Prediction of bubbles in presence of α-stable aggregates moving averages

Gilles de Truchis[1], Sébastien Fries[2], Arthur Thomas[3]

September 7, 2026

Abstract

This paper introduces a novel framework based on α-stable moving average aggregates to model financial bubbles with heterogeneous growth and crash dynamics. We establish the theoretical properties of the model, showing in particular that it admits a semi-norm representation, which enables the prediction of extreme trajectories from past observations whenever each latent component is anticipative. We develop a minimum distance estimator based on the joint empirical characteristic function of consecutive observations and establish its consistency and asymptotic normality under suitable regularity conditions. Monte Carlo simulations confirm reliable finite-sample performance, and a subsampling procedure confirms convergence to the asymptotic Gaussian distribution, with confidence interval coverage reaching near-nominal levels for every parameter as the sample size grows. An empirical illustration to the CBOE Crude Oil ETF Volatility Index decomposes observed volatility dynamics into distinct latent components with different persistence properties, showing that what appears as a single explosive episode actually consists of multiple superimposed processes with heterogeneous growth rates and crash probabilities. A pattern-matching forecasting exercise based on this decomposition anticipates both the timing and magnitude of the March 2020 volatility spike.
Keywords: Aggregated processes, Stable random vectors, Spectral representation, Anticipative processes, Financial bubbles

[1]Université d'Orléans, Laboratoire d'Economie d'Orléans (LEO), 45000 Orléans, France, gilles.detruchis@univ-orleans.fr.
[2]Department of Econometrics and Data Science, Vrije Universiteit Amsterdam, Amsterdam, Netherlands
[3]Corresponding author: Université Paris Dauphine, Université PSL, LEDA, CNRS, IRD, 75016 Paris , France, arthur.thomas@dauphine.psl.eu.

1. Introduction

Financial markets regularly witness dramatic episodes where asset prices undergo rapid growth followed by abrupt collapses. These phenomena, termed rational asset pricing bubbles when they diverge from fundamental values (Blanchard and Watson , 1982; Tirole , 1985), have become increasingly prominent alongside well-documented features such as heavy-tailed distributions and volatility clustering, and emerge as solutions to linear rational expectation models that admit multiple stationary equilibria through infinite variance innovations (Gouriéroux et al. , 2020). Such patterns are incompatible with standard linear time series models, which fail to simultaneously accommodate heavy-tailed innovations, explosive growth, and the abrupt reversals characteristic of locally explosive episodes. From an empirical perspective, anticipative (or noncausal) models appear as good candidates to account for the non-linear dynamics of bubbles and the non-Gaussian environment. Such future-oriented models may generate intermittent periods of explosive growth and relative stability within a stationary linear framework, while also admitting a regular time representation involving nonlinear dynamics or non-i.i.d. innovations (Gourieroux and Jasiak , 2026; Fries and Zakoian , 2019). Beyond estimation, this class of models has been studied for its forecasting properties (Gouriéroux and Jasiak , 2016, 2018; Gourieroux and Jasiak , 2026) and its applications ranging from inflation and macroeconomic dynamics (Lanne and Saikkonen , 2011, 2013; Hecq et al. , 2020) to bubble modelling and tail risk (Fries and Zakoian , 2019; Gouriéroux et al. , 2025; Giancaterini et al. , 2026). This framework also exhibits intriguing properties, such as a predictive distribution with lighter tails than the marginal distribution, which enables more accurate predictions of higher-order moments (see e.g. Fries , 2022) and forecasts based on pattern recognition (see de Truchis et al. , 2026a), both critical for informed investment decisions.

A natural class of such models is provided by anticipative stable two-sided moving averages: stationary linear processes of the form

$$X_t = \sum_{k\in\mathbb{Z}} d_k \varepsilon_{t+k}, \qquad \varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{S}(\alpha,\beta,\sigma,0), \tag{1.1}$$

where (ε_t) is an i.i.d. sequence of α-stable innovations with tail index $\alpha \in (0,2]$, skewness $\beta \in [-1,1]$, scale $\sigma > 0$, and $(d_k)_{k\in\mathbb{Z}}$ is a real deterministic sequence satisfying appropriate summability conditions. Because the coefficient d_k is nonzero for $k > 0$, the process X_t depends on future innovations, making it anticipative. The simplest instance is the purely anticipative stable AR(1) studied in Gouriéroux and Zakoian (2017), which corresponds to $d_k = \psi^k \mathbf{1}_{k\geq 0}$, $|\psi| < 1$, and admits the two-sided moving-average representation

$$X_t = \sum_{k=0}^{+\infty} \psi^k \varepsilon_{t+k}, \qquad |\psi| < 1, \tag{1.2}$$

showing explicitly that X_t depends on current and future innovations only, with geometrically decaying coefficients at rate ψ. This framework extends naturally to mixed-causal autoregressive processes of orders (r,s), denoted MAR(r,s), defined by $\Phi(L)\Psi(L^{-1})X_t = \varepsilon_t$, where $\Phi(L) = \prod_{i=1}^{r}(1-\lambda_i L)$ with $|\lambda_i| < 1,\ \forall i \in \{1,\dots,r\}$ and $\Psi(L^{-1}) = \prod_{j=1}^{s}(1-\zeta_j L^{-1})$ with $|\zeta_j| < 1, \forall j \in \{1,\dots,s\}$. Under these root conditions, every MAR(r,s) process admits a two-sided MA(∞) representation of the form (1.1), with coefficients (d_k) that decay geometrically on each side of the origin (Gouriéroux and Jasiak , 2016; de Truchis and Thomas , 2026). The purely causal case ($s = 0$) reduces to a standard AR(r) model driven only by past innovations, while the purely noncausal case ($r = 0$) yields a process driven exclusively by future innovations.

However, anticipative models impose a similar increase rate for all bubbles, fully determined by the noncausal autoregressive coefficients (Gouriéroux and Zakoian , 2017).[4] This lack of flexibility may conflict with empirical evidence

[4] Lanne and Luoto (2013) introduce time-varying parameters in noncausal models not to allow growth rates to vary directly over time, but their approach, in a sense, addresses this issue.

on financial markets where the surge of explosive episodes can exhibit very different patterns. A natural motivation for heterogeneous bubble dynamics also arises from heterogeneous agent models: when fundamentalist and chartist traders interact, their competing strategies can generate distinct speculative components with different persistence properties (Agliari et al. , 2018), further supporting aggregation as a natural device for capturing the empirical complexity of bubble dynamics. A natural remedy is to consider processes resulting from the linear aggregation of $J \geq 1$ independent latent stable moving averages, each driven by its own innovation sequence. Formally, an α-stable aggregate is defined as

$$\mathcal{X}_t = \sigma \sum_{j=1}^{J} \pi_j X_{j,t}, \qquad X_{j,t} = \sum_{k \in \mathbb{Z}} d_{j,k}\, \varepsilon_{j,t+k}, \qquad \varepsilon_{j,t} \overset{\text{i.i.d.}}{\sim} \mathcal{S}(\alpha, \beta_j, 1, 0), \tag{1.3}$$

where $\sigma > 0$ is an overall scale parameter, $(\pi_j)_{j=1}^J$ are positive mixing weights summing to one, $(d_{j,k})_k$ are distinct deterministic coefficient sequences characterising the dynamics of each latent component $X_{j,t}$, and the innovation sequences $(\varepsilon_{j,t})$ are mutually independent. The tail index α is assumed common across all components. This restriction rests on two grounds. First, it follows from the stability under linear combinations: the aggregate $\mathcal{X}_t$ admits a well-defined α-stable distribution only when all components share the same tail index, since the tail of a sum of independent stable variables is governed by the heaviest one and the resulting distribution is tractable only in the equal-α case. Second, a common α preserves identifiability of the remaining parameters: the components are distinguished by their dynamic coefficients $(d_{j,k})_k$ and mixing weights (π_j), so that heterogeneity in bubble growth rates is fully captured by the autoregressive structure rather than by tail heterogeneity. As a consequence, all components share the same tail decay rate while differing in their persistence and crash dynamics; the asymmetry parameters β_j may however vary across components.

For example, the simplest specification takes each latent component $X_{j,t}$ to be a purely anticipative stable AR(1) process as in (1.2), with $d_{j,k} = \psi_j^k\, \mathbf{1}_{k \geq 0}$, $\psi_j \in (0, 1)$, so that

$$\mathcal{X}_t = \sigma \sum_{j=1}^{J} \pi_j X_{j,t}, \qquad X_{j,t} = \sum_{k=0}^{+\infty} \psi_j^k\, \varepsilon_{j,t+k}. \tag{1.4}$$

Because the ψ_j differ across components, each latent process generates a distinct bubble growth pattern, and their superposition produces richer and more heterogeneous explosive dynamics than any single anticipative process. Furthermore, as noted by Gouriéroux et al. (2021), aggregation introduces multiple independent sources of noise, making the stable aggregate structurally different from any mixed-causal or two-sided moving average model, and therefore better suited for derivative pricing where each risk factor must be separately identified and hedged.

The literature on the estimation of anticipative stable processes is relatively sparse. For non-aggregated mixed-causal AR models, an early semi-parametric approach is developed by Gassiat (1990, 1993), who establish consistency, asymptotic normality, and adaptivity for noncausal AR(p) processes with infinite-variance innovations, thereby covering the α-stable case; however, their procedure requires knowledge of the innovation density up to a scale parameter. Building on this line of work, Breidt et al. (1991) establish maximum likelihood theory for noncausal ARMA processes, which Andrews et al. (2009) extend to α-stable AR models in full generality. Velasco and Lobato (2018) develop frequency-domain minimum distance estimation for potentially non-invertible and noncausal ARMA models. Gourieroux and Jasiak (2023) propose the Generalized Covariance estimator (GCov), which identifies MAR(r, s) parameters by minimizing serial dependence in polynomial transformations of the pseudo-residuals; however, its asymptotic theory requires innovations with finite moments of a sufficiently high order, a condition incompatible with α-stable distributions. For aggregated processes, Gouriéroux and Zakoian (2017) propose a Cauchy-specific ($\alpha = 1$, $\beta = 0$) minimum distance estimator (MDE) based on the empirical characteristic function (ECF), deriving identification results for both continuous and discrete mixing distributions; nonetheless, their framework is confined to the Cauchy family and to purely anticipative AR(1) latent components. The ECF approach is grounded in the minimum distance theory of Knight and Yu (2002) and Xu and Knight (2010), who establish $\sqrt{n}$-consistency and asymptotic normality for ECF estimators in stationary time series, albeit under finite-variance assumptions that do not directly extend to the α-stable setting. Finally, Gouriéroux et al. (2021) study a Gaussian–Cauchy aggregate for WTI crude oil prices but do not establish formal asymptotic properties: no consistency result or central limit theorem is provided for the component-weight estimator, nor are the mixing conditions required under infinite-variance innovations formally verified.

In this paper, we make two contributions to the literature on econometric modelling of financial bubbles. First, we introduce the aggregation model (1.3), a novel flexible framework that overcomes a key limitation of existing anticipative heavy-tailed models, which impose uniform growth patterns across different bubble episodes. We derive the theoretical tail properties characterizing such an aggregation model. We demonstrate that the model admits a semi-norm representation[5], except when one of the underlying components exhibits purely non-anticipative behavior. This structural property allows us to predict extreme trajectories with heterogeneous growth patterns.

The main object underlying prediction in the α-stable framework is the spectral measure, which encodes the dependence structure of the process across time. In standard representations, this measure distributes mass in all directions of the space, mixing past and future indistinguishably. The semi-norm representation, introduced in de Truchis et al. (2026a), circumvents this limitation by replacing the standard norm with a semi-norm that assigns zero weight to future coordinates: the spectral mass is then entirely supported on directions that depend on the past alone. Economically, this means that the shape of current observations, specifically the magnitude and duration of the price surge, fully determines the asymptotic conditional distribution of the future trajectory. When this representation exists, the conditional probability that a bubble follows a given path converges to a well-defined limit as observations grow large, providing a coherent early-warning system for predicting bubble peaks.

Building on de Truchis et al. (2026a), we characterise when the aggregate admits such a representation. The main condition is simple: each latent component must be genuinely anticipative, in the sense that its forward-looking moving average coefficients do not vanish over arbitrarily long horizons. Intuitively, a purely causal component contributes innovations that are entirely invisible from any finite window of past observations, so that no amount of history can signal the bubble it drives, such a bubble always arrives as a complete surprise.

Second, we develop a minimum distance estimator based on the joint empirical characteristic function of consecutive observations $(\mathcal{X}_t, \mathcal{X}_{t+1})$, and we demonstrate its consistency and asymptotic normality under suitable regularity conditions. Departing from Gouriéroux and Zakoian (2017), who focus on continuous support distributions for the mixing weights in the specific Cauchy case, our approach handles discrete mixing weights and the full α-stable family with $\alpha \in (1, 2)$.[6] We first consider aggregates of anticipative AR(1) processes, then extend the framework to MAR$(1, 1)$ aggregates, and finally cover the general MAR(r, s) case, building on the exact two-sided MA(∞) coefficient representation derived in de Truchis and Thomas (2026); when no closed-form expression for the characteristic function is available, we show that, under a non-cancellation condition on the moving-average coefficients (Assumption 9), a numerically truncated approximation achieves a geometrically decaying approximation error at controlled computational cost.

Our methodology draws from Knight and Yu (2002) and Xu and Knight (2010), who developed asymptotic theory

[5] similarly to non-aggregated processes (de Truchis et al. , 2026a)

[6] The restriction $\alpha > 1$ ensures that the objective function $D_{\mathcal{X}}(\theta)$ belongs to $C^2(\Theta)$, a regularity condition required for the asymptotic theory of the estimator (Lemma 2.1). It also guarantees that the innovations have finite moments of all orders $\delta \in (0, \alpha)$, which is needed for the strong mixing property of the latent processes and the central limit theorem underlying asymptotic normality (Knight and Yu , 2002; Xu and Knight , 2010).

for minimum distance estimation using the empirical characteristic function in stationary time series, but we extend their approach to handle heavy-tailed stable distributions. We establish the asymptotic properties of our estimator under suitable regularity conditions, proving consistency and asymptotic normality. To numerically validate the finite-sample convergence toward the limiting Gaussian distribution, we conduct an extensive simulation study combining Monte Carlo experiments and subsampling diagnostics following Politis and Romano (1994) and Politis, Romano and Wolf (1999). The Monte Carlo analysis demonstrates that the estimator exhibits reliable convergence properties across various parameter configurations, though with moderate finite-sample biases. At $n = 500$ and $n = 1{,}000$, subsampling coverage is well below nominal for every parameter, and it improves steadily as n grows: by $n = 100{,}000$ all five, π_1 included, sit within a few points of nominal, and the same holds at $n = 500{,}000$, where estimation variance also shrinks almost exactly at the $1/\sqrt{n}$ rate. α is consistently the best covered of the five. None of the parameters, the autoregressive coefficients (ψ_j) and π_1 included, can be trusted in absolute terms below $n = 10{,}000$.

As an empirical illustration, we estimate an aggregation of purely anticipative stable AR(1) processes using the CBOE Crude Oil ETF Volatility Index (OVX) data collected at weekly frequency over May 23, 2015–April 27, 2025 ($T = 518$), and we demonstrate that the observed volatility patterns can be effectively decomposed into multiple latent stable components with heterogeneous persistence properties. The empirical analysis reveals that what initially appears as a single explosive episode actually consists of several superimposed processes with distinct autoregressive parameters and crash probabilities, each driven by an independent innovation sequence, pointing to the coexistence of multiple speculative components within a single observed bubble.

The remainder of this paper is organised as follows. Section 2 introduces the stable aggregates model and develops a new minimum distance estimator based on the characteristic functions of the unobserved latent components. Section 3 extends the representation theorem of de Truchis et al. (2026a) to stable aggregates and theoretically derives the conditions under which the forecast of a stable aggregate is possible. Section 4 documents the finite-sample performance of the minimum distance estimator through Monte Carlo simulations and implements a subsampling methodology to numerically verify the asymptotic normality of the estimator. An application to the CBOE Crude Oil ETF Volatility Index is proposed in Section 5. Section 6 concludes. All proofs are relegated to Appendix A, while the Online Supplement provides Monte Carlo results, subsampling diagnostics, convergence analysis, and additional results for the empirical illustration.

2. Estimating stable-aggregate of moving average

Consider X_t an α-stable moving average defined by

$$X_t = \sum_{k\in\mathbb{Z}} d_k \varepsilon_{t+k}, \qquad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha,\beta,\sigma,0) \tag{2.1}$$

with $d_0 > 0$, (d_k) a real deterministic sequence such that if $\alpha \neq 1$ or $(\alpha,\beta) = (1,0)$,

$$\sum_{k\in\mathbb{Z}} |d_k|^s < +\infty, \quad \text{for some} \quad s \in (0,\alpha)\cap(0,1], \tag{2.2}$$

and if $\alpha = 1$ and $\beta \neq 0$,

$$0 < \sum_{k\in\mathbb{Z}} |d_k| \Big| \ln |d_k| \Big| < +\infty. \tag{2.3}$$

For $d_k = \psi^k$, X_t is a simple strictly stationary anticipative AR(1). For X_t the strictly stationary solution of $\Psi(F)\Phi(B)X_t = \Theta(F)H(B)\varepsilon_t$, with F and B the lead and lag operators, the process belongs to the class of mixed-phase ARMA. Furthermore, if $\Theta = H = 1$, X_t is called a mixed causal–noncausal or MAR(r,s) process, where

$r = \deg(\Phi)$ and $s = \deg(\Psi)$. Adding the $(\alpha, \beta) = (1, 0)$ restrictions (let say $\mathcal{S}1\mathcal{S}$), X_t actually comes down to the so-called anticipative Cauchy AR(1) studied, e.g., in Gouriéroux and Jasiak (2018). As emphasized in the introduction, stable moving averages of the form (2.1) 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) in the context of explosive bubbles. They suggest to alleviate this restriction by considering processes resulting from the linear combination of different models.

Definition 2.1. *Let $(X_{1,t}), \ldots, (X_{J,t})$ be $J \geq 1$ stable moving averages, each satisfying (2.1)-(2.3), for some distinct coefficients sequences $(d_{j,k})_k$ and mutually independent error sequences $\varepsilon_{j,t} \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, \beta_j, 1, 0)$, $j = 1, \ldots, J$. Let also $(\pi_j)_{j=1,\ldots,J}$ be positive numbers summing to 1, $\sigma > 0$ be a scale parameter and define*

$$\mathcal{X}_t = \sigma \sum_{j=1}^{J} \pi_j X_{j,t}, \qquad \text{for} \quad t \in \mathbb{Z}.$$

We will call such process $\mathcal{X}_t$ a stable aggregate, and call $X_{j,t}$, $j = 1, \ldots, J$ the latent components of $\mathcal{X}_t$.

