600-617.

[95] Lindner, A., and R., Maller. 2005. Lévy integrals and the stationarity of generalised Ornstein–Uhlenbeck processes. *Stochastic processes and their applications*, 115, 1701-1722.

[96] Ling, S. 2007. A double AR(p) model: structure and estimation. *Statistica Sinica*, 17, 161-175.

[97] Maller, R. A., Müller, G., and A., Szimayer. 2009. Ornstein–Uhlenbeck processes and extensions. *Handbook of Financial Time Series*, 421-437.

[98] Mann, H. B., and A, Wald. 1943. On the statistical treatment of linear stochastic difference equations. *Econometrica*, Journal of the Econometric Society, 173-220.

[99] Markovich, N. M. 2014. Modelling clusters of extreme values. *Extremes*, 17, 97-125.

[100] Mason, D. 1982. Laws of large numbers for sums of extreme values. *Annals of Probability* 10, 754-764.

[101] McCulloch, J. H. 1986. Simple consistent estimators of stable distribution parameters. *Communication in Statistics–Computation and Simulation* 15, 1109–36.

[102] Meinguet, T., and J., Segers. 2010. Regularly varying time series in Banach spaces. arXiv preprint arXiv:1001.3262.

[103] Miller, G. 1978. Properties of certain symmetric stable distributions. *Journal of Multivariate Analysis*, 8, 346–360.

[104] Nielsen, H. B., and A., Rahbek. 2014. Unit root vector autoregression with volatility induced stationarity. *Journal of Empirical Finance*, 29, 144-167.

[105] Phillips, P.C.B., Shi, S. and J. Yu. 2015. Testing for multiple bubbles: historical episodes of exuberance and collapse in the S&P500. *International Economic Review*, 56, 1043-1077.

[106] Phillips, P.C.B., Wu, Y. and J. Yu. 2011. Explosive behaviour in the 1990's Nasdaq: when did exuberance escalate asset values? *International Economic Review*, 52, 201-226.

[107] Planinić, H., and P., Soulier. 2017. The tail process revisited. *Extremes*, 1-29.

[108] Rao, M. M. 1961. Consistency and limit distributions of estimators of parameters in explosive stochastic difference equations. *The Annals of Mathematical Statistics*, 195-218.

[109] Robert, C. Y., Segers, J., and C. A., Ferro. 2009. A sliding blocks estimator for the extremal index. *Electronic Journal of Statistics*, 3, 993-1020.

[110] Rootzén, H. 1978. Extremes of moving averages of stable processes. *The Annals of Probability*, 847-869.

[111] Rosenblatt, M. 1995. Prediction and non-Gaussian autoregressive stationary sequences. *The Annals of Applied Probability*, 5, 239-247.

[112] Rosenblatt, M. 2000. *Gaussian and non-Gaussian linear, time series and random fields*, Springer Verlag, New-York.

[113] Royden, H. L., and P. M., Fitzpatrick. 2010. *Real analysis*. Prentice Hall, fourth edition.

[114] Rubin, H. 1950. Consistency of maximum likelihood estimates in the explosive case. *Statistical Inference in Dynamic Economic Models*, 356-364.

[115] Saikkonen, P., and R., Sandberg. 2016. Testing for a unit root in noncausal autoregressive models. *Journal of Time Series Analysis*, 37, 99-125.

[116] Samorodnitsky, G., and M. S., Taqqu. 1991. Conditional moments and linear regression for stable random variables. *Stochastic processes and their applications*, 39, 183-199.

[117] Samorodnitsky, G., and M. S., Taqqu. 1994. *Stable non-Gaussian random processes*, Chapman & Hall, London.

[118] Titchmarsh, E. C. 1948. *Introduction to the theory of Fourier integrals*. Second edition, Oxford University Press.

[119] Scargle, J. D. 1981. Phase-sensitive deconvolution to model random processes, with special reference to astronomical data. In *Applied Time Series Analysis II* (D. F. Findley, ed.) 549-564. Academic Press, New York.

[120] Stigum, B. P. 1974. Asymptotic properties of dynamic stochastic parameter estimates (III). Journal of Multivariate Analysis, 4(4), 351-381.

[121] Vaughan, S. 2013. Random time series in astronomy. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 371, 20110549.

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

[123] White, J. S. 1958. The limiting distribution of the serial correlation coefficient in the explosive case. *The Annals of Mathematical Statistics*, 1188-1197.

[124] Wiggins, R. A. 1978. Minimum entropy deconvolution. *Geoexploration*, 16, 21-35.

[125] Wolf, R. 1848. Nachrichten von der Sternwarte in Berne. Sonnenflacken Beobachten. *Mittheilungen der Naturforschenden Gesellschaft in Bern*, 169-173.

[126] Wolfer, A. 1925. Observed sunspot relative numbers, 1749–1924. *Terrestrial Magnetism and Atmospheric Electricity*, 30, 83-86.