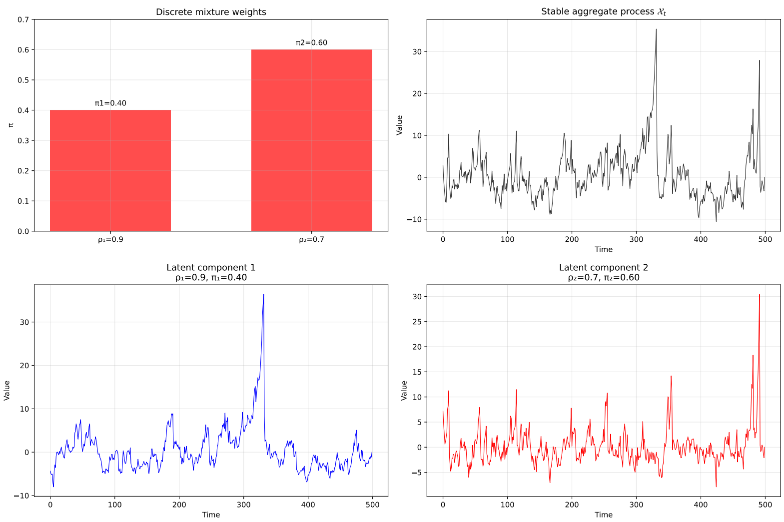

Figure 1: Simulated stable aggregate dynamics with two components. Top left: Distribution of weights for the two components with $\psi_1 = 0.90$, $\pi_1 = 0.40$ for the first component and $\psi_2 = 0.70$, $\pi_2 = 0.60$ for the second component. Top right: The resulting trajectory of the aggregated process $\mathcal{X}_t$. Middle and bottom panels: The individual latent component processes with different persistence parameters.

The estimator we propose is valid for any strictly stationary stable aggregate satisfying Definition 2.1, but in practice, it requires to formally derive the characteristic function of the latent components which can be tedious.

We provide the derivation for four parametric specifications. The first three admit a closed-form characteristic function: the aggregation of purely anticipative AR(1) processes, mixed causal-noncausal MAR(1, 1) processes , and the Gaussian-plus-AR(1) mixture of Gouriéroux et al. (2021). Notice that even in these specific frameworks, these aggregations feature much richer dynamics than single-component stable processes, as illustrated in Figure 1. The fourth specification extends the framework to general MAR(r, s) processes; in this case, no closed-form expression for the characteristic function is available, but we show that, under mild regularity conditions on the moving-average coefficients,[7] a truncated approximation of the two-sided MA(∞) representation of de Truchis and Thomas (2026) achieves a geometrically decaying approximation error, at a controlled computational cost. To disentangle the components of $\mathcal{X}_t$, our method leverages the independence of the latent processes and the resulting structure of the joint

[7] Namely, well-specified components (pairwise distinct causal and noncausal roots) and the non-cancellation condition of Assumption 9; we also restrict attention to $\alpha \in (1, 2)$ and $\beta_j = 0$ in this specification.

characteristic function:

$$\varphi_{\mathcal{X}}(u,v) = \mathbb{E}\Big(\exp\{i(u\mathcal{X}_t + v\mathcal{X}_{t+1})\}\Big) = \prod_{j=1}^{J} \varphi_{X_j}(\sigma\pi_j u, \sigma\pi_j v) \tag{2.4}$$

where φ_{X_j} is the joint characteristic function of a single latent component. In the rest of this section, we focus on $\beta_j = \beta$ for simplicity.

2.1. Case 1: Aggregation of Anticipative AR(1) Processes

We first restrict our attention to the case where each latent component $X_{j,t}$ is a purely anticipative AR(1) process. Its moving average representation is given by $d_{j,k} = \psi_j^k \mathbf{1}_{k\geq 0}$, with $0 < \psi_j < 1$. This restriction ensures that the asymmetry parameter β is preserved through the infinite summation defining each latent component $X_{j,t}$. From an economic perspective, positive autoregressive coefficients correspond to monotonic bubble growth patterns without oscillations, which is the empirically relevant case for financial applications modeling speculative bubbles.[8] The process is thus defined by $X_{j,t} = \sum_{k=0}^{\infty} \psi_j^k \varepsilon_{j,t+k}$. The joint characteristic function of the vector $(X_{j,t}, X_{j,t+1})$ is given by

$$\varphi_{X_j}(u,v) = \mathbb{E}\Big(\exp i(uX_{j,t} + vX_{j,t+1})\Big) = \mathbb{E}\Big(\exp i\big((u\psi_j + v)X_{j,t+1} + u\varepsilon_{j,t}\big)\Big), \tag{2.5}$$

for $(u,v) \in \mathbb{R}^2$. Due to the independence of the innovations, this simplifies to

$$\varphi_{X_j}(u,v) = \mathbb{E}\Big(\exp i(u\psi_j + v)X_{j,t+1}\Big)\mathbb{E}\Big(\exp iu\varepsilon_{j,t}\Big),$$

Assuming for simplicity a common asymmetry parameter $\beta_j = \beta$, we have for $\alpha \neq 1$

$$\log \mathbb{E}\Big(\exp i(u\psi_j + v)X_{j,t+1}\Big) = -\frac{|u\psi_j + v|^\alpha}{1 - |\psi_j|^\alpha}\left(1 - i\beta\,\mathrm{sign}(u\psi_j + v)\tan\left(\frac{\pi\alpha}{2}\right)\right)$$
$$\log \mathbb{E}(\exp iu\varepsilon_{j,t}) = -\left(1 - i\beta\,\mathrm{sign}(u)\tan\left(\frac{\pi\alpha}{2}\right)\right)|u|^\alpha.$$

The log-characteristic function of the aggregate is then obtained by substituting these expressions into Equation (2.4)

$$\log\varphi_{\mathcal{X}}(u,v) = -\sigma^\alpha \sum_{j=1}^{J} \pi_j^\alpha \left(\frac{|u\psi_j + v|^\alpha}{1 - |\psi_j|^\alpha}\left(1 - i\beta\mathrm{sign}(u\psi_j + v)\tan\frac{\pi\alpha}{2}\right) + |u|^\alpha\left(1 - i\beta\mathrm{sign}(u)\tan\frac{\pi\alpha}{2}\right)\right).$$

The Cauchy case examined in Gouriéroux and Zakoian (2017) is recovered for $\alpha = 1$, $\beta = 0$, leading to $\log \mathbb{E}(\exp iu\varepsilon_{j,t}) = -|u|$ and

$$\log\varphi_{\mathcal{X}}(u,v) = -\sigma \sum_{j=1}^{J} \pi_j \left(\frac{|u\psi_j + v|}{1 - |\psi_j|} + |u|\right).$$

As each latent component satisfies $|\psi_j| < 1$, the per-component summability condition (2.2) holds and the aggregate $\mathcal{X}_t$ is strictly stationary. Since J is finite in Definition 2.1, the following sum is automatically finite; we record it because it is the form of the strict stationarity condition that remains meaningful when the number of components grows (continuous aggregation, $J \to \infty$, see de Truchis et al. (2026b)):

$$\sum_{j=1}^{J} \frac{\pi_j^s}{1 - |\psi_j|^s} < \infty \quad \text{for } s \in (0,\alpha) \cap (0,1]. \tag{2.6}$$

[8] To see why this matters, recall that for a sum $\sum_{k=0}^{\infty} c_k u_k$ with $u_k \overset{i.i.d.}{\sim} \mathcal{S}(\alpha,\beta,\sigma,0)$, the resulting distribution is $\mathcal{S}(\alpha,\beta',\sigma',0)$ where $\beta' = \frac{\sum_{k=0}^{\infty}|c_k|^\alpha \mathrm{sign}(c_k)}{\sum_{k=0}^{\infty}|c_k|^\alpha} \cdot \beta$. When $\psi_j > 0$, all coefficients $c_k = \psi_j^k$ are positive, yielding $\beta' = \beta$. However, when $\psi_j < 0$, the coefficients alternate in sign, leading to $\beta' \neq \beta$. The case $\psi_j < 0$ would thus require a component-specific modified asymmetry parameter β'_j in the characteristic function.

2.2. *The minimum distance estimator*

As suggested by Knight and Yu (2002) and Gouriéroux and Zakoian (2017), one can rely on the empirical counterpart of the joint characteristic function (ECF) to build a minimum distance estimator (MDE). The ECF is simply defined as

$$\varphi_n(u,v) = \frac{1}{n-1}\sum_{t=1}^{n-1}\exp(i(u\mathcal{X}_t + v\mathcal{X}_{t+1})) \tag{2.7}$$

which can be decomposed into real and imaginary parts:

$$\varphi_n(u,v) = \frac{1}{n-1}\sum_{t=1}^{n-1}\cos(u\mathcal{X}_t + v\mathcal{X}_{t+1}) + \frac{1}{n-1}\sum_{t=1}^{n-1} i\sin(u\mathcal{X}_t + v\mathcal{X}_{t+1}) \tag{2.8}$$

By the law of large numbers, $\varphi_n(u,v) \xrightarrow{P} \varphi(u,v;\theta_0)$ as $n \to \infty$, where θ_0 denotes the true parameter values. For the sake of simplicity, we illustrate the parameter identification logic for the special case of the anticipative AR(1). The identification of $\theta = (\varsigma_1,\ldots,\varsigma_J,\psi_1,\ldots,\psi_J,\alpha,\beta)$, with $\varsigma_j := \sigma\pi_j$ (see Assumption 1 below), rests on the asymptotic behavior of the population characteristic function $\varphi(\cdot;\theta_0)$ as $(u,v)\to(0,0)$ along different directions.[9] For small values of u, the behavior of $\varphi(\cdot;\theta_0)$ is dominated by the α-stable distribution's properties. Specifically, for $u>0$,

$$\alpha = \lim_{u\to 0}\frac{\log\log|\varphi(u,0;\theta_0)|^{-1}}{\log|u|} \tag{2.9}$$

and

$$\beta = -\lim_{u\to 0^+}\frac{\mathrm{Im}(\log\varphi(u,0;\theta_0))}{\mathrm{Re}(\log\varphi(u,0;\theta_0))}\cdot\cot\frac{\pi\alpha}{2}. \tag{2.10}$$

For the identification of the remaining parameters, we exploit the behavior of the function

$$g(\lambda) = \lim_{u\to 0}\frac{\log|\varphi(u,\lambda u;\theta_0)|}{|u|^\alpha} = -\sum_{j=1}^{J}\varsigma_j^\alpha\left(\frac{|\psi_j+\lambda|^\alpha}{1-|\psi_j|^\alpha}+1\right) \tag{2.11}$$

for $v=\lambda u$ and $\lambda\in\mathbb{R}$. Two features of (2.11) deserve comment. First, the limit is not doing any real work: factoring $|u|^\alpha$ out of $\log|\varphi(u,\lambda u;\theta_0)|$ shows that the ratio $\log|\varphi(u,\lambda u;\theta_0)|/|u|^\alpha$ already equals $-\sum_j\varsigma_j^\alpha\big(|\psi_j+\lambda|^\alpha/(1-|\psi_j|^\alpha)+1\big)$ for every $u\neq 0$; (2.11) is thus an identity, not an asymptotic statement. Second, $(\varsigma_j,\psi_j)_{j\le J}$ are not recovered by evaluating g at a handful of points: since J itself is unknown, there is no finite system to solve, so instead one locates where g stops being smooth. Away from $\{-\psi_1,\ldots,-\psi_J\}$, g is a finite sum of real-analytic pieces, hence real-analytic itself; at $\lambda=-\psi_j$ each summand $|\psi_j+\lambda|^\alpha$ only reaches C^1 regularity when $\alpha\in(1,2)$, its second derivative blowing up since the exponent $\alpha-2$ is negative:

$$g''(\lambda) = -\alpha(\alpha-1)\sum_{j=1}^{J}\frac{\varsigma_j^\alpha}{1-|\psi_j|^\alpha}|\psi_j+\lambda|^{\alpha-2} \xrightarrow[\lambda\to-\psi_k]{} -\infty. \tag{2.12}$$

Because the ψ_j are pairwise distinct (Definition 2.1), the singular support of g'' is exactly $\{-\psi_1,\ldots,-\psi_J\}$: it delivers $\{\psi_1,\ldots,\psi_J\}$ as a set, with no need to know J in advance. Near each singularity only the matching summand blows up, so rescaling by $|\lambda+\psi_k|^{2-\alpha}$ kills every other term and isolates

$$\lim_{\lambda\to-\psi_k}|\lambda+\psi_k|^{2-\alpha}\,g''(\lambda) = -\frac{\alpha(\alpha-1)\,\varsigma_k^\alpha}{1-|\psi_k|^\alpha}, \tag{2.13}$$

[9] These limits are taken on $\varphi(\cdot;\theta_0)$ itself, not on the empirical characteristic function φ_n at a fixed sample size n: since $\log|\varphi_n(u,0)| = -S_n u^2/2 + O(u^4)$ as $u\to 0$, with S_n the sample variance of the observations, the corresponding ratios computed on φ_n would instead recover the Gaussian values 2 for α, and a ratio diverging like u^{-1} for β irrespective of θ_0. Equivalently, the limits in u and in n may be taken, provided $n\to\infty$ is taken first.

which pins down $\varsigma_k > 0$ once α and ψ_k are known. Since $w > 0$ everywhere, $D_0(\theta) = 0$ forces $\varphi(\cdot;\theta) = \varphi(\cdot;\theta_0)$ w-almost everywhere, hence everywhere by continuity; then $\alpha = \alpha_0$ and $\beta = \beta_0$ by (2.9)–(2.10), $g \equiv g_0$ on $\mathbb{R}$, and (2.12)–(2.13) determine the unordered family $\{(\varsigma_j, \psi_j)\}_{j\leq J}$. Combined with the ordering restriction of Assumption 1, this determines θ uniquely, so that Assumption 5 holds in this specification rather than being merely posited. Note that $\alpha > 1$ is what makes this work: for $\alpha \leq 1$ the singularity already sits in g', and the same argument goes through one derivative down.

Now we can define the MDE estimator as the minimizer of the objective distance measure

$$D_{\mathcal{X}}(\theta) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |\varphi_n(u,v) - \varphi(u,v;\theta)|^2 w(u,v)dudv \tag{2.14}$$

where $w(u,v)$ is a weighting function ensuring the convergence of the integral. The MDE estimator is then defined as

$$\hat{\theta}_n = \arg\min_{\theta} D_{\mathcal{X}}(\theta). \tag{2.15}$$

Knight and Yu (2002), show that under the following regularity conditions, the MDE estimator has standard limit theory. They suggest that it could accommodate α-stable models. Actually, some of their assumptions, listed hereafter, do not readily extend to the α-stable case. The characteristic functions of α-stable distributions are likely to exhibit singularities in their derivatives when $\alpha \in (0,2)$, particularly near points where $|\psi_j u + v|^\alpha$ vanishes. Without appropriate regularization through the weight function, these singularities can cause the integrals defining the first and second derivatives of (2.14) to diverge. Moreover, when $\alpha = 1$, an additional source of divergence arises in the first derivative of the characteristic exponent with respect to α, since the term $\tan(\pi\alpha/2)$ appearing in the standard parameterization satisfies $\frac{\partial}{\partial\alpha}\tan(\pi\alpha/2) = \frac{\pi/2}{\cos^2(\pi\alpha/2)} \to \infty$ as $\alpha \to 1$, regardless of the behavior of $|u|^\alpha$. The decisive reason is however structural rather than analytic: at $\alpha = 1$ with $\beta \neq 0$ the Samorodnitsky–Taqqu representation changes functional form, the term $\beta\tan(\pi\alpha/2)\,\mathrm{sign}(u)$ being replaced by $-\beta\frac{2}{\pi}\ln|u|$, so the parametric family is not even continuous in α at $\alpha = 1$ in the parameterization used here. Together with the requirement $\alpha - 2 > -1$ of the C^2 analysis, this is what makes $\alpha \in (1,2)$, rather than $\alpha \in (0,1) \cup (1,2)$, the natural domain. The following lemma establishes the precise conditions under which their regularity assumptions remain valid for α-stable aggregates.

Lemma 2.1. *Consider the MDE objective function defined by*

$$D_{\mathcal{X}}(\theta) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |\varphi_n(u,v) - \varphi(u,v;\theta)|^2 w(u,v)\,du\,dv \tag{2.16}$$

where $w(u,v) = (\kappa/\pi)\exp(-\kappa(u^2+v^2))$ with $\kappa > 0$ a positive constant.

Then, for any $\alpha \in (1,2)$, the objective function $D_{\mathcal{X}}(\theta)$ belongs to $C^2(\Theta)$ and Assumptions 3, 6, 8, as well as the integrability and continuity conditions on the second-order derivatives required by Assumption 7, are satisfied. Here the nonsingularity of $\Sigma(\theta_0)$ demanded by Assumption 7 is not an extra restriction: it follows from the model itself, and Appendix A.1.3 proves it. We keep it as an assumption in the general statement only because the same argument breaks down for the MAR(r,s) aggregates of Case 4.

Lemma 2.1, shows that we need to reduce the parameter space of α by introducing Assumption 2, and to replace the pointwise domination requirement (A6) of Knight and Yu (2002) by its integrated counterpart in Assumption 7, in order to recover their asymptotic theory in presence of α-stable distributions. It also reveals the critical role of the decaying exponential weights $w(u,v)$. Assumption 7 states the domination requirement in integrated rather than pointwise form, and this is not a matter of convenience: pointwise domination, as in condition (A6) of Knight and Yu (2002), cannot hold for α-stable scores. As ψ_k ranges over its compact domain, the singular line $\{\psi_k u + v = 0\}$ pivots and sweeps out a cone of positive Lebesgue measure, on which $\sup_{\theta\in\Theta}|\partial^2\varphi/\partial\psi_k^2| = +\infty$; no single w-integrable envelope can dominate the second derivatives uniformly over θ. What survives is weaker, but sufficient: for each fixed θ the w-integral of the second derivative is finite, and it varies continuously with θ. This is exactly what Appendix A.1.3 establishes. Assumption 4 is satisfied under the condition given by (2.6) or (2.24): these deliver strict stationarity, while ergodicity comes for free. $\mathcal{X}_t$ is a measurable function of a two-sided i.i.d. shift, and such a shift is ergodic, so any measurable factor of it is ergodic too, no summability condition is needed for that part. The stronger strong-mixing property, needed elsewhere for the central limit theorem, is established separately in the validation of Assumption 8. Assumption 5 is likewise not an extra posit in the specification of Case 1: it follows from the identification argument built above, (2.9) and (2.10) for α and β, (2.12)–(2.13) for the unordered family $\{(\varsigma_j, \psi_j)\}$, once combined with the ordering restriction of Assumption 1 that fixes the labelling. In the other specifications no such argument is available, and Assumption 5 is maintained as a genuine assumption.

Assumption 1. *$\theta = (\varsigma_1, \ldots, \varsigma_J, \psi_1, \ldots, \psi_J, \alpha, \beta) \in \Theta$, with $\varsigma_j := \sigma\pi_j > 0$ for $j = 1, \ldots, J$, where the parameter space $\Theta \subset \mathbb{R}^{2J+2}$ is a compact set with $\theta_0 \in Int(\Theta)$. The original parameters are recovered through $\sigma = \sum_{j=1}^{J} \varsigma_j$ and $\pi_j = \varsigma_j/\sigma$. We further require $\Theta \subset \{\theta : \psi_1 < \psi_2 < \cdots < \psi_J\}$.*

The ordering restriction is what rules out label switching. Formula (A.2) writes $\log \varphi_{\mathcal{X}}$ as a sum of one and the same function evaluated at each pair (ς_j, ψ_j), and a sum does not care about the order of its terms: for any permutation π of $\{1, \ldots, J\}$, relabelling θ into $\theta^\pi := (\varsigma_{\pi(1)}, \ldots, \varsigma_{\pi(J)}, \psi_{\pi(1)}, \ldots, \psi_{\pi(J)}, \alpha, \beta)$ leaves $\varphi(\cdot\,; \theta^\pi)$ identical to $\varphi(\cdot\,; \theta)$. If Θ treated all relabellings alike, $D_0(\theta) = 0$ would have $J!$ solutions instead of one, and Assumption 5 would be hopeless. Definition 2.1 does not help here: it rules out two components coinciding, not two distinct components being listed in the wrong order. What does help is forcing an order on them: $\psi_1 < \cdots < \psi_J$ singles out exactly one labelling per set of components. The constraint is open, and since the ψ_j are pairwise distinct by Definition 2.1, it can always be met at θ_0 by relabelling; choosing Θ as a compact set on which the inequalities stay strict throughout keeps both compactness and $\theta_0 \in \mathrm{Int}(\Theta)$ intact. Case 2 uses the same device with a lexicographic order on (ϕ_j, ψ_j), and Case 4 with an order on (r_j, s_j, θ_j).

The reduced parameterization $\varsigma_j = \sigma\pi_j$ is dictated by the structure of the model rather than adopted for computational convenience: the characteristic function (2.4) depends on $(\sigma, \pi_1, \ldots, \pi_J)$ only through the products ς_j. Carrying σ and $(\pi_1, \ldots, \pi_J)$ as $J+1$ free coordinates, with no relation linking them, would leave the scores tied by the exact, pointwise relation $\sigma\,\partial\varphi/\partial\sigma = \sum_{j=1}^{J} \pi_j\,\partial\varphi/\partial\pi_j$, so that $\Sigma(\theta_0)$ of Assumption 7 would be singular at every θ_0. Eliminating one coordinate through $\pi_1 + \ldots + \pi_J = 1$ would remove this particular degeneracy, but only at the cost of singling out one component as the one solved for, an arbitrary asymmetry the model itself does not impose. Working directly with $\varsigma = (\varsigma_1, \ldots, \varsigma_J)$ avoids both defects at once: it removes the redundancy without privileging any component, and (σ, π_j) are recovered afterwards.

Assumption 2. *The tail parameter space is such that $\alpha \in (1, 2)$ and $w(u, v)$ is the normalised exponential weight function of form $w(u, v) = (\kappa/\pi)\exp(-\kappa(u^2 + v^2))$ with $\kappa > 0$ a positive constant.*

The specific constant κ/π is not arbitrary: since $\iint_{\mathbb{R}^2} \exp\{-\kappa(u^2+v^2)\}\,du\,dv = \pi/\kappa$, it is exactly what turns w into a genuine $\mathcal{N}(0, (2\kappa)^{-1}I_2)$ probability measure, so Assumption (A8) of Knight and Yu (2002) holds by construction rather than by a separate argument. Yet nothing in the estimator or its limiting distribution actually depends on it[10]: the choice is made purely for convenience, and its payoff is that the weight constants $C_{w,\alpha}$ and $I_r(\alpha, \kappa)$ used below come out in closed form.

[10] Replacing w by cw multiplies $D_{\mathcal{X}}$ by c, hence leaves $\hat{\theta}_n$ unchanged, while it multiplies $\Sigma(\theta_0)$ by c and $\Omega(\theta_0)$ by c^2, so that $\Sigma(\theta_0)^{-1}\Omega(\theta_0)\Sigma(\theta_0)^{-1}$ is invariant.

Assumption 3. *With probability one, $D_{\mathcal{X}}(\theta)$ is twice continuously differentiable under the integral sign with respect to θ over Θ.*

Assumption 4. *The sequence $\{\mathcal{X}_t\}$ is strictly stationary and ergodic.*

Assumption 5. *Let $D_0(\theta) = \iint |\varphi(u,v;\theta_0) - \varphi(u,v;\theta)|^2 w(u,v)dudv$ and $D_0(\theta) = 0$ only if $\theta = \theta_0$.*

Assumption 6. *$K(x;\theta)$ is a measurable function of $x = (x_1, x_2) \in \mathbb{R}^2$ for all θ and bounded, where*

$$K(x;\theta) = \iint \Bigg[(\cos(ux_1 + vx_2) - Re\ \varphi(u,v;\theta))\frac{\partial Re\ \varphi(u,v;\theta)}{\partial\theta} \\ +(\sin(ux_1 + vx_2) - Im\ \varphi(u,v;\theta))\frac{\partial Im\ \varphi(u,v;\theta)}{\partial\theta}\Bigg] w(u,v)dudv. \tag{2.17}$$

For the observed series, we set $K_t := K((\mathcal{X}_t, \mathcal{X}_{t+1});\theta)$, $t = 1,\ldots,n-1$; note that the observation index is denoted t to avoid any confusion with the component index $j = 1,\ldots,J$.

Assumption 7. *The $p \times p$ matrix*

$$\Sigma(\theta_0) = \iint \left(\frac{\partial\varphi(u,v;\theta_0)}{\partial\theta}\right)\left(\frac{\partial\overline{\varphi}(u,v;\theta_0)}{\partial\theta'}\right) w(u,v)dudv$$

is nonsingular [11]*; for each $\theta \in \Theta$,*

$$\iint \left\|\frac{\partial^2\varphi(u,v;\theta)}{\partial\theta\partial\theta'}\right\| w(u,v)\,du\,dv < \infty,$$

and the map

$$\theta \longmapsto \frac{\partial^2\varphi(\cdot\,;\theta)}{\partial\theta\partial\theta'} \in L^1\big(w\,du\,dv\big)$$

is continuous, that is,

$$\lim_{\tilde\theta\to\theta} \iint \left\|\frac{\partial^2\varphi(u,v;\tilde\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2\varphi(u,v;\theta)}{\partial\theta\partial\theta'}\right\| w(u,v)\,du\,dv = 0 \qquad \text{for every } \theta \in \Theta.$$

Here $p = \dim\Theta$ is the dimension of the parameter vector introduced in Assumption 1, and it is not the same across specifications: $p = 2J+2$ for the anticipative AR(1) aggregate of Case 1, $p = 3J+2$ once the causal roots ϕ_j of Case 2's MAR(1,1) aggregate are added, and $p = J+1+\sum_{j=1}^J (r_j + s_j)$ for the MAR(r,s) aggregate of Case 4, where each component contributes its own order.

Assumption 8. *Let $\mathcal{F}_t$ be a σ-algebra such that $\{K_t, \mathcal{F}_t\}$ is an adapted stochastic sequence, where $K_t = K((x_t, x_{t+1});\theta_0)$. We can think of $\mathcal{F}_t$ as being the σ-algebra generated by the entire current and past history of K_t. Let $\nu_\ell = \mathbb{E}[K_0|K_{-\ell}, K_{-\ell-1},\ldots] - \mathbb{E}[K_0|K_{-\ell-1}, K_{-\ell-2},\ldots]$ for $\ell \geq 0$. Assume that $\mathbb{E}(K_0|\mathcal{F}_{-m})$ converges in mean square to 0 as $m \to \infty$ and $\sum_{\ell=0}^\infty \mathbb{E}[\nu_\ell'\nu_\ell]^{1/2} < \infty$.*

Assumption 8 is the projection condition (A7) of Knight and Yu (2002). ν_ℓ measures what the ℓ-th past observation adds to K_0: conditioning further back, on $K_{-\ell-1}, K_{-\ell-2},\ldots$, versus conditioning only up to $K_{-\ell}$, isolates that one increment, and requiring these increments to be summable is what pins down the weak dependence of the score sequence. Everything here is evaluated at θ_0, where $\mathbb{E}[K_0] = 0$. One notational note: with j already tied to the latent components, t denotes the observation index and ℓ the projection lag throughout.

[11] Two deliberate departures from the printed form of (A6) of Knight and Yu (2002) are worth recording. First, we write the Gram matrix with the complex conjugate, $\iint(\partial\varphi/\partial\theta)(\partial\overline{\varphi}/\partial\theta')w$, whereas Knight and Yu (2002) display it without: differentiating $|\varphi_n - \varphi(\theta)|^2$ twice produces $2\,\mathrm{Re}[\partial_\theta\varphi\,\overline{\partial_{\theta'}\varphi}]$, so the conjugate is the correct form; it is also what makes $\Sigma(\theta_0)$ positive semi-definite. By the evenness of w and the Hermitian symmetry $\varphi(-u,-v;\theta) = \overline{\varphi(u,v;\theta)}$, the resulting matrix is moreover real (see Remark 2.1), so no ambiguity arises in the sandwich. Second, the domination requirement is stated here in integrated form rather than as the pointwise envelope of (A6); as shown in Appendix A.1.3, no pointwise w-integrable envelope exists for α-stable aggregates, while the integrated form is verifiable and suffices for the proof of their Theorem 2.1.

Proposition 2.1. *Under Assumptions 1-8*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N\big(0, \Sigma(\theta_0)^{-1}\Omega(\theta_0)\Sigma(\theta_0)^{-1}\big) \tag{2.18}$$

where $\Sigma(\theta_0)$ is defined in Assumption 7, and $\Omega(\theta_0)$ is the long-run variance matrix of the score function $K_t = K((x_t, x_{t+1}); \theta_0)$ from Assumption 6

$$\Omega(\theta_0) = \Gamma(0) + \sum_{h=1}^{\infty} \big(\Gamma(h) + \Gamma(h)'\big), \qquad \Gamma(h) := Cov\big(K_1, K_{1+h}\big)$$

Remark 2.1. *Writing $f_\theta := \varphi_n - \varphi(\cdot\,;\theta)$, the first two derivatives of the criterion (2.14) are*

$$\partial_\theta D_{\mathcal{X}}(\theta) = -2 \iint \mathrm{Re}\big[\overline{f_\theta}\, \partial_\theta \varphi(\cdot\,;\theta)\big]\, w\, du\, dv, \tag{2.19}$$

$$\partial^2_{\theta\theta'} D_{\mathcal{X}}(\theta) = 2 \iint \mathrm{Re}\big[\partial_\theta \varphi\, \overline{\partial_{\theta'} \varphi}\big] w\, du\, dv \;-\; 2 \iint \mathrm{Re}\big[\overline{f_\theta}\, \partial^2_{\theta\theta'} \varphi\big] w\, du\, dv. \tag{2.20}$$

Since $\mathrm{Re}[\overline{f_\theta}\,\partial_\theta\varphi] = (\mathrm{Re}\,\varphi_n - \mathrm{Re}\,\varphi)\,\partial_\theta \mathrm{Re}\,\varphi + (\mathrm{Im}\,\varphi_n - \mathrm{Im}\,\varphi)\,\partial_\theta \mathrm{Im}\,\varphi$ and $\mathrm{Re}\,\varphi_n(u,v) = (n-1)^{-1}\sum_{t=1}^{n-1}\cos(u\mathcal{X}_t + v\mathcal{X}_{t+1})$, comparison of (2.19) with (2.17) gives the exact identity

$$\nabla D_{\mathcal{X}}(\theta_0) = -\,2\,\bar{K}_n, \qquad \bar{K}_n := \frac{1}{n-1}\sum_{t=1}^{n-1} K_t, \tag{2.21}$$

while the second integral in (2.20) vanishes in probability at $\theta = \theta_0$, so that

$$\nabla^2 D_{\mathcal{X}}(\theta_0) \;\longrightarrow\; 2\,\Sigma(\theta_0). \tag{2.22}$$

Substituting (2.21) and (2.22) into the usual first-order expansion of $\nabla D_{\mathcal{X}}(\hat{\theta}_n) = 0$ around θ_0 gives $\sqrt{n}(\hat{\theta}_n - \theta_0) = -\frac{1}{2}\Sigma(\theta_0)^{-1}\sqrt{n}\,\nabla D_{\mathcal{X}}(\theta_0) + o_P(1) = \Sigma(\theta_0)^{-1}\sqrt{n}\,\bar{K}_n + o_P(1)$: the $-\frac{1}{2}$ inherited from inverting $2\Sigma(\theta_0)$ exactly cancels the -2 inherited from (2.21), leaving a unit coefficient on $\bar{K}_n$. This is why $\Omega(\theta_0)$ can simply be the long-run variance of K_t itself, with no leftover scaling to carry through the sandwich. A second fact worth recording here is that $\Sigma(\theta_0)$ is real, symmetric and positive semi-definite: w is even and $\varphi(-u,-v;\theta) = \overline{\varphi(u,v;\theta)}$, so $\partial_i\varphi(-u,-v;\theta) = \overline{\partial_i\varphi(u,v;\theta)}$, and substituting $(u,v) \mapsto (-u,-v)$ in the defining integral shows $[\Sigma(\theta_0)]_{ij} = \overline{[\Sigma(\theta_0)]_{ij}}$. Nonsingularity therefore carries no complex-versus-real ambiguity, and $\Sigma(\theta_0)^{-1}\Omega(\theta_0)\Sigma(\theta_0)^{-1}$ comes out as a genuine real covariance matrix. There is a third departure from Knight and Yu (2002) to record here, alongside the two already noted under Assumption 7. Their Theorem 2.1 prints the long-run variance as $\mathrm{var}(K(x_1;\theta_0)) + 2\sum_{j\geq 2}\mathrm{cov}(K(x_1;\theta_0), K(x_j;\theta_0)) = \Gamma(0) + 2\sum_{h\geq 1}\Gamma(h)$, doubling $\Gamma(h)$ rather than adding its transpose; this only defines a symmetric matrix when $\Gamma(h) = \Gamma(h)'$ for every h, which holds automatically in their scalar setting but not once K_t is vector-valued, as it is here. The form actually used, $\Omega(\theta_0) = \Gamma(0) + \sum_{h\geq 1}\big(\Gamma(h) + \Gamma(h)'\big)$, is the correct long-run variance for a vector process and reduces to theirs exactly when $p = 1$.

The proof of this theorem is omitted as, by Lemma 2.1, it follows from Theorem 2.1 of Knight and Yu (2002), whose Hessian step is shown in Appendix A.1.3 to go through under the integrated domination condition of Assumption 7. Notice that in our α-stable framework, unlike Xu and Knight (2010), $\Sigma(\theta_0)$ and $\Omega(\theta_0)$ have no closed-form solutions.

2.3. Case 2: Aggregation of Mixed Causal-Noncausal MAR(1,1) Processes

We now consider a richer dynamic structure where each latent component $X_{j,t}$ is a mixed causal-noncausal MAR(1,1) process defined by $(1 - \phi_j L)(1 - \psi_j L^{-1})X_{j,t} = \varepsilon_{j,t}$, with $|\phi_j| < 1$ and $|\psi_j| < 1$. In this specification the parameter vector of Assumption 1 becomes

$$\theta = (\varsigma_1, \ldots, \varsigma_J, \phi_1, \ldots, \phi_J, \psi_1, \ldots, \psi_J, \alpha, \beta) \in \Theta \subset \mathbb{R}^{3J+2}, \qquad \varsigma_j = \sigma\pi_j,$$

of dimension $p = 3J + 2$, the causal roots ϕ_j adding one coordinate per component relative to Case 1. As in Case 1, the products ς_j rather than (σ, π_j) are the identified quantities. The corresponding MA(∞), $d_{j,k}$, coefficients are given by $\psi_j^k(1 - \phi_j\psi_j)^{-1}$ if $k \geq 0$ and $\phi_j^{|k|}(1 - \phi_j\psi_j)^{-1}$ for $k < 0$. The log-characteristic function for a single component X_j is derived from the linear combination of innovations

$$uX_{j,t} + vX_{j,t+1} = \sum_{k=-\infty}^{\infty} (ud_{j,k} + vd_{j,k-1})\varepsilon_{j,t+k}.$$

In the symmetric ($\mathcal{S}\alpha\mathcal{S}$) case, the log-characteristic function is

$$\log \varphi_{X_j}(u, v) = -\sum_{k=-\infty}^{\infty} |ud_{j,k} + vd_{j,k-1}|^\alpha.$$

We split the sum into its causal ($k \leq 0$) and noncausal ($k \geq 1$) parts. For the causal part ($k \leq 0$), the generic term is $ud_{j,k} + vd_{j,k-1} = (1 - \phi_j\psi_j)^{-1}(u\phi_j^{|k|} + v\phi_j^{|k-1|}) = (u + v\phi_j)(1 - \phi_j\psi_j)^{-1}\phi_j^{|k|}$. For the noncausal part ($k \geq 1$), the generic term is $ud_{j,k} + vd_{j,k-1} = (1 - \phi_j\psi_j)^{-1}(u\psi_j^k + v\psi_j^{k-1}) = (u\psi_j + v)(1 - \phi_j\psi_j)^{-1}\psi_j^{k-1}$. The sum becomes the sum of two geometric series

$$\begin{aligned}
\sum_{k=-\infty}^{\infty} |ud_{j,k} + vd_{j,k-1}|^\alpha &= \frac{1}{|1 - \phi_j\psi_j|^\alpha}\left(\sum_{k=-\infty}^{0} |(u + v\phi_j)\phi_j^{|k|}|^\alpha + \sum_{k=1}^{\infty} |(u\psi_j + v)\psi_j^{k-1}|^\alpha\right) \\
&= \frac{1}{|1 - \phi_j\psi_j|^\alpha}\left(|u + v\phi_j|^\alpha \sum_{l=0}^{\infty}(|\phi_j|^\alpha)^l + |u\psi_j + v|^\alpha \sum_{l=0}^{\infty}(|\psi_j|^\alpha)^l\right) \\
&= \frac{1}{|1 - \phi_j\psi_j|^\alpha}\left(\frac{|u + v\phi_j|^\alpha}{1 - |\phi_j|^\alpha} + \frac{|u\psi_j + v|^\alpha}{1 - |\psi_j|^\alpha}\right).
\end{aligned}$$