[127] Wu, R. 2011. Least absolute deviation estimation for general ARMA time series models with infinite variance. *Statistica Sinica*, 779-805.

[128] Wu, W., and S., Cambanis. 1991. Conditional variance of symmetric stable variables. *Stable Processes and Related Topics*, 85-99, Birkhäuser Boston.

[129] Wu, R., and R. A., Davis. 2010. Least absolute deviation estimation for general autoregressive moving average time-series models. *Journal of Time Series Analysis*, 31, 98-112.

[130] Yule, G. U. 1927. On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Philosophical transactions of the royal society A*, 226, 267-298.

[131] Zolotarev, V. M. 1986. *One-dimensional stable distributions.* Vol. 65, American Mathematical Society.

Titre : Processus linéaires α-stables anticipatifs pour l'analyse des séries temporelles: dynamique conditionnelle et estimation

Mots clés : Séries temporelles linéaires, Processus anticipatifs, Distributions stables multivariées, Dynamique conditionnelle

Résumé : Dans le contexte des séries temporelles linéaires, on étudie les processus strictement stationnaires dits *anticipatifs* dépendant potentiellement de tous les termes d'une suite d'erreurs α-stables indépendantes et identiquement distribuées. On considère en premier lieu les processus autoregressifs (AR) et l'on montre que des moments conditionnels d'ordres plus élevés que les moments marginaux existent dès lors que le polynôme caractéristique admet au moins une racine à l'intérieur du cercle unité. Des formules fermées sont obtenues pour les moments d'ordre un et deux dans des cas particuliers. On montre que la méthode des moindres carrés permet d'estimer une représentation *all-pass* causale du processus dont la validité peut être vérifiée par un test de type portmanteau, et l'on propose une méthode fondée sur des propriétés d'*extreme clustering* pour retrouver la représentation AR originale. L'AR(1) stable anticipatif est étudié en détails dans le cadre des vecteurs stables bivariés et des formes fonctionnelles pour les quatre premiers moments conditionnels sont obtenues pour toute paramétrisation admissible. Lors des évènements extrêmes, il est montré que ces moments deviennent équivalents à ceux d'une distribution de Bernoulli chargeant deux évolutions futures opposées: accroissement exponentiel ou retour aux valeurs centrales. Des résultats parallèles sont obtenus pour l'analogue de l'AR(1) en temps continu, le processus d'Ornstein-Uhlenbeck stable anticipatif. Pour des moyennes mobiles α-stables infinies, la distribution conditionnelle des chemins futurs sachant la trajectoire passée est obtenue lors des évènements extrêmes par le biais d'une nouvelle représentation des vecteurs stables multivariés sur des cylindres unités relatifs à des semi-normes. Contrairement aux normes, ce type de représentation donne lieu à une propriété de variations régulières des queues de distribution utilisable dans un contexte de prévision, mais tout vecteur stable n'admet pas une telle représentation. Une caractérisation est donnée et l'on montre qu'un chemin fini de moyenne mobile α-stable sera représentable pourvu que le processus soit "suffisamment anticipatif". L'approche s'étend aux processus résultant de la combinaison linéaire de moyennes mobiles α-stables, et la distribution conditionnelle des chemins futurs s'interprète naturellement en termes de reconnaissance de formes.

Title : Anticipative α-stable linear processes for time series analysis: conditional dynamics and estimation

Keywords : Linear time series, Anticipative processes, Multivariate stable distributions, Conditional dynamics

Abstract : In the framework of linear time series analysis, we study a class of so-called *anticipative* strictly stationary processes potentially depending on all the terms of an independent and identically distributed α-stable errors sequence. Focusing first on autoregressive (AR) processes, it is shown that higher order conditional moments than marginal ones exist provided the characteristic polynomials admits at least one root inside the unit circle. The forms of the first and second order moments are obtained in special cases. The least squares method is shown to provide a consistent estimator of an all-pass causal representation of the process, the validity of which can be tested by a portmanteau-type test. A method based on extreme residuals clustering is proposed to determine the original AR representation. The anticipative stable AR(1) is studied in details in the framework of bivariate α-stable random vectors and the functional forms of its first four conditional moments are obtained under any admissible parameterisation. It is shown that during extreme events, these moments become equivalent to those of a two-point distribution charging two polarly-opposite future paths: exponential growth or collapse. Parallel results are obtained for the continuous time counterpart of the AR(1), the anticipative stable Ornstein-Uhlenbeck process. For infinite α-stable moving averages, the conditional distribution of future paths given the observed past trajectory during extreme events is derived on the basis of a new representation of stable random vectors on unit cylinders relative to semi-norms. Contrary to the case of norms, such representation yield a multivariate regularly varying tails property appropriate for prediction purposes, but not all stable vectors admit such a representation. A characterisation is provided and it is shown that finite length paths of a stable moving average admit such representation provided the process is "anticipative enough". Processes resulting from the linear combination of stable moving averages are encompassed, and the conditional distribution has a natural interpretation in terms of pattern identification.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France