Finally, substituting this result into the aggregate function from Equation (2.4), we obtain the log-characteristic function for the MAR(1,1) aggregate for the $\mathcal{S}\alpha\mathcal{S}$ case . For the asymmetric case with $\alpha \neq 1$ we impose, for simplicity, that every component satisfies $\phi_j > 0$ and $\psi_j > 0$. The restriction has teeth once $\beta \neq 0$: when $\phi_j < 0$, the coefficients $(u + v\phi_j)\phi_j^{|k|}$ change sign as k varies, so the causal block picks up its own effective skewness, distinct from β (the same mechanism as the footnote of Case 1); the noncausal block behaves the same way through ψ_j. Left unrestricted, the general formula would have to carry these block- and component-specific skewness parameters, $\beta'_{j,\mathcal{C}}$ and $\beta'_{j,\mathcal{A}}$, separately. Under this restriction we obtain

$$\log \varphi_{\mathcal{X}}(u, v) = -\sum_{j=1}^{J} \frac{\varsigma_j^\alpha}{|1 - \phi_j\psi_j|^\alpha}\left(\mathcal{C}_j(u, v) + \mathcal{A}_j(u, v)\right) \tag{2.23}$$

where $\mathcal{C}_j(u, v)$ and $\mathcal{A}_j(u, v)$ represent the complex-valued contributions from the causal and noncausal dynamics of each component j, respectively

$$\begin{aligned}
\mathcal{C}_j(u, v) &= \frac{|u + v\phi_j|^\alpha}{1 - |\phi_j|^\alpha}\left(1 - i\beta\,\text{sign}(u + v\phi_j)\tan\left(\frac{\pi\alpha}{2}\right)\right), \\
\mathcal{A}_j(u, v) &= \frac{|u\psi_j + v|^\alpha}{1 - |\psi_j|^\alpha}\left(1 - i\beta\,\text{sign}(u\psi_j + v)\tan\left(\frac{\pi\alpha}{2}\right)\right).
\end{aligned}$$

Each component being strictly stationary ($|\phi_j| < 1$, $|\psi_j| < 1$), the aggregate is strictly stationary; the corresponding summability condition, again automatic for finite J, but meaningful under continuous aggregation, now reads

$$\sum_{j=1}^{J} \frac{\pi_j^s}{|1 - \phi_j\psi_j|^s}\left(\frac{1}{1 - |\psi_j|^s} + \frac{|\phi_j|^s}{1 - |\phi_j|^s}\right) < \infty \quad \text{for } s \in (0, \alpha) \cap (0, 1]. \tag{2.24}$$

This specification also needs $\phi_1, \ldots, \phi_J$ pairwise distinct and $\psi_1, \ldots, \psi_J$ pairwise distinct, on top of what Definition 2.1 already gives: that definition only asks the sequences $(d_{j,k})_k$ to differ, which leaves room for $\phi_1 = \phi_2$ as long as $\psi_1 \neq \psi_2$. With the added condition, the $2J$ lines along which the score fails to be analytic,

$$\mathcal{L}_j^{\mathcal{C}} = \{u + v\phi_j = 0\}, \qquad \mathcal{L}_j^{\mathcal{A}} = \{\psi_j u + v = 0\}, \qquad j = 1, \ldots, J,$$

are pairwise distinct. Two causal lines coincide only if $\phi_j = \phi_\ell$, two noncausal lines only if $\psi_j = \psi_\ell$, and a causal line can meet a noncausal one only if $\phi_j\psi_\ell = 1$, which the root conditions $|\phi_j|, |\psi_\ell| < 1$ already rule out on their own. The argument of Appendix A.1.3 then goes through unchanged, working on the connected components of $\mathbb{R}^2 \setminus \left(\{u=0\} \cup \{v=0\} \cup \bigcup_j (\mathcal{L}_j^{\mathcal{C}} \cup \mathcal{L}_j^{\mathcal{A}})\right)$. Both axes have to go this time, not just $u = 0$: differentiating with respect to ϕ_k brings down a factor v, exactly as differentiating with respect to ψ_k brings down a factor u, so either axis would silently kill the corresponding derivative. The direction transverse to $\mathcal{L}_k^{\mathcal{C}}$ is ∂_u (since $\partial_u(u + v\phi_k) = 1$) and to $\mathcal{L}_k^{\mathcal{A}}$ is ∂_v; one derivative in that direction isolates c_{ϕ_k} or c_{ψ_k}, a second brings in the logarithmic factor that isolates c_α, and comparing the two sides of the line then forces $c_{\varsigma_k} = c_\beta = 0$, just as in Case 1. One more restriction, $\phi_j \neq 0$ and $\psi_j \neq 0$ for every j (the boundary case of a purely noncausal component already belongs to Case 1), keeps every one of the $2J$ lines off the coordinate axes themselves. Under these conditions, $\Sigma(\theta_0)$ is nonsingular in Case 2 too.

2.4. Case 3: Aggregation of Mixed Stable and Gaussian Processes

Our estimation framework can also be extended to accommodate aggregates mixing α-stable and Gaussian components, an approach explored in Gouriéroux and Zakoian (2017) and Gouriéroux et al. (2021) but only for the Cauchy case. Consider a process $\mathcal{X}_t$ resulting from the aggregation of an α-stable MAR$(r,1)$, $r \in \{0,1\}$ with $\alpha \in (1,2)$ and a Gaussian AR(1) component $X_{\mathcal{N},t}$. As the distinction between causal and noncausal dynamics is unidentifiable when $\alpha = 2$, we adopt the standard causal specification for the Gaussian component. The log-characteristic function of the Gaussian AR(1) component $X_{\mathcal{N},t} = \phi_\mathcal{N} X_{\mathcal{N},t-1} + \eta_t$, $\eta_t \sim \mathcal{N}(0,1)$, for the vector $(X_{\mathcal{N},t}, X_{\mathcal{N},t+1})$ is given by

$$\log \varphi_\mathcal{N}(u,v) = -\frac{(u + v\phi_\mathcal{N})^2}{2(1-\phi_\mathcal{N}^2)} - \frac{v^2}{2}$$

The resulting aggregate log-characteristic function, $\log \varphi_\mathcal{X}(u,v)$, is the sum of the stable component's characteristic functions $\log \varphi_{X_j}(u,v)$ and $\log \varphi_\mathcal{N}(u,v)$, scaled by their respective aggregation weights as in Equation (2.4). A Gaussian log-characteristic function is quadratic in (u,v), so $\log \varphi_\mathcal{N}(\varsigma u, \varsigma v) = \varsigma^2 \log \varphi_\mathcal{N}(u,v)$: it scales with degree 2, not with degree α as the stable blocks do, $\log \varphi_{X_j}(\varsigma_j u, \varsigma_j v) = \varsigma_j^\alpha \tilde{\varphi}_j(u,v)$. Since $\alpha < 2$ throughout, these two exponents differ, and the aggregate has to carry both, $\log \varphi_\mathcal{X} = \sum_j \varsigma_j^\alpha \tilde{\varphi}_j + \varsigma_\mathcal{N}^2 \log \varphi_\mathcal{N}$, with $\varsigma_\mathcal{N}$ entering θ as its own coordinate rather than as a share of the stable components' scale. This is also why the recovery formulas $\sigma = \sum_j \varsigma_j$, $\pi_j = \varsigma_j/\sigma$ of Assumption 1 only make sense restricted to the stable block: a single σ feeding two blocks that scale at different powers would not multiply their contributions in the same proportion, so there is no coherent sense in which $\varsigma_\mathcal{N}$ could be read as $\sigma\pi_\mathcal{N}$ for a shared σ. This composite function can be directly employed in the MDE objective function (2.14). The estimator $\hat{\theta}_n$ defined in (2.15) remains valid because the stability index $\alpha = 2$ for the Gaussian component is fixed and not estimated. The Gaussian block $\log \varphi_\mathcal{N}$ is C^∞ in its parameters, and all its derivatives are polynomials in (u,v) times $\varphi_\mathcal{N}$, hence dominated by w-integrable functions uniformly on compacts. Adding it to $\log \varphi_\mathcal{X}$ therefore preserves the conclusion of Lemma 2.1: the extended criterion is again $C^2(\Theta)$ and the differentiability and integrability conditions required by Proposition 2.1 are satisfied, with $\Sigma(\theta_0)$ now formed from the score of the extended parameter vector. It should be noted that the lemma concerns the regularity of $D_\mathcal{X}$, not of $\log \varphi_\mathcal{X}$, which is not C^2 in θ: for fixed (u,v) with $u \neq 0$, $\psi_k \mapsto |\psi_k u + v|^\alpha$ fails to be twice differentiable at $\psi_k = -v/u$, and it is exactly this failure that makes the integrated treatment of Assumption 7 necessary. Nonsingularity of $\Sigma(\theta_0)$ delivers local identification of the extended parameter vector, the score family having full rank at θ_0; global identification remains Assumption 5, now bearing on the extended vector.

2.5. Case 4: General MAR(r,s) Aggregates

We now extend the estimation framework to aggregates of general MAR(r_j, s_j) processes with arbitrary causal order $r_j \geq 1$ and noncausal order $s_j \geq 1$. Consider a process $\mathcal{X}_t$ resulting from the aggregation of J independent

MAR(r_j, s_j) components:

$$\Phi_j(L)\Psi_j(L^{-1})X_{j,t} = \varepsilon_{j,t}, \quad j = 1, \ldots, J, \tag{2.25}$$

where $\Phi_j(L) = \prod_{i=1}^{r_j}(1 - \lambda_{j,i}L)$ with $|\lambda_{j,i}| < 1$, $\Psi_j(L^{-1}) = \prod_{l=1}^{s_j}(1 - \zeta_{j,l}L^{-1})$ with $|\zeta_{j,l}| < 1$, and $\varepsilon_{j,t} \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, 0, 1, 0)$. We restrict to the symmetric case $\beta_j = 0$ for all j to simplify the exposition. We take each component to be well-specified: the r_j causal roots pairwise distinct, the s_j noncausal roots pairwise distinct, a condition guaranteed below by $\mathfrak{s} > 0$ and needed for the partial-fraction decomposition (2.26) to reduce to simple poles. Nothing needs to be imposed across the two sides: $\Phi_j(z)$'s roots have modulus $|\lambda_{j,i}|^{-1} > 1$ while $\Psi_j(z^{-1})$'s have modulus $|\zeta_{j,l}| < 1$, so the polynomials never share a factor, and no causal root can be the reciprocal of a noncausal one. The stationarity restrictions $|\lambda_{j,i}|, |\zeta_{j,l}| < 1$ already do this work on their own, ruling out the transfer-function cancellation that would otherwise leave the order (r_j, s_j) over-specified.

Each component's roots serve as its parameters, $\theta_j = (\lambda_{j,1}, \ldots, \lambda_{j,r_j}, \zeta_{j,1}, \ldots, \zeta_{j,s_j}) \in \Theta_j$, with Θ_j a compact set of root configurations closed under conjugation: a real root is its own coordinate, a non-real one contributes the pair $(\operatorname{Re}\zeta, \operatorname{Im}\zeta)$ for one representative of its conjugate pair, so Θ_j sits in $\mathbb{R}^{r_j+s_j}$ either way and Φ_j, Ψ_j keep real coefficients. Every bound derived below is stated through the moduli $|\lambda_{j,i}|$, $|\zeta_{j,l}|$ and the separation $\mathfrak{s}$, so none of it cares whether a given root happens to be real or complex, and the full parameter vector to be estimated is $\theta = (\varsigma_1, \ldots, \varsigma_J, \theta_1, \ldots, \theta_J, \alpha) \in \mathbb{R}^{J+1+\sum_{j=1}^{J}(r_j+s_j)}$, collecting everything across components: the scaled mixing weights $\varsigma_j = \sigma\pi_j$ (as in Assumption 1, since the characteristic function depends on (σ, π_j) only through these products), the causal and noncausal roots of each component, and the shared tail index α.

Unlike the MAR$(1,1)$ case, the MA(∞) coefficients exhibit a multi-mode structure. By de Truchis and Thomas (2026), these coefficients admit the closed-form representation:

$$d_{j,k}(\theta_j) = \begin{cases} \displaystyle\sum_{l=1}^{s_j} A_{j,l}(\theta_j)\,\zeta_{j,l}^{k}, & k \geq 1, \\ \displaystyle\sum_{i=1}^{r_j} B_{j,i}(\theta_j)\,\lambda_{j,i}^{|k|}, & k \leq 0, \end{cases} \tag{2.26}$$

where $A_{j,l}$ and $B_{j,i}$ are rational functions of the roots given by partial-fraction expansion.[12] The well-specification condition ensures that all poles are simple and separated from one another, so these weights are uniformly bounded over Θ:

$$\sup_{\theta\in\Theta} |d_{j,k}(\theta_j)| \leq C_0\,\rho^{|k|}, \qquad k \in \mathbb{Z}, \tag{2.27}$$

where $C_0 := \max(r_j, s_j)\,\rho^{\min(r_j,s_j)-1}/\mathfrak{s}^{r_j+s_j-1}$ (the denominator of each partial-fraction weight contains $r_j + s_j - 1$ factors, each bounded below by $\mathfrak{s}$, and the numerators are bounded by ρ^{s_j-1} and ρ^{r_j-1} respectively, both dominated by $\rho^{\min(r_j,s_j)-1}$ since $\rho < 1$), $\rho := \sup_{\theta\in\Theta} \max_{1\leq j\leq J} \max(\max_i |\lambda_{j,i}|, \max_l |\zeta_{j,l}|) < 1$ (strictly below one by compactness of Θ inside the open stability region, parallel to the definition of $\mathfrak{s}$ as an infimum) and $\mathfrak{s} := \inf_{\theta\in\Theta} \min_{1\leq j\leq J} \min\{\min_{l\neq m} |\zeta_{j,l} - \zeta_{j,m}|, \min_{i\neq i'} |\lambda_{j,i} - \lambda_{j,i'}|, \min_{i,l} |\lambda_{j,i}\zeta_{j,l} - 1|\} > 0$ denotes the minimum separation distance between roots over the parameter space Θ. The constants are used uniformly over the components. Set

$$\bar{q} := \max_{1\leq j\leq J} \max(r_j, s_j), \qquad \underline{m} := \min_{1\leq j\leq J} \min(r_j, s_j), \qquad \bar{n} := \max_{1\leq j\leq J} (r_j + s_j),$$

[12] Specifically, by de Truchis and Thomas (2026),

$$A_{j,l}(\theta_j) = \frac{(-1)^{r_j}\,\zeta_{j,l}^{s_j-1}}{\prod_{m\neq l}(\zeta_{j,l} - \zeta_{j,m}) \cdot \prod_{i=1}^{r_j}(\lambda_{j,i}\zeta_{j,l} - 1)} \text{ and } B_{j,i}(\theta_j) = \frac{(-1)^{s_j}\,\lambda_{j,i}^{r_j-1}}{\prod_{i'\neq i}(\lambda_{j,i} - \lambda_{j,i'}) \cdot \prod_{l=1}^{s_j}(\lambda_{j,i}\zeta_{j,l} - 1)}.$$

and assume without loss of generality that $\mathfrak{s} \le 1$ (otherwise replace $\mathfrak{s}$ by $\mathfrak{s} \wedge 1$, which only weakens the bounds). With this notation (2.27) holds with the explicit constant

$$C_0 = \bar{q}\, \rho^{\underline{m}-1}\, \mathfrak{s}^{-(\bar{n}-1)}. \tag{2.28}$$

Differentiating (2.26) yields analogous bounds for the derivatives:

$$\sup_{\theta\in\Theta} |\partial_\theta d_{j,k}(\theta_j)| \le C_1(1+|k|)\rho^{|k|}, \quad \sup_{\theta\in\Theta} |\partial^2_{\theta\theta'} d_{j,k}(\theta_j)| \le C_2(1+k^2)\rho^{|k|}, \tag{2.29}$$

with

$$C_1 = \bar{q}\, \mathcal{A}_1 + 2\rho^{\underline{m}-2}\, \mathfrak{s}^{-(\bar{n}-1)}, \qquad C_2 = \bar{q}\big(\mathcal{A}_2 + 2\rho^{-1}\mathcal{A}_1\big) + 4\rho^{\underline{m}-3}\, \mathfrak{s}^{-(\bar{n}-1)}, \tag{2.30}$$

and where

$$\mathcal{A}_1 := \bar{q}\, \mathfrak{s}^{-(\bar{n}-1)} + \bar{n}\, 2^{\bar{n}-1}\mathfrak{s}^{-2(\bar{n}-1)}, \qquad \mathcal{A}_2 := \bar{q}^{\,2}\mathfrak{s}^{-(\bar{n}-1)} + \big(2\bar{q}\bar{n} + \bar{n}^2\big)2^{\bar{n}-1}\mathfrak{s}^{-2(\bar{n}-1)} + 2\bar{n}^2 2^{2\bar{n}-2}\mathfrak{s}^{-3(\bar{n}-1)}$$

bound the first and second derivatives of the partial-fraction weights; the derivation is a routine, if lengthy, application of the quotient and Leibniz rules to the explicit weights $A_{j,l}, B_{j,i}$, and is deferred to Appendix A.2. Case 4 loses the proportionality $d_{j,k}/d_{j,k-1} =$ const. that gave closed-form characteristic functions in Cases 1-2 as soon as $\max(r_j, s_j) > 1$; we fall back on a truncated characteristic function instead.

Every bound used below concerns the pair $(d_{j,k}, d_{j,k-1})$, not $d_{j,k}$ on its own, which is why it pays to fold the cost of the shift into the constants directly rather than track it separately at each occurrence. Define

$$\bar{C}_0 := \rho^{-1}C_0, \qquad \bar{C}_1 := 2\rho^{-1}C_1, \qquad \bar{C}_2 := 3\rho^{-1}C_2, \tag{2.31}$$

so that, for all $k \in \mathbb{Z}$, $j \in \{1, \dots, J\}$ and $\theta \in \Theta$,

$$\max\big(|d_{j,k}|, |d_{j,k-1}|\big) \le \bar{C}_0\, \rho^{|k|}, \quad \max\big(|\partial_\theta d_{j,k}|, |\partial_\theta d_{j,k-1}|\big) \le \bar{C}_1(1+|k|)\rho^{|k|}, \quad \max\big(\|\partial^2_{\theta\theta'} d_{j,k}\|, \|\partial^2_{\theta\theta'} d_{j,k-1}\|\big) \le \bar{C}_2(1+k^2)\rho^{|k|}. \tag{2.32}$$

Where do the factors ρ^{-1}, $2\rho^{-1}$ and $3\rho^{-1}$ come from? Split on the sign of k. When $k \ge 1$, $|k-1| = |k| - 1$, so trading $d_{j,k}$ for $d_{j,k-1}$ costs one power of ρ^{-1} and nothing else. When $k \le 0$ the opposite happens: $\rho^{|k-1|} \le \rho^{|k|}$, so the exponential part is free, but the polynomial envelope is not, since $1 + |k-1| = 2 + |k| \le 2(1 + |k|)$ and $1 + (k-1)^2 = 2 + k^2 + 2|k| \le 3(1 + k^2)$, the latter inequality tightest at $k = -1$. Because $\rho^{-1} > 1$, multiplying rather than choosing between the two penalties still gives a valid, and simpler, bound in both regimes at once, which is (2.32). As an immediate corollary, $\|\boldsymbol{d}^1_{j,k}\| \le \sqrt{2}\, \bar{C}_0\, \rho^{|k|}$.

For notational brevity, set $\Delta_{j,k}(u, v; \theta_j) := u\, d_{j,k}(\theta_j) + v\, d_{j,k-1}(\theta_j)$ and $\boldsymbol{d}^1_{j,k}(\theta_j) := (d_{j,k}(\theta_j), d_{j,k-1}(\theta_j)) \in \mathbb{R}^2$. The exact joint log-CF of $(X_{j,t}, X_{j,t+1})$ is $\log\varphi_{X_j}(u, v; \theta_j) = -\sum_{k\in\mathbb{Z}} |\Delta_{j,k}|^\alpha$. For $M \ge 1$, define the truncated log-CF:

$$\log\varphi^{(M)}_{X_j}(u, v; \theta_j) := -\sum_{k=-M}^{M} |\Delta_{j,k}(u, v; \theta_j)|^\alpha, \tag{2.33}$$

the aggregate truncated CF:

$$\varphi^{(M)}_{\mathcal{X}}(u, v; \theta) = \prod_{j=1}^{J} \exp\big\{\log\varphi^{(M)}_{X_j}(\varsigma_j u,\ \varsigma_j v;\ \theta_j)\big\}, \tag{2.34}$$

and the MDE objective:

$$D^{(M)}_{\mathcal{X}}(\theta) = \iint_{\mathbb{R}^2} \big|\varphi_n(u, v) - \varphi^{(M)}_{\mathcal{X}}(u, v; \theta)\big|^2 w(u, v)\, du\, dv,$$

with $w(u, v) = (\kappa/\pi) \exp\{-\kappa(u^2 + v^2)\}$ as in Assumption 2.

The analysis of the second-order derivatives of the truncated objective function additionally requires a geometric lower bound on the norm of consecutive pairs of coefficients. Such a bound cannot be derived from the partial-fraction decomposition (2.26) alone: the moduli of the dominant roots on the causal and noncausal sides may differ, so that one side decays strictly faster than $\rho^{|k|}$; and cancellations between terms sharing the same modulus (e.g., complex-conjugate pairs of roots, which the real-coefficient MAR(r_j, s_j) model allows for $\max(r_j, s_j) \geq 2$) may drive $d_{j,k}$ and $d_{j,k-1}$ simultaneously close to zero along subsequences of lags. We therefore impose the non-cancellation property as an explicit assumption.

Assumption 9. *There exist $c_v > 0$ and $\rho_1 \in (0, 1)$ satisfying the rate condition $2\log(1/\rho) > (2 - \underline{\alpha}) \log(1/\rho_1)$, where $\underline{\alpha} := \min\{\alpha : \theta \in \Theta\} > 1$, such that*

$$\|\boldsymbol{d}^1_{j,k}(\theta_j)\| \geq c_v\, \rho_1^{|k|}, \qquad \text{for all } k \in \mathbb{Z},\ j \in \{1, \dots, J\},\ \theta \in \Theta,$$

the constants c_v and ρ_1 being independent of k, j and θ.

Remark 2.2. *Assumption 9 is phrased directly in terms of the coefficients $d_{j,k}$, which have no closed form once $\max(r_j, s_j) > 1$ and are therefore awkward to check on a given specification. The three conditions below bear only on the roots themselves and imply the assumption.*

(ι) Dominant root. On each side the root of maximal modulus is real, simple and dominant uniformly in the gap: writing, for each θ and each j,

$$\rho_\zeta(\theta_j) := \max_{1 \leq l \leq s_j} |\zeta_{j,l}(\theta_j)|, \qquad \rho_\lambda(\theta_j) := \max_{1 \leq i \leq r_j} |\lambda_{j,i}(\theta_j)|,$$

there exist $\epsilon > 0$ and an indexing, both independent of θ, such that

$$|\zeta_{j,1}(\theta_j)| = \rho_\zeta(\theta_j) \quad \text{and} \quad |\zeta_{j,l}(\theta_j)| \leq \rho_\zeta(\theta_j) - \epsilon \quad (l \geq 2), \qquad \text{for all } \theta \in \Theta,$$

and symmetrically for the $\lambda_{j,i}$. It is the gap ϵ that is uniform over Θ, not the dominant modulus itself: requiring the latter to be constant would confine Θ to a level set and contradict $\theta_0 \in \mathrm{Int}(\Theta)$.

($\iota\iota$) Roots bounded away from the origin. $\inf_{\theta \in \Theta} \min_{j,l,i} \big(|\zeta_{j,l}| \wedge |\lambda_{j,i}|\big) =: \underline{r} > 0$.

($\iota\iota\iota$) No cancellation at short lags. $\inf_{\theta \in \Theta} \min_j \min_{|k| < k_*} \|\boldsymbol{d}^1_{j,k}(\theta_j)\| > 0$, with k_* as defined below.

Together, (ι) and ($\iota\iota$) settle the long lags; ($\iota\iota\iota$) mops up the finitely many that are left. With $a := \inf_{\theta \in \Theta} \min_{1 \leq j \leq J} |A_{j,1}|$ and $\bar{A} := \sup_{\theta \in \Theta} \max_{1 \leq j \leq J} \sum_{l \geq 2} |A_{j,l}|$, condition ($\iota\iota$) already gives $a > 0$, since the numerator $\zeta_{j,1}^{s_j - 1}$ is bounded below by $\underline{r}^{\,s_j - 1}$ and the denominator above on the compact Θ, while $\bar{A} < \infty$ is immediate from (2.27). For $k \geq 1$,

$$|d_{j,k}| \;\geq\; |A_{j,1}|\, \rho_\zeta^k - \sum_{l \geq 2} |A_{j,l}|\, (\rho_\zeta - \epsilon)^k \;\geq\; \rho_\zeta^k \Big[a - \bar{A}\big(1 - \epsilon/\rho_\zeta\big)^k \Big] \;\geq\; \tfrac{a}{2}\, \rho_\zeta^k$$

as soon as $k \geq k_\zeta := \big\lceil \sup_{\theta \in \Theta} \max_j \log(2\bar{A}/a) \,/\, \log\big(\rho_\zeta(\theta_j)/(\rho_\zeta(\theta_j) - \epsilon)\big) \big\rceil$, finite by compactness and by condition ($\iota\iota$), with the convention $k_\zeta := 1$ when $s_j = 1$ for every j, in which case $|d_{j,k}| = |A_{j,1}| \rho_\zeta(\theta_j)^k$ exactly. The causal side is symmetric, with its own threshold k_λ, and we set $k_* := \max(k_\zeta, k_\lambda)$. Define

$$\rho_1 := \inf_{\theta \in \Theta} \min_{1 \leq j \leq J} \min\big(\rho_\zeta(\theta_j), \rho_\lambda(\theta_j)\big) \;>\; 0,$$

positive by compactness and ($\iota\iota$). Then $|d_{j,k}| \geq \frac{a}{2} \rho_\zeta(\theta_j)^k \geq \frac{a}{2} \rho_1^k$ for $k \geq k_*$, and c_v taken as the smallest of $a/2$, its causal counterpart, and the quantity in ($\iota\iota\iota$) delivers Assumption 9. Finally, (2.27) forces $\rho_1 \leq \rho$ in every case.

Lemma 2.3 and its proof never call on the rate condition in its logarithmic form; what recurs is a single constant, so we introduce it now. Set

$$\bar{\rho} := \rho^2\, \rho_1^{\underline{\alpha}-2} \;=\; \sup_{\theta\in\Theta} \rho^2 \rho_1^{\alpha-2},$$

which no longer depends on θ. Assumption 9's rate condition is exactly $\bar{\rho} < 1$, and since $\rho_1 \le \rho$ with $\underline{\alpha} - 2 < 0$, we also get $\bar{\rho} \ge \rho^{\underline{\alpha}} \ge \rho^{\alpha}$ for every α admissible under Assumption 1.

Lemma 2.2. *For every $k \in \mathbb{Z}$, j, $\theta \in \Theta$, and $\alpha \in (1,2)$:*

$$\iint_{\mathbb{R}^2} |\Delta_{j,k}|^{\alpha}\, w(u,v)\, du\, dv = C_{w,\alpha}\, \|\boldsymbol{d}^1_{j,k}\|^{\alpha}, \tag{2.35}$$

where $C_{w,\alpha} := \frac{\kappa}{\pi} \int_{\mathbb{R}} |s|^{\alpha} e^{-\kappa s^2}\, ds \cdot \int_{\mathbb{R}} e^{-\kappa t^2}\, dt = \frac{\Gamma\left(\frac{\alpha+1}{2}\right)}{\sqrt{\pi}\,\kappa^{\alpha/2}} < \infty$.

Lemma 2.3. *Let Assumption 9 hold and recall $\bar{\rho} = \rho^2 \rho_1^{\underline{\alpha}-2} \in [\rho^{\alpha}, 1)$. For $\alpha \in (1,2)$, the following bounds hold for every $M \ge 1$, $j \in \{1, \dots, J\}$, and $\theta \in \Theta$:*

$$\iint \left|\log \varphi_{X_j} - \log \varphi^{(M)}_{X_j}\right| w\, du\, dv \le \mathfrak{c}_0\, \rho^{\alpha M}, \tag{2.36}$$

$$\iint \left|\partial_\theta [\log \varphi_{X_j} - \log \varphi^{(M)}_{X_j}]\right| w\, du\, dv \le \mathfrak{c}_1 (1+M)\, \rho^{\alpha M}, \tag{2.37}$$

$$\iint \left|\partial^2_{\theta\theta'} [\log \varphi_{X_j} - \log \varphi^{(M)}_{X_j}]\right| w\, du\, dv \le \mathfrak{c}_2 (1+M)^2\, \bar{\rho}^{\,M}, \tag{2.38}$$

where the constants $\mathfrak{c}_0$, $\mathfrak{c}_1$, $\mathfrak{c}_2$ are given in closed form by (A.28), (A.30) and (A.33) below, and depend on $(\alpha, \kappa, \rho, \rho_1, c_v, \mathfrak{s})$ and on the orders (r_j, s_j) through $\bar{q}$, $\underline{m}$, $\bar{n}$, but not on M, k or j.

The bounds above are for the unscaled $\log \varphi_{X_j}(u, v; \theta_j)$. The scaled quantity $\log \varphi_{X_j}(\varsigma_j u, \varsigma_j v; \theta_j)$, which is what Lemma 2.4 actually needs, obeys the same bounds after multiplication by the explicit factors in $V_\gamma := \sup_{\theta\in\Theta} \max_j \varsigma_j^{\gamma}$ and $\ell_\varsigma := \sup_\Theta \max_j |\ln \varsigma_j|$ recorded in the proof, and summing over the components multiplies them by J.

Lemma 2.4. *Under Assumption 2 and Assumption 9, with $\alpha \in (1,2)$ and $\beta_j = 0$ for all j:*

(ι) *For every fixed $M \ge 1$, $D^{(M)}_{\mathcal{X}} \in C^2(\Theta)$, and Assumptions 3, 6, 8, as well as the integrability and continuity conditions of Assumption 7, are satisfied with $\varphi_{\mathcal{X}}$ replaced by $\varphi^{(M)}_{\mathcal{X}}$; moreover, if $\Sigma(\theta_0)$ is nonsingular, there exists $M_0 \ge 1$ such that the truncated Gram matrix $\Sigma^{(M)}(\theta_0)$ is nonsingular for every $M \ge M_0$.*

($\iota\iota$) *Let*

$$G(\theta) := \iint \partial_\theta \left|\varphi_n - \varphi_{\mathcal{X}}(\cdot\,;\theta)\right|^2 w\, du\, dv, \qquad H(\theta) := \iint \partial^2_{\theta\theta'} \left|\varphi_n - \varphi_{\mathcal{X}}(\cdot\,;\theta)\right|^2 w\, du\, dv,$$

both finite for every $\theta \in \Theta$, and let $G^{(M)}$, $H^{(M)}$ be the corresponding integrals for $\varphi^{(M)}_{\mathcal{X}}$, so that $G^{(M)} = \nabla D^{(M)}_{\mathcal{X}}$ and $H^{(M)} = \nabla^2 D^{(M)}_{\mathcal{X}}$ by (ι). Then, as $M \to \infty$,

$$\sup_{\theta\in\Theta} |D^{(M)}_{\mathcal{X}}(\theta) - D_{\mathcal{X}}(\theta)| = O(\rho^{\alpha M}),$$

$$\sup_{\theta\in\Theta} \|G^{(M)}(\theta) - G(\theta)\| = O\big((1+M)\rho^{\alpha M}\big),$$

$$\sup_{\theta\in\Theta} \|H^{(M)}(\theta) - H(\theta)\| = O\big((1+M)^2 \bar{\rho}^{\,M}\big).$$

By ($\iota\nu$) below, $G = \nabla D_{\mathcal{X}}$ and $H = \nabla^2 D_{\mathcal{X}}$, so that the last two displays are indeed statements about the derivatives of the untruncated criterion.

($\iota\iota\iota$) *The truncated scores converge to the full scores in $L^2_w(\mathbb{R}^2)$: as $M \to \infty$,*

$$\sup_{\theta\in\Theta} \iint_{\mathbb{R}^2} \left\|\partial_\theta \varphi^{(M)}_{\mathcal{X}}(u,v;\theta) - \partial_\theta \varphi_{\mathcal{X}}(u,v;\theta)\right\|^2 w(u,v)\, du\, dv = O\big((1+M)^2 \rho^{2\alpha M}\big).$$

($\iota\nu$) *The untruncated criterion satisfies the same conditions: $D_{\mathcal{X}} \in C^2(\Theta)$, so that Assumption 3 holds; the score $K(\cdot\,;\theta)$ built from $\varphi_{\mathcal{X}}$ is bounded, so that Assumption 6 holds; Assumption 8 holds; and $\theta \mapsto \partial^2_{\theta\theta'}\varphi_{\mathcal{X}}(\cdot\,;\theta)$ is finite and continuous in $L^1(w\, du\, dv)$, so that the integrability and continuity clauses of Assumption 7 hold. Also, $\theta \mapsto \partial_\theta \varphi_{\mathcal{X}}(\cdot\,;\theta)$ is continuous from Θ into $L^2_w(\mathbb{R}^2)$*

Proposition 2.2. *Let $\mathcal{X}_t$ be a strictly stationary $\mathcal{S}\alpha\mathcal{S}$ aggregate as in Definition 2.1 whose latent components are MAR(r_j, s_j) processes with $\beta_j = 0$ and $\alpha \in (1,2)$. Assume 1 (in the Case-4 parameterization), 2, 4, 5, 9, that $\Sigma(\theta_0)$ is non-singular. Assumptions 3, 6, 8 and the integrability and continuity clauses of Assumption 7 then hold for $\varphi^{(M)}_{\mathcal{X}}$ by Lemma 2.4(ι), and for $\varphi_{\mathcal{X}}$ by Lemma 2.4($\iota\nu$); the proof in Appendix A.2 uses the truncated form for the estimator and the untruncated form for the limiting objects. Moreover, let $(M_n)_{n\geq 1}$ satisfy*

$$M_n \geq c_\star \log n, \qquad c_\star > \frac{1}{2\log(1/\bar{\rho})}, \qquad \bar{\rho} = \rho^2 \rho_1^{\frac{\alpha-2}{}}. \tag{2.39}$$

Then $\hat{\theta}^{(M_n)}_n = \arg\min_{\theta\in\Theta} D^{(M_n)}_{\mathcal{X}}(\theta)$ satisfies:

$$\hat{\theta}^{(M_n)}_n \xrightarrow{P} \theta_0, \qquad \sqrt{n}\big(\hat{\theta}^{(M_n)}_n - \theta_0\big) \xrightarrow{d} \mathcal{N}\big(0,\, \Sigma(\theta_0)^{-1}\Omega(\theta_0)\Sigma(\theta_0)^{-1}\big),$$

with $\Sigma(\theta_0)$, $\Omega(\theta_0)$ defined analogously to Proposition 2.1, i.e., with the Case-4 score $\partial\varphi_{\mathcal{X}}/\partial\theta$.

3. Forecasting aggregation of moving averages

This section begins by summarizing relevant findings from de Truchis et al. (2026a), DFT henceforth, concerning the description of stable random vectors on the unit cylinder.[13] Let the vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ be an α-stable random vector, Γ a finite spectral measure on the Euclidean unit sphere S_d and $\boldsymbol{\mu}^0$ a non-random vector in $\mathbb{R}^d$, such that,

$$\mathbb{E}\Big(e^{i\langle \boldsymbol{u},\boldsymbol{X}\rangle}\Big) = \exp\left\{ -\int_{S_d} |\langle \boldsymbol{u},\boldsymbol{s}\rangle|^\alpha \Big(1 - i\,\mathrm{sign}(\langle \boldsymbol{u},\boldsymbol{s}\rangle) w(\alpha, \langle \boldsymbol{u},\boldsymbol{s}\rangle)\Big)\Gamma(d\boldsymbol{s}) + i\,\langle \boldsymbol{u},\boldsymbol{\mu}^0\rangle \right\}, \quad \forall \boldsymbol{u} \in \mathbb{R}^d, \tag{3.1}$$

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 $s \in \mathbb{R}$. Drawing on DFT, we explore alternative representations of $\boldsymbol{X}$ where the integration is performed over a unit cylinder $C^{\|\cdot\|}_d := \{\boldsymbol{s} \in \mathbb{R}^d : \ \|\boldsymbol{s}\| = 1\}$, defined by a semi-norm $\|\cdot\|$ on $\mathbb{R}^d$, in presence of stable aggregates. The reason why we are interested in alternative representations is that, in the presence of the Euclidean norm, the spectral measure encodes information in all directions of $\mathbb{R}^d$ and does not allow us to predict future elements of the vector $\boldsymbol{X}$ while ensuring that these future elements are not themselves carriers of information for prediction. By contrast, the semi-norm $\|\cdot\|$ is flexible enough to force some directions $\mathbb{R}^d$ to vanish.

We will say that $\boldsymbol{X}$ is representable on $C^{\|\cdot\|}_d$ if $\boldsymbol{X}$ can be written as in (3.1) with $(S_d, \Gamma, \boldsymbol{\mu}^0)$ replaced by $(C^{\|\cdot\|}_d, \Gamma^{\|\cdot\|}, \boldsymbol{\mu}^0_{\|\cdot\|})$. As demonstrated in DFT for the single-component model, $\boldsymbol{X}$ is representable on $C^{\|\cdot\|}_d \iff \Gamma(K^{\|\cdot\|}) = 0$ when $\alpha \neq 1$ or if $\boldsymbol{X}$ is $\mathcal{S}1\mathcal{S}$. Moreover, $\Gamma^{\|\cdot\|}(d\boldsymbol{s}) = \|\boldsymbol{s}\|_e^{-\alpha}\,\Gamma \circ T^{-1}_{\|\cdot\|}(d\boldsymbol{s})$ with $T_{\|\cdot\|} : S_d \setminus K^{\|\cdot\|} \longrightarrow C^{\|\cdot\|}_d$ defined by $T_{\|\cdot\|}(\boldsymbol{s}) = \boldsymbol{s}/\|\boldsymbol{s}\|$. Importantly, this new representation inherits from the traditional representation the

[13] We exclude the Gaussian case from further discussion as anticipative dynamics are not identifiable when $\alpha = 2$.

following asymptotic conditional tail property: for any Borel sets $A, B \subset C_d^{\|\cdot\|}$ with $\Gamma^{\|\cdot\|}\Big(\partial(A\cap B)\Big) = \Gamma^{\|\cdot\|}(\partial B) = 0$, and $\Gamma^{\|\cdot\|}(B) > 0$,

$$\mathbb{P}_x^{\|\cdot\|}(\boldsymbol{X}, A|B) \underset{x\to+\infty}{\longrightarrow} \frac{\Gamma^{\|\cdot\|}(A\cap B)}{\Gamma^{\|\cdot\|}(B)}, \tag{3.2}$$

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

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

To build a forecasting strategy upon these theoretical results, DFT considers vectors of the form $\boldsymbol{X_t} = (X_{t-m}, \dots, X_t, X_{t+1}, \dots, X_{t+h})$, $m \geq 0$, $h \geq 1$, derived from a stable moving average process and choose, without loss of generality, semi-norms satisfying

$$\|(x_{-m}, \dots, x_0, x_1, \dots, x_h)\| = 0 \iff x_{-m} = \ldots = x_0 = 0, \tag{3.3}$$

for any $(x_{-m}, \dots, x_h) \in \mathbb{R}^{m+h+1}$. They show that for $\alpha \neq 1$ 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. More precisely, they find that sequences of consecutive zero values in must either be of finite length or extend infinitely to the left :

$$\forall k \in \mathbb{Z}, \quad \Big[(d_{k+m}, \dots, d_k) = \boldsymbol{0} \quad \Longrightarrow \quad \forall \ell \leq k-1, \quad d_\ell = 0\Big]. \tag{3.4}$$

This result surprisingly establishes that the anticipativeness of a stable moving average is a necessary condition (and sufficient for $\alpha \neq 1$ and $(\alpha, \beta) = (1, 0)$) to make use of (3.2) in order to feasibly predict $\boldsymbol{X_t}$. The more non-anticipative a moving average is (i.e., the larger the gaps of zeros in its forward-looking coefficients), the larger m must be to achieve representability of $(X_{t-m}, \dots, X_t, X_{t+1}, \dots, X_{t+h})$ on the appropriate unit cylinder.

3.1. Extending the representation to stable aggregates

To extend these results to stable aggregates, we first provide the spectral representation of paths of the aggregated process $\mathcal{X}_t$ on the Euclidean unit sphere.

Lemma 3.1. *Let $\mathcal{X}_t$ be an α-stable aggregate with latent moving averages $(X_{1,t}), \dots, (X_{J,t})$ as in Definition 2.1, but now allowing $\beta_j \in [-1, 1]$ to vary across components, and $\boldsymbol{X_t} = (\mathcal{X}_{t-m}, \dots, \mathcal{X}_t, \mathcal{X}_{t+1}, \dots, \mathcal{X}_{t+h})$ 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

$$\Gamma = \sigma^\alpha \sum_{j=1}^{J} \sum_{\vartheta\in S_1} \sum_{k\in\mathbb{Z}} w_{j,\vartheta} \pi_j^\alpha \|\boldsymbol{d}_{j,k}\|_e^\alpha d_{\left\{\frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|_e}\right\}}, \tag{3.5}$$

$$\boldsymbol{\mu}^0 = \begin{cases} \boldsymbol{0}, & \text{if } \alpha \neq 1 \\ -\frac{2\sigma}{\pi} \sum_{j=1}^J \sum_{k\in\mathbb{Z}} \pi_j \beta_j \boldsymbol{d}_{j,k} \ln \|\sigma \pi_j \boldsymbol{d}_{j,k}\|_e, & \text{if } \alpha = 1 \end{cases}$$

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 $k \in \mathbb{Z}$, $j = 1, \dots, J$, δ is the Dirac mass, $\vartheta \in S_1$ with $S_1 = \{-1, +1\}$, and if $\boldsymbol{d}_{j,k} = \boldsymbol{0}$, the term vanishes by convention from the sums.

Notice that $\Gamma = \sigma^\alpha \sum_{j=1}^J \pi_j^\alpha \Gamma_j$, where Γ_j denotes the spectral measure of the path $\boldsymbol{X}_{j,t}$ from the moving average $(X_{j,t})$, $j = 1, \dots, J$. 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 $\sigma^\alpha \sum_{j=1}^J \pi_j^\alpha \Big(\Gamma_j(A) - \Gamma_j(-A)\Big) = 0$ 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 3.2. *Let $\mathcal{X}_t$ be an α-stable aggregate with latent moving averages $(X_{1,t}),\ldots,(X_{J,t})$ as in Definition 2.1, where each component j has asymmetry parameter $\beta_j \in [-1,1]$. Let $m \geq 0$, $h \geq 1$ and $\|\cdot\|$ be a semi-norm on $\mathbb{R}^{m+h+1}$ satisfying (3.3). When either $\alpha \neq 1$ or $\boldsymbol{X}_t \sim \mathcal{S}1\mathcal{S}$, the vector $\boldsymbol{X}_t$ is representable on $C^{\|\cdot\|}_{m+h+1}$ if and only if condition (3.4) holds with m for all coefficient sequences $(d_{j,k})_k$, $j=1,\ldots,J$. For $\alpha=1$ and $\boldsymbol{X}_t$ asymmetric, the vector $\boldsymbol{X}_t$ is representable on $C^{\|\cdot\|}_{m+h+1}$ if and only if (3.4) holds and*

$$\sum_{k\in\mathbb{Z}} \|\boldsymbol{d}_{j,k}\|_e \left| \ln\left(\|\boldsymbol{d}_{j,k}\|/\|\boldsymbol{d}_{j,k}\|_e\right)\right| < +\infty, \quad \forall j \in \{1,\ldots,J\} \tag{3.6}$$

hold with m and h for all sequences $(d_{j,k})_k$, $j=1,\ldots,J$.

The next proposition extends to stable aggregated processes the notion of past-representability introduced in DFT and helps to understand to what extent anticipativeness is crucial in this more general framework.

Proposition 3.1. *Let $\mathcal{X}_t$ be an α-stable aggregate with latent moving averages $(X_{1,t}),\ldots,(X_{J,t})$ as in Definition 2.1, where $\mathcal{X}_t = \sigma\sum_{j=1}^J \pi_j X_{j,t}$ with scale parameter $\sigma > 0$.*

(ι) Define for $j=1,\ldots,J$ the sets $\mathcal{M}_j = \{m \geq 1 : \ \exists k \in \mathbb{Z},\ \ d_{j,k+m} = \ldots = d_{j,k+1} = 0,\ \ d_{j,k} \neq 0\}$, and

$$m_{0,j} = \begin{cases} \sup \mathcal{M}_j, & \text{if} \quad \mathcal{M}_j \neq \emptyset, \\ 0, & \text{if} \quad \mathcal{M}_j = \emptyset. \end{cases} \tag{3.7}$$

(a) For $\alpha \neq 1$, the aggregated process $\mathcal{X}_t$ is past-representable if and only if $(X_{j,t})$ is past-representable for all $j=1,\ldots,J$, i.e.,

$$\sup_{j=1,\ldots,J} m_{0,j} < +\infty. \tag{3.8}$$

Moreover, letting $m \geq 0$, $h \geq 1$, $\mathcal{X}_t$ is (m,h)-past-representable if and only if (3.8) holds and $m \geq \max_{j=1,\ldots,J} m_{0,j}$.

(b) For $\alpha = 1$, the process $\mathcal{X}_t$ is past-representable if and only if (3.8) holds and there exists a pair (m,h), $m \geq \max_{j=1,\ldots,J} m_{0,j}$, $h \geq 1$ such that either

$$\boldsymbol{X}_t \ \text{ is } \ \mathcal{S}1\mathcal{S}, \qquad \text{or,} \qquad \boldsymbol{X}_t \text{ asymmetric and (3.6) holds for all sequences } (d_{j,k})_k.$$

If such a pair exists, then the process $\mathcal{X}_t$ is (m,h)-past-representable.

($\iota\iota$) Let $\|\cdot\|$ be a semi-norm satisfying (3.3) and assume that $\mathcal{X}_t$ is (m,h)-past-representable for some $m \geq 0$, $h \geq 1$. The spectral representation $(\Gamma^{\|\cdot\|}, \boldsymbol{\mu}^{\|\cdot\|})$ of the vector $\boldsymbol{X_t} = (\mathcal{X}_{t-m},\ldots,\mathcal{X}_t,\mathcal{X}_{t+1},\ldots,\mathcal{X}_{t+h})$ on $C^{\|\cdot\|}_{m+h+1}$ is given by:

$$\Gamma^{\|\cdot\|} = \sigma^\alpha \sum_{j=1}^J \sum_{\vartheta\in S_1} \sum_{k\in\mathbb{Z}} w_{j,\vartheta} \pi_j^\alpha \|\boldsymbol{d}_{j,k}\|^\alpha \delta_{\left\{\frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|}\right\}}, \tag{3.9}$$

$$\boldsymbol{\mu}^{\|\cdot\|} = \begin{cases} \boldsymbol{0}, & \text{if } \alpha \neq 1 \\ -\frac{2\sigma}{\pi}\sum_{j=1}^J \sum_{k\in\mathbb{Z}} \pi_j\beta_j \boldsymbol{d}_{j,k} \ln\|\sigma\pi_j \boldsymbol{d}_{j,k}\|, & \text{if } \alpha = 1 \end{cases} \tag{3.10}$$

where $\boldsymbol{d}_{j,k} = (d_{j,k+m},\ldots,d_{j,k},d_{j,k-1},\ldots,d_{j,k-h})$, $w_{j,\vartheta} = (1+\vartheta\beta_j)/2$, for any $k \in \mathbb{Z}$, $j=1,\ldots,J$, δ is the Dirac mass, $\vartheta \in S_1$ with $S_1 = \{-1,+1\}$, and if $\boldsymbol{d}_{j,k} = \boldsymbol{0}$, the term vanishes by convention from the sums.

The necessary condition (3.8) extends what was noticed in the Proposition 3 of DFT, namely, that anticipativeness is a minimal requirement for past-representability. Importantly, notice that a single non-anticipative latent moving average is enough to render the aggregated process not past-representable, regardless of the other latent components. Also, 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, $\beta_1 = \ldots = \beta_J = 0$, then the paths $\boldsymbol{X}_t$ are $\mathcal{S}1\mathcal{S}$ 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 3.1. *For any $j = 1, \dots, J$, let $(X_{j,t})$ be the ARMA strictly stationary solution of $\Psi_j(F)\Phi_j(B)X_{j,t} = \Theta_j(F)H_j(B)\varepsilon_{j,t}$, with mutually independent sequences $\varepsilon_{j,t} \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, \beta_j, 1, 0)$. Define $\mathcal{X}_t = \sigma \sum_{j=1}^{J} \pi_j X_{j,t}$ for any positive weights π_j summing to 1 and $\sigma > 0$. Then, for any $\alpha \in (0, 2)$, $(\beta_1, \dots, \beta_J) \in [-1, 1]^J$, the following statements are equivalent:*

(ι) $(\mathcal{X}_t)$ *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 (3.7). Moreover, letting $m \geq 0$, $h \geq 1$, the aggregated process $(\mathcal{X}_t)$ is (m, h)-past-representable if and only if for any $j = 1, \dots, J$, $m_{0,j} < +\infty$, and $m \geq \max_j m_{0,j}$.

3.2. Tail conditional distribution of stable aggregates

Now, we derive the tail conditional distribution of linear stable aggregates. The case of a general past-representable stable aggregate is considered. We also pay a particular attention to the anticipative $\mathcal{G}\alpha\mathcal{S}$ AR(1) because to the best of our knowledge, no deconvolution estimation techniques exists for stable aggregates as defined in 2.1, except for the anticipative $\mathcal{G}\alpha\mathcal{S}$ AR(1) discussed in Section 2. To be relevant for the prediction framework, the Borel set B appearing in Equation 3.2 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 $\mathcal{X}_{t+1}, \dots, \mathcal{X}_{t+h}$. For $\|\cdot\|$ a semi-norm on $\mathbb{R}^{m+h+1}$ satisfying (3.3), denote $S_{m+1}^{\|\cdot\|} = \{(s_{-m}, \dots, s_0) \in \mathbb{R}^{m+1} : \quad \|(s_{-m}, \dots, s_0, 0, \dots, 0)\| = 1\}$.[14] 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

$$f : \begin{array}{ccc} \mathbb{R}^{m+h+1} & \longrightarrow & \mathbb{R}^{m+1} \\ (x_{-m}, \dots, x_0, x_1, \dots, x_h) & \longmapsto & (x_{-m}, \dots, x_0) \end{array}. \tag{3.11}$$

Let $\mathcal{X}_t$ an α-stable aggregate as in Definition 2.1. Assume $\mathcal{X}_t$ is (m, h)-past-representable, for some $m \geq 0$, $h \geq 1$. Also, we know by Proposition 3.1 ($\iota\iota$), that $\Gamma^{\|\cdot\|}$ is of the form

$$\Gamma^{\|\cdot\|} = \sigma^\alpha \sum_{j=1}^{J} \sum_{\vartheta \in S_1} \sum_{k \in \mathbb{Z}} w_{j,\vartheta} \pi_j^\alpha \|\boldsymbol{d}_{j,k}\|^\alpha \delta_{\left\{ \frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|} \right\}}. \tag{3.12}$$

Proposition 3.2. *Let $\mathcal{X}_t$ be an α-stable aggregate as in Definition 2.1. Assume $\mathcal{X}_t$ is (m, h)-past-representable, for some $m \geq 0$, $h \geq 1$. Also, we know by Proposition 3.1 ($\iota\iota$), that $\Gamma^{\|\cdot\|}$ is of the form*

$$\Gamma^{\|\cdot\|} = \sigma^\alpha \sum_{j=1}^{J} \sum_{\vartheta \in S_1} \sum_{k \in \mathbb{Z}} w_{j,\vartheta} \pi_j^\alpha \|\boldsymbol{d}_{j,k}\|^\alpha \delta_{\left\{ \frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|} \right\}}.$$

[14] 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

$$\mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t}, A\Big|B(V)\Big) \underset{x\to+\infty}{\longrightarrow} \frac{\Gamma^{\|\cdot\|}\left(\left\{\dfrac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|}\in A:\quad \dfrac{\vartheta f(\boldsymbol{d}_{j,k})}{\|\boldsymbol{d}_{j,k}\|}\in V\right\}\right)}{\Gamma^{\|\cdot\|}\left(\left\{\dfrac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|}\in C_{m+h+1}^{\|\cdot\|}:\quad \dfrac{\vartheta f(\boldsymbol{d}_{j,k})}{\|\boldsymbol{d}_{j,k}\|}\in V\right\}\right)}, \tag{3.13}$$

for any Borel sets $A\subset C_{m+h+1}^{\|\cdot\|}$, $V\subset S_{m+1}^{\|\cdot\|}$ *such that* $\left\{\dfrac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|}\in C_{m+h+1}^{\|\cdot\|}:\quad \dfrac{\vartheta f(\boldsymbol{d}_{j,k})}{\|\boldsymbol{d}_{j,k}\|}\in V\right\}\neq\emptyset$, $\Gamma^{\|\cdot\|}\Big(\partial(A\cap B(V))\Big)=\Gamma^{\|\cdot\|}(\partial B(V))=0$, *where* $B(V)=V\times\mathbb{R}^h$ *and* f *is as in* (3.11).

Observe that setting $V = S_{m+1}^{\|\cdot\|}$, and A an arbitrarily small closed neighbourhood of all the points $(\vartheta\boldsymbol{d}_{j,k}/\|\boldsymbol{d}_{j,k}\|)_{\vartheta,j,k}$, as in the single-component case we have $\lim_{x\to+\infty}\mathbb{P}\Big(\boldsymbol{X_t}/\|\boldsymbol{X_t}\|\in A\Big|\|\boldsymbol{X_t}\|>x\Big)=1$. 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 $(d_{j,k+m},\ldots,d_{j,k},d_{j,k-1},\ldots,d_{j,k-h})$ 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.

Interestingly, as in DFT in the non-aggregated case, the observed path $(\mathcal{X}_{t-m},\ldots,\mathcal{X}_{t-1},\mathcal{X}_t)/\|\boldsymbol{X_t}\|$ will *a fortiori* be of the same shape as some $\vartheta(d_{j,k+m},\ldots,d_{j,k+1},d_{j,k})/\|\boldsymbol{d}_{j,k}\|$ 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. In practice, we anticipate that matching an observed path to a particular pattern j among the collection of J patterns will be challenging, even for a small number of latent components.

3.3. Example: Aggregation of Anticipative AR(1) Processes

We now consider the aggregation of stable anticipative AR(1) processes discussed in Section 2. We assume without loss of generality that the ψ_j's are distinct. For each anticipative AR(1) with parameter ψ_j, the moving average coefficients are of the form $(\psi_j^k\mathbb{1}_{\{k\geq 0\}})_k$, and thus, $m_{0,j}=0$ for all j, where the $m_{0,j}$'s are given in (3.7). By Corollary (3.1), we know for any $m\geq 0$, $h\geq 1$, the aggregated process $\mathcal{X}_t$ 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 3.3. *Let* $\mathcal{X}_t$ *be an aggregation of* α*-stable anticipative AR(1) processes as in Definition 2.1 with* $d_{j,k}=\psi_j^k$ *and general scale parameter* $\sigma>0$.

Letting $\boldsymbol{X_t}=(\mathcal{X}_{t-m},\ldots,\mathcal{X}_t,\mathcal{X}_{t+1},\ldots,\mathcal{X}_{t+h})$ *for* $m\geq 0$, $h\geq 1$, *its spectral measure on* $C_{m+h+1}^{\|\cdot\|}$ *for a seminorm satisfying* (3.3) *is given by*

$$\Gamma^{\|\cdot\|}=\sum_{\vartheta\in S_1}\left[w_\vartheta\delta_{\{(\vartheta,0,\ldots,0)\}}+\sum_{j=1}^J\sigma^\alpha\pi_j^\alpha\left(w_{j,\vartheta}\sum_{k=-m+1}^{h-1}\|\boldsymbol{d}_{j,k}\|^\alpha\delta_{\left\{\frac{\vartheta\boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|}\right\}}+\frac{\bar{w}_{j,\vartheta}}{1-|\psi_j|^\alpha}\|\boldsymbol{d}_{j,h}\|^\alpha\delta_{\left\{\frac{\vartheta\,\boldsymbol{d}_{j,h}}{\|\boldsymbol{d}_{j,h}\|}\right\}}\right)\right], \tag{3.14}$$

where for all $\vartheta \in S_1$, $j \in \{1,\dots,J\}$ and $-m+1 \le k \le h$,

$$
\begin{aligned}
\boldsymbol{d}_{j,k} &= (\psi_j^{k+m}\mathbb{1}_{\{k\ge -m\}},\dots,\psi_j^{k}\mathbb{1}_{\{k\ge 0\}},\psi_j^{k-1}\mathbb{1}_{\{k\ge 1\}},\dots,\psi_j^{k-h}\mathbb{1}_{\{k\ge h\}}),\\
w_{j,\vartheta} &= (1+\vartheta\beta_j)/2,\\
w_\vartheta &= \sum_{j=1}^{J}\sigma^\alpha\pi_j^\alpha w_{j,\vartheta},\\
\bar{w}_{j,\vartheta} &= (1+\vartheta\bar{\beta}_j)/2,\\
\bar{\beta}_j &= \beta_j\frac{1-\psi_j^{<\alpha>}}{1-|\psi_j|^\alpha},
\end{aligned}
$$

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 ψ_j's are positive. Let us first introduce useful neighbourhoods of the distinct charged points of $\Gamma^{\|\cdot\|}$. Denote $\boldsymbol{d}_{0,-m} = (\overbrace{1,0,\dots,0}^{m+h+1})$ 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 $\mathcal{I} := S_1 \times \Big(\{1,\dots,J\}\times\{-m,h\}\cup\{(0,-m)\}\Big)$. With f as in (3.11), 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

$$
\forall(\vartheta',j',k')\in\mathcal{I},\qquad \frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|}\in V_0 \implies \frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})}{\|\boldsymbol{d}_{j_0,k_0}\|}, \tag{3.15}
$$

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,

$$
\forall(\vartheta',j',k')\in\mathcal{I},\qquad \frac{\vartheta'\boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|}\in A_{\vartheta,j,k} \implies (\vartheta',j',k')=(\vartheta,j,k). \tag{3.16}
$$

Proposition 3.3. *Let $\mathcal{X}_t$ be an aggregation of α-stable anticipative AR(1) processes as in Definition 2.1 with $d_{j,k} = \psi_j^k \in (0,1)$ 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 3.3, for any $m\ge 0$, $h\ge 1$. 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 (3.15) 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 (3.16), the following hold.*

(ι) ***Case** $m\ge 1$.*

(a) If $0\le k_0\le h$:

$$
\mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t}, A_{\vartheta,j,k}\Big|B(V_0)\Big) \underset{x\to\infty}{\longrightarrow} \begin{cases} |\psi_{j_0}|^{\alpha k}(1-|\psi_{j_0}|^\alpha)\delta_{\vartheta_0}(\vartheta)\delta_{j_0}(j), & 0\le k\le h-1,\\[2ex] |\psi_{j_0}|^{\alpha h}\delta_{\vartheta_0}(\vartheta)\delta_{j_0}(j), & k=h. \end{cases}
$$

(b) If $-m\le k_0\le -1$:

$$
\mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t}, A_{\vartheta,j,k}\Big|B(V_0)\Big) \underset{x\to\infty}{\longrightarrow} \delta_{\vartheta_0}(\vartheta)\delta_{j_0}(j)\delta_{k_0}(k).
$$

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

$$
\mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t}, A_{\vartheta,j,k}\Big|B(V_0)\Big) \underset{x\to\infty}{\longrightarrow} \begin{cases} \dfrac{\sum_{i=1}^J \pi_i^\alpha w_{i,\vartheta_0}}{\sum_{i=1}^J p_{i,\vartheta_0}}\delta_{\{\vartheta_0\}}(\vartheta), & k=0\\[3ex] \dfrac{p_{j,\vartheta_0}}{\sum_{i=1}^J p_{i,\vartheta_0}}|\psi_j|^{\alpha k}(1-|\psi_j|^\alpha)\delta_{\{\vartheta_0\}}(\vartheta), & 1\le k\le h-1,\\[3ex] \dfrac{p_{j,\vartheta_0}}{\sum_{i=1}^J p_{i,\vartheta_0}}|\psi_j|^{\alpha h}\delta_{\{\vartheta_0\}}(\vartheta), & k=h, \end{cases}
$$

with $p_{j,\vartheta_0} = \pi_j^\alpha w_{j,\vartheta_0}/(1-|\psi_j|^\alpha)$.

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 $k_0 \in \{0,\ldots,h\}$ or $k_0 \in \{-m,\ldots,-1\}$. 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),

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

We can notice that all the above sequences are pieces of explosive exponentials, terminated at some coordinate. For $k \in \{0,\ldots,h\}$, the first zero component, i.e. the crash of the bubble, is situated at or after the $(m+2)^{\text{th}}$ component, whereas for $k \in \{-m,\ldots,-1\}$, it is situated at or before the $(m+1)^{\text{th}}$. Using the homogeneity of the semi-norm, we have on the other hand that

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

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 $k_0 \in \{-m,\ldots,-1\}$ 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 $k_0 \in \{0,\ldots,h\}$ 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 ψ_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 $|\psi_j|^{\alpha k}(1-|\psi_j|^\alpha)$ corresponds to the probability that the bubble will peak in exactly k periods ($0 \leq k < h$), and $|\psi_j|^{\alpha h}$ corresponds to the probability that the bubble will last at least h more periods.

The previous statement confirms the interpretation of the conditional moments proposed in Fries (2022) 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 3.3 furthermore shows that asymptotically, as few as two observations are sufficient to identify the growth rate ψ_j^{-1} of an ongoing extreme episode,[15] 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, as long as the latent components parameters are identified.

4. Simulation Results

This section documents the finite-sample performance of the minimum distance estimator developed in Section 2. Section 4.1 presents Monte Carlo evidence on estimation accuracy. Section 4.2 implements a subsampling methodology to empirically verify the asymptotic normality predicted by Proposition 2.1. Complete results, including detailed tables, graphical diagnostics, and sensitivity analyses, are provided in the Online Supplement, Sections S.1 and S.2.

4.1. Estimation accuracy

The observed process is generated by the aggregation of two independent α-stable AR(1) processes:

$$\mathcal{X}_t = \pi_1 X_{1,t} + \pi_2 X_{2,t}, \tag{4.1}$$

$$X_{j,t} = \psi_j X_{j,t+1} + \varepsilon_{j,t}, \quad \varepsilon_{j,t} \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, \beta, 1, 0), \tag{4.2}$$

where $j \in \{1, 2\}$ and $\psi_j \in (0, 1)$. We fix $\sigma = 1.6$, $\pi_1 = 7/16$, $\pi_2 = 9/16$, yielding combined scale parameters $\varsigma_1 = 0.7$ and $\varsigma_2 = 0.9$. Three distributional settings are considered: (i) Cauchy ($\mathcal{S}1\mathcal{S}$); (ii) symmetric α-stable ($\mathcal{S}\alpha\mathcal{S}$, $\alpha = 1.5$); and (iii) general α-stable ($\mathcal{G}\alpha\mathcal{S}$, $\alpha = 1.5$, $\beta = 0.3$). For each case, we perform 1,000 replications with sample sizes $T \in \{250, 500, 1{,}000\}$.

Results (Table S.1 in the Online Supplement) confirm the good finite-sample behavior of the estimator. The dominant autoregressive coefficient ψ_1 is precisely estimated across all settings, with mean relative error (MRE) below 13% even in the $\mathcal{G}\alpha\mathcal{S}$ case for $T = 250$. The tail index α is recovered with high accuracy (MRE around 9% at $T = 250$, declining to 4-5% at $T = 1{,}000$). The smaller coefficient ψ_2 and the combined scale parameters exhibit higher relative errors, reflecting the difficulty in disentangling individual component contributions from the aggregate signal. The asymmetry parameter β is the most challenging to estimate (MRE of 69% at $T = 250$), although its identification is not required for the autoregressive and scale parameters. Section 2.3 extends this design to mixed causal-noncausal MAR(1,1) aggregates, adding one causal root per component; Table S.2 of the Online Supplement reports the corresponding accuracy results, which are qualitatively similar but uniformly noisier, the added causal root proving the hardest parameter to pin down throughout.

4.2. Subsampling-based verification of asymptotic normality

To empirically verify Proposition 2.1, we implement a subsampling methodology following Politis and Romano (1994) and Politis, Romano and Wolf (1999). Given a full sample of size n, we construct non-overlapping subsamples of size $b = \lfloor n^{2/3} \rfloor$:

$$\mathcal{X}_b^{(i)} = \{\mathcal{X}_{(i-1)b+1}, \dots, \mathcal{X}_{ib}\}, \quad i = 1, \dots, N_b = \lfloor n/b \rfloor. \tag{4.3}$$

[15] 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.

[16] Since the individual scale parameters ς_1 and ς_2 exhibit slow finite-sample convergence (as already noted in Section 4.1 above, and documented in Table S.1 of the Online Supplement), we apply a post-estimation reparameterization:

$$\vartheta = (\psi_1, \psi_2, \sigma, \pi_1, \alpha), \quad \text{where} \quad \sigma = \varsigma_1 + \varsigma_2, \quad \pi_1 = \frac{\varsigma_1}{\sigma}. \tag{4.4}$$

We conduct $M = 200$ Monte Carlo replications for each sample size $n \in \{500, 1{,}000, 10{,}000, 100{,}000, 500{,}000\}$.[17]

Coverage at $n = 500$ and $n = 1{,}000$ is far below nominal for all five reparameterized coordinates, reflecting the small number of non-overlapping blocks available at these sample sizes. By $n = 10{,}000$–$100{,}000$ coverage approaches nominal for all of them, α fastest and most reliably (Online Supplement, Table S.4); this still holds at $n = 500{,}000$, where the standard deviation of every parameter also shrinks almost exactly at the $1/\sqrt{n}$ rate.

Two other things stand out. Shapiro-Wilk and Jarque-Bera rejection rates jump at $n = 100{,}000$ for the three parameters confined to a bounded region of the parameter space, while σ and α show nothing of the sort; this traces to a small share of subsample blocks landing on the boundary of the compact parameter space, and has no bearing on the asymptotic normality established for the full-sample estimator by Proposition 2.1 (Online Supplement, Section S.2.2 and Table S.8). At $n = 500{,}000$ this recedes rather than getting worse, e.g. Shapiro-Wilk rejection for ψ_1 falls from 0.595 to 0.260 (Online Supplement, Table S.9), so it looks like a transient artifact of the block boundary rather than a growing problem with the estimator. Normal-approximation intervals also beat the quantile-based ones at small n, consistent with how few non-overlapping blocks are available for the quantile calculation there.

5. Illustration on financial markets

We apply our framework to the CBOE Crude Oil ETF Volatility Index (OVX), a forward-looking measure of expected crude oil price volatility often referred to as a *fear index.* The literature on heterogeneous agent models (e.g. Agliari et al. , 2018) suggests that fundamentalist/chartist dynamics may generate distinct volatility components, particularly during periods of market stress. We collect the OVX from the FRED website at weekly frequency over 23/05/2015-27/04/2025 ($T = 518$), linearly detrended following Hecq and Voisin (2021).

We estimate a general α-stable specification ($\mathcal{G}\alpha\mathcal{S}$) with two latent AR(1) components.[18] The estimates reveal distinct dynamics: the first component ($\hat{\psi}_1 = 0.80$, $\hat{\pi}_1 = 0.28$) captures abrupt, short-lived volatility bursts, while the second ($\hat{\psi}_2 = 0.85$, $\hat{\pi}_2 = 0.72$) drives more persistent explosive episodes. The two autoregressive coefficients are nearly equal; it is the marked difference in weights, not in persistence, that generates this contrast in dynamics, as illustrated by the deconvolution via de Truchis et al. (2026b), which shows that extreme market stress periods, most visibly the 2020 disruption, feature a superposition of both dynamics with heterogeneous persistence properties. The tail index $\hat{\alpha} = 1.47$ confirms heavy-tailed behavior well beyond Gaussian accommodation.

To demonstrate forecasting potential, we conduct an in-sample prediction exercise for the 2020 oil market disruption. Setting January 2020 as the cut-off, we apply Proposition 3.3 with pattern matching length $m = 20$. For each component, we identify the matched position k_0 and compute cumulative crash probabilities $1 - |\psi_{j_0}|^{\alpha h}$ and survival probabilities $|\psi_{j_0}|^{\alpha h}$ for all future horizons h.[19]

[16] Following Politis, Romano and Wolf (1999), the choice $b = \lfloor n^{2/3} \rfloor$ ensures $b \to \infty$, $b/n \to 0$, and satisfies the higher-order requirements for the subsampling approximation to be valid (Online Supplement, Section S.2.1); no alternative subsample-size rule is reported.

[17] Table S.3 summarizes the Monte Carlo design. Table S.4 summarizes results across all five sample sizes; complete diagnostics for each sample size are reported in Tables S.5–S.9. Visual diagnostics are displayed in various figures throughout the online supplement.

[18] Estimation results for the $\mathcal{S}\alpha\mathcal{S}$ and $\mathcal{S}1\mathcal{S}$ specifications, along with the complete deconvolution analysis, are reported in Section S.3 of the Online Supplement.

[19] Detailed per-component crash probability profiles, forecast trajectories across risk thresholds (90%, 95%, 99%), and the complete forecasting algorithm are provided in Section S.3 of the Online Supplement.

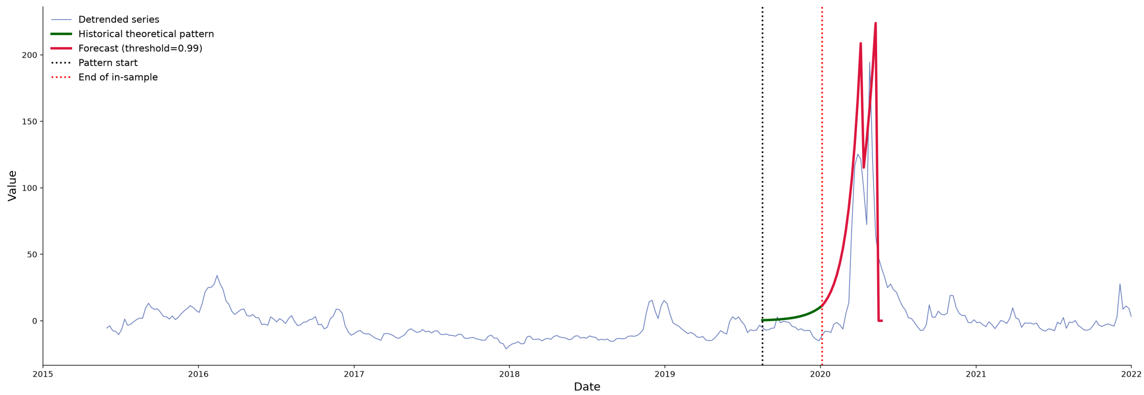

Figure 2: Combined in-sample forecast at the 99% risk threshold for the 2020 OVX bubble. The blue line shows the detrended series, the green segment indicates the matched historical theoretical pattern starting at the black dotted vertical line, and the red curve displays the out-of-sample forecast beyond the January 2020 cut-off (red dotted vertical line).

Figure 2 presents the combined forecast at the 99% threshold. The aggregate trajectory closely tracks the realized path during the March 2020 spike, reaching approximately 220 before the predicted crash remarkably close to the observed peak of around 195.

6. Conclusion

This paper addresses a fundamental limitation in the empirical modeling of rational asset bubbles in financial markets by introducing a novel framework based on α-stable moving average aggregates. Traditional approaches to bubble modeling based on anticipative heavy-tailed processes impose uniform bubble patterns across different episodes, contradicting the observed heterogeneity in market dynamics. Our contribution is both theoretical and methodological. Theoretically, we develop a flexible model built on α-stable moving average aggregates that accommodates diverse bubble growth patterns and crash dynamics. We establish that this model admits a semi-norm representation on a unit cylinder, similar to non-aggregated moving averages, thereby enabling the forecasting of bubble episodes with heterogeneous growth trajectories. We extend the spectral representation of stable processes to aggregated components and derive conditions under which the tail conditional distribution can be used for prediction, showing that anticipativeness remains a necessary condition for past-representability even in the aggregated case. Methodologically, we develop a minimum distance estimation procedure based on the joint characteristic function that effectively identifies the parameters of stable aggregates, and we establish its consistency and asymptotic normality. Unlike existing approaches limited to the Cauchy case with continuous support distributions, our framework extends to the general α-stable family with discrete support, making it more suitable for empirical applications. Monte Carlo simulations confirm robust finite-sample performance, and a subsampling procedure supports the asymptotic normality of the estimator, with confidence interval coverage converging to near-nominal levels for every parameter of the reparameterized space, including the mixing proportion. An empirical illustration using the CBOE OVX index reveals the presence of multiple anticipative components with distinct persistence properties and asymmetric weights. The deconvolution analysis shows that the 2020 oil market disruption actually comprises multiple superimposed processes with heterogeneous growth rates and crash probabilities, and our forecasting framework successfully anticipates both the timing and magnitude of the March 2020 volatility spike.

References

Agliari, A., Naimzada, A., Pecora, N. (2018). Boom-bust dynamics in a stock market participation model with heterogeneous traders. *Journal of Economic Dynamics and Control*, 91, 458-468.

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

Blanchard, O.J., and Watson, M.W. 1982. Bubbles, Rational Expectations and Financial Markets. *NBER Working Papers*, 0945.

Breidt, F. J., Davis, R. A., Lii, K.-S., and Rosenblatt, M. 1991. Maximum likelihood estimation for noncausal autoregressive processes. *Journal of Multivariate Analysis*. 36, 175–198.

Brockwell, P.J., and Davis, R.A. 1991. *Time series: theory and methods*. Springer science & business media.

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

Cubadda, G., Hecq, A., Telg, S. (2019). Detecting Co-Movements in Non-Causal Time Series. *Oxford Bulletin of Economics and Statistics*, 81(3), 697-715.

de Truchis, G., Fries, S., Thomas, A. (2026). Forecasting Extreme Trajectories Using Seminorm Representations *Working Paper 2025-06*, Chaire Économie du Climat, Paris.

de Truchis, G., Thomas, A., L. Vaudrée (2026). Deconvolution and Filtering of noncausal Alpha-Stable Processes *On going paper*.

de Truchis, G. and Thomas, A. (2026). Laurent series expansion for MA(∞) representation of mixed causal-noncausal autoregressive processes. *Working Paper*.

Doukhan, P. 1994. Mixing: Properties and Examples. *Lecture Notes in Statistics, Vol. 85. New York: Springer-Verlag*.

Erdős, P. and A. H., Stone. 1970. On the sum of two Borel sets. *Proceedings of the American Mathematical Society*, 25, 304-306.

Fries, S. (2022) Conditional Moments of Noncausal Alpha-Stable Processes and the Prediction of Bubble Crash Odds. *Journal of Business & Economic Statistics*, 40:4, 1596-1616

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

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.

Gassiat, E. 1993. Adaptive estimation in noncausal stationary AR processes. *The Annals of Statistics*. 21, 2022–2042.

Giancaterini, F., Hecq, A., Jasiak, J., and Manafi Neyazi, A. (2026). Bubble detection with application to green bubbles: A noncausal approach. *arXiv preprint arXiv:2505.14911*.

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

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

Gourieroux, C., Jasiak, J. (2018). Misspecification of noncausal order in autoregressive processes. *Journal of Econometrics*, 205(1), 226-248.

Gourieroux, C. and Jasiak, J. 2023. Generalized Covariance Estimator. *Journal of Business & Economic Statistics*. 41(4), 1315–1327.

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

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

Gouriéroux, C., Lu, Y., and Robert, C.-Y. 2025. The Causal-Noncausal Tail Processes: An Introduction. *arXiv preprint arXiv:2506.04046*.

Gourieroux, C., and Jasiak, J. (2026). Nonlinear fore(back)casting and innovation filtering for causal–noncausal VAR models. *Journal of Financial Econometrics*, 24(2), nbag005.

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

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.

Hecq, A., Telg, S., and L., Lieb. 2017. Do seasonal adjustments induce noncausal dynamics in inflation rates? *Econometrics*, 5, 48.

Hecq, A., Telg, S., and Lieb, L. 2017. Simulation, estimation and selection of mixed causal-noncausal autoregressive models: The MARX Package. *SSRN*. https://ssrn.com/abstract=3015797.

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

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

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

Knight, J. L., Yu, J. (2002). Empirical characteristic function in time series estimation. *Econometric Theory*, 18(3), 691-721.

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

Lanne, M. and Saikkonen, P. 2013. Noncausal vector autoregression. *Econometric Theory*. 29, 447–481.

Lanne, M., and Luoto, J. (2013). A noncausal autoregressive model with time-varying parameters: An application to U.S. inflation. Discussion Papers of DIW Berlin 1285, DIW Berlin, German Institute for Economic Research.

Rootzén, H. (1978): Extremes of moving averages of stable processes. *The Annals of Probability*, 847-869.

Rosenblatt, M. (2000): *Gaussian and Non-Gaussian Linear Time Series and Random Fields*. Springer Series in Statistics, Springer-Verlag, New York.

Samorodnitsky, G., and M. S., Taqqu. 1994. *Stable non-Gaussian random processes*, Chapman & Hall, London, 516-536

Tirole, J. 1985. Asset bubbles and overlapping generations. *Econometrica*, 1499-1528.

Politis, D.N. and Romano, J.P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics*, 22(4), 2031–2050.

Politis, D.N., Romano, J.P. and Wolf, M. (1999). *Subsampling.* Springer Series in Statistics, Springer-Verlag, New York.

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

Xu, D. and Knight, J. (2010). Continuous empirical characteristic function estimation of mixtures of normal parameters. *Econometric Reviews*, 30(1) 25–50.

A. Proofs

A.1. Proof of Lemma 2.1

We first establish the $C^2(\Theta)$ regularity of (2.14), the MDE objective function. The proof proceeds by analyzing the theoretical characteristic function structure and establishing precise control over its derivatives under Assumptions 1 and 2. For simplicity, the proof is only developed for the MAR(0,1) case although it also holds for the MAR(1,1) case. Case 3 is covered by direct extension of this lemma's conclusion, as noted where the Gaussian block is introduced; Case 4 is treated separately by Lemma 2.4 in Appendix A.2, since its untruncated characteristic function admits no closed form. In accordance with Assumption 1, the parameter vector is $\theta = (\varsigma_1, \dots, \varsigma_J, \psi_1, \dots, \psi_J, \alpha, \beta)$ with $\varsigma_j = \sigma\pi_j$.

A.1.1. $C^2(\Theta)$ regularity and validation of Assumption 3

Recall that for the α-stable MAR(0,1) component, the variable $uX_{j,t} + vX_{j,t+1}$ decomposes into two independent parts: $(\psi_j u + v)X_{j,t+1}$ and $u\varepsilon_{j,t}$. The joint log-characteristic function is the sum of their log-characteristic functions. Since $\alpha \in (1, 2)$ by Assumption 2, we have

$$\tilde{\varphi}_j(u, v; \theta) = -\frac{|\psi_j u + v|^\alpha}{1 - |\psi_j|^\alpha} \mathcal{A}(\psi_j u + v) - |u|^\alpha \mathcal{A}(u), \tag{A.1}$$

where $\mathcal{A}(x) = 1 - i\beta \operatorname{sign}(x) \tan(\pi\alpha/2)$, so that $\tilde{\varphi}_j$ is the log-characteristic function of the pair $(X_{j,t}, X_{j,t+1})$ at unit scale. The scale enters through the arguments in (2.4): since $\tilde{\varphi}_j$ is positively homogeneous of degree α, $\tilde{\varphi}_j(\varsigma_j u, \varsigma_j v) = \varsigma_j^\alpha \, \tilde{\varphi}_j(u, v)$, and therefore

$$\log \varphi_{\mathcal{X}}(u, v; \theta) = \sum_{j=1}^{J} \varsigma_j^\alpha \, \tilde{\varphi}_j(u, v; \theta), \qquad \varsigma_j = \sigma\pi_j, \tag{A.2}$$

in accordance with Assumption 1, it is this scaled form that the argument of Appendix A.1.3 relies on. Let $\mathcal{K} \subset \Theta$ be any compact subset satisfying the uniform bounds: $\inf_{j,\theta\in\mathcal{K}}(1-|\psi_j|) \geq \delta' > 0$, $\inf_{\theta\in\mathcal{K}} \alpha \geq \alpha_0 > 1$, $\sup_{\theta\in\mathcal{K}} \sigma \leq M < \infty$, and $\sup_{\theta\in\mathcal{K}} |\beta| \leq B < \infty$. From these assumptions, we can establish a uniform lower bound for $1 - |\psi_j|^\alpha$. Since $|\psi_j| \leq 1 - \delta'$, it follows that $1 - |\psi_j|^\alpha \geq 1 - (1 - \delta')^\alpha$, which is increasing in α (since $1 - \delta' \in (0, 1)$). Therefore, its minimum value over $\mathcal{K}$ is attained at α_0. We can thus define a single constant $\delta = 1 - (1 - \delta')^{\alpha_0} > 0$, which ensures that for all $\theta \in \mathcal{K}$, we have $1 - |\psi_j|^\alpha \geq \delta$. The envelopes constructed below retain the factor $|\varphi(u, v; \theta)|$, and it is worth recording once what that factor buys. For $(u, v) \neq 0$ write $r = \|(u, v)\|$. If $|u| \geq r/2$ then $\max(|u|, |\psi_1 u + v|) \geq r/2$; and if $|u| < r/2$ then $|v| > \frac{\sqrt{3}}{2} r$, whence $|\psi_1 u + v| \geq |v| - |u| > \frac{1}{3} r$ since $|\psi_1| < 1$. In either case $\max(|u|, |\psi_1 u + v|) \geq r/3$, so that, $1 - |\psi_j|^\alpha$ being at most one,

$$\left|\varphi(u, v; \theta)\right| \;\leq\; \exp\left\{-c \, \|(u, v)\|^\alpha\right\}, \qquad c := \inf_{\theta\in\mathcal{K}} \left(\frac{\underline{\varsigma}}{3}\right)^\alpha > 0, \tag{A.3}$$

uniformly on $\mathcal{K}$, where $\underline{\varsigma} := \inf_{\theta\in\mathcal{K}} \min_j \varsigma_j > 0$. Consequently every envelope below of the form $|\varphi| \times$ (polynomial in $|u|, |v|$ with logarithmic factors) is not merely w-integrable but bounded and rapidly decreasing; in particular the products of first derivatives that arise in $\partial^2_{\theta\theta'}\varphi = \varphi(\partial^2_{\theta\theta'} \log \varphi + \partial_\theta \log \varphi \, \partial_{\theta'} \log \varphi)$ are dominated by the same envelopes as the second derivatives of $\log \varphi$, and we shall use the crude bound $|\varphi| \leq 1$ only where no such gain is needed.

(ι) The derivatives with respect to the scale coordinates ς_k are computed from (A.2), $\log \varphi(u, v; \theta) = \sum_{j=1}^{J} \varsigma_j^\alpha \tilde{\varphi}_j(u, v)$, in which ς_k appears only through the factor ς_k^α. Hence

$$\frac{\partial \varphi}{\partial \varsigma_k}(u, v; \theta) = \varphi(u, v; \theta) \cdot \alpha \, \varsigma_k^{\alpha-1} \, \tilde{\varphi}_k(u, v), \qquad \frac{\partial^2 \varphi}{\partial \varsigma_k^2}(u, v; \theta) = \varphi(u, v; \theta) \cdot \left[\alpha(\alpha - 1)\varsigma_k^{\alpha-2}\tilde{\varphi}_k + \alpha^2 \varsigma_k^{2\alpha-2} \tilde{\varphi}_k^2\right].$$

Substituting $\tilde{\varphi}_k(u,v) = -\big[|\psi_k u+v|^\alpha \mathcal{A}(\psi_k u+v)/(1-|\psi_k|^\alpha)+|u|^\alpha\mathcal{A}(u)\big]$ and using the uniform bounds on the compact set $\mathcal{K}$, namely $|\mathcal{A}(\cdot)| \le 1+B|\tan(\pi\alpha_0/2)| =: M_{\mathcal{A}}$ with $\alpha_0 = \inf_{\theta\in\mathcal{K}}\alpha > 1$, $1-|\psi_k|^\alpha \ge \delta$ and $\varsigma_k \in [\underline{\varsigma},\overline{\varsigma}] \subset (0,\infty)$, we obtain

$$\left|\frac{\partial\varphi}{\partial\varsigma_k}(u,v;\theta)\right| \;\le\; C_\varsigma\,|\varphi(u,v;\theta)|\left[\frac{|\psi_k u+v|^\alpha}{\delta}+|u|^\alpha\right] \;=:\; C_\varsigma\,G_\varsigma(u,v),$$

with $C_\varsigma = \alpha\,\overline{\varsigma}^{\,\alpha-1}M_{\mathcal{A}}$. The bounding function G_ς is of polynomial growth of degree α and is therefore integrable against w for any $\alpha > 1$. Note that no derivative with respect to σ or π_k is required: these are not coordinates of θ under Assumption 1, and differentiating along them would follow the direction of exact degeneracy $\sigma\,\partial_\sigma\varphi = \sum_j \pi_j\,\partial_{\pi_j}\varphi$ discussed there.

($\iota\iota$) The derivative with respect to α is computed from the j-th summand $\varsigma_j^\alpha\tilde{\varphi}_j$ of (A.2), with $\tilde{\varphi}_j$ given by (A.1). Since $\partial_\alpha[\varsigma_j^\alpha\tilde{\varphi}_j] = \varsigma_j^\alpha\ln\varsigma_j\,\tilde{\varphi}_j + \varsigma_j^\alpha\,\partial_\alpha\tilde{\varphi}_j$, we have

$$\begin{aligned}\frac{\partial}{\partial\alpha}\big[\varsigma_j^\alpha\tilde{\varphi}_j\big] &= -\varsigma_j^\alpha\ln\varsigma_j\left[\frac{|\psi_j u+v|^\alpha\mathcal{A}(\psi_j u+v)}{1-|\psi_j|^\alpha}+|u|^\alpha\mathcal{A}(u)\right]\\ &\quad-\varsigma_j^\alpha\left[\frac{|\psi_j u+v|^\alpha\ln|\psi_j u+v|}{1-|\psi_j|^\alpha}\mathcal{A}(\psi_j u+v)+|u|^\alpha\ln|u|\mathcal{A}(u)\right]\\ &\quad-\varsigma_j^\alpha\left[\frac{|\psi_j u+v|^\alpha|\psi_j|^\alpha\ln|\psi_j|}{(1-|\psi_j|^\alpha)^2}\right]\mathcal{A}(\psi_j u+v)\\ &\quad-\varsigma_j^\alpha\left[\frac{|\psi_j u+v|^\alpha}{1-|\psi_j|^\alpha}\frac{\partial\mathcal{A}(\psi_j u+v)}{\partial\alpha}+|u|^\alpha\frac{\partial\mathcal{A}(u)}{\partial\alpha}\right],\end{aligned}$$

which is the j-th summand of (A.12) in Appendix A.1.3. Using the uniform bounds on the compact set $\mathcal{K}$, specifically $1-|\psi_j|^\alpha \ge \delta$ and the fact that the derivative $\partial_\alpha\mathcal{A}(x) = -i\beta\,\mathrm{sign}(x)\,\frac{\pi}{2}\sec^2(\pi\alpha/2)$ is uniformly bounded on any compact $\mathcal{K} \subset \{\alpha > 1\}$ since $\sec^2(\pi\alpha/2)$ is continuous and finite for $\alpha \in (1,2)$, we obtain the following majoration,

$$\begin{aligned}\big|\partial_\alpha[\varsigma_j^\alpha\tilde{\varphi}_j]\big| &\le C_1\left[\frac{|\psi_j u+v|^\alpha}{\delta}+|u|^\alpha\right]\\ &\quad+C_2\left[\frac{|\psi_j u+v|^\alpha|\ln|\psi_j u+v||}{\delta}+|u|^\alpha|\ln|u||\right],\end{aligned}$$

where C_1 and C_2 are finite constants depending only on $\mathcal{K}$. This leads to a bounding function of the form

$$\left|\frac{\partial\varphi}{\partial\alpha}(u,v;\theta)\right| \le C_\alpha|\varphi(u,v;\theta)|\left[\frac{H_\alpha(\psi_j u+v)}{\delta}+H_\alpha(u)\right] = C_\alpha G_\alpha(u,v),$$

where $H_\alpha(x) = |x|^\alpha(1+|\ln|x||)$. To conclude on the integrability of $G_\alpha(u,v)$ against $w(u,v)$, we use a continuity argument. Since $\alpha \ge \alpha_0 > 1$ on $\mathcal{K}$, the function $x \mapsto H_\alpha(x)$ is continuous on $\mathbb{R}$ (prolonged by 0 at $x = 0$ since $\lim_{x\to0}|x|^\alpha\ln|x| = 0$). Consequently, $H_\alpha(x)$ is bounded on any compact set and grows polynomially at infinity. At this stage, we need Assumption 2 as it imposes $w(u,v) = (\kappa/\pi)\exp(-\kappa(u^2+v^2))$. As a consequence, the growth is dominated by the exponential decay of $w(u,v)$, ensuring that $\int G_\alpha(u,v)w(u,v)dudv < \infty$.

($\iota\iota\iota$) The derivative with respect to β is obtained by differentiating the asymmetry terms $\mathcal{A}(x)$ within the log-characteristic function expression (A.1). We have $\partial_\beta\mathcal{A}(x) = -i\mathrm{sign}(x)\tan(\pi\alpha/2)$ for $\alpha > 1$. Thus,

$$\begin{aligned}\frac{\partial}{\partial\beta}\big[\varsigma_j^\alpha\tilde{\varphi}_j\big] &= -\frac{\varsigma_j^\alpha}{1-|\psi_j|^\alpha}|\psi_j u+v|^\alpha\frac{\partial\mathcal{A}(\psi_j u+v)}{\partial\beta}-\varsigma_j^\alpha|u|^\alpha\frac{\partial\mathcal{A}(u)}{\partial\beta}\\ &= i\varsigma_j^\alpha\tan(\pi\alpha/2)\left[\frac{|\psi_j u+v|^\alpha\mathrm{sign}(\psi_j u+v)}{1-|\psi_j|^\alpha}+|u|^\alpha\mathrm{sign}(u)\right]\end{aligned}$$

The derivative of the full characteristic function is $\frac{\partial\varphi}{\partial\beta} = \varphi\sum_{j=1}^J\frac{\partial}{\partial\beta}\big[\varsigma_j^\alpha\tilde{\varphi}_j\big]$, by (A.2). Taking the modulus and applying the triangle inequality along with the uniform bounds on the compact set $\mathcal{K}$ (specifically $|\mathrm{sign}(\cdot)| \le 1$ and

$|\tan(\pi\alpha/2)| \le \sup_{\alpha\in\mathcal{K}} |\tan(\pi\alpha/2)| := M_\omega$), we obtain

$$\left|\frac{\partial\varphi}{\partial\beta}(u,v;\theta)\right| \le |\varphi(u,v;\theta)| \sum_{j=1}^{J} \varsigma_j^\alpha M_\omega \left[\frac{|\psi_j u+v|^\alpha}{\delta} + |u|^\alpha\right]$$
$$\le C_\beta |\varphi(u,v;\theta)| \left[\sum_{j=1}^{J} \frac{|\psi_j u+v|^\alpha}{\delta} + J|u|^\alpha\right].$$

The functional form of this bound is identical to that found for the derivative with respect to ς_k (a polynomial of degree α in u, v multiplied by the characteristic function). Consequently, it is integrable against the exponential weight $w(u,v)$ for any $\alpha \in (1,2)$.

($\iota\nu$) Finally, we turn to the most critical case, the derivative with respect to ψ_k. For $\alpha > 1$ the map $x \mapsto |x|^\alpha \mathcal{A}(x) = |x|^\alpha - i\beta\tan(\pi\alpha/2)\, x|x|^{\alpha-1}$ is continuously differentiable on all of $\mathbb{R}$, with derivative $\alpha|x|^{\alpha-1}\big(\mathrm{sign}(x) - i\beta\tan(\pi\alpha/2)\big)$. No distributional subtlety therefore arises, provided the numerator is differentiated as a whole rather than factor by factor: differentiating $\mathcal{A}(\psi_k u+v)$ separately would produce $\partial_{\psi_k}\mathrm{sign}(\psi_k u+v)$, a Dirac mass on the line $\psi_k u + v = 0$, which the product form avoids. Writing $x = \psi_k u + v$ and recalling from (A.2) that component k contributes $\varsigma_k^\alpha \tilde{\varphi}_k$,

$$\begin{aligned}\frac{\partial}{\partial\psi_k}\left[\varsigma_k^\alpha \tilde{\varphi}_k\right] &= -\,\varsigma_k^\alpha \frac{\partial}{\partial\psi_k}\left[\frac{|x|^\alpha \mathcal{A}(x)}{1-|\psi_k|^\alpha}\right] \\ &= -\,\varsigma_k^\alpha \left[\frac{\alpha\, u\, |x|^{\alpha-1}\big(\mathrm{sign}(x) - i\beta\tan\frac{\pi\alpha}{2}\big)}{1-|\psi_k|^\alpha} + \frac{\alpha\, |\psi_k|^{\alpha-1}\mathrm{sign}(\psi_k)\, |x|^\alpha \mathcal{A}(x)}{(1-|\psi_k|^\alpha)^2}\right].\end{aligned} \tag{A.4}$$

The first term is the one that fails to be differentiable along the line $x = 0$; its exponent $\alpha - 1 \in (0,1)$ being positive, the term itself vanishes there, though its derivative blows up. The second term, of order $|x|^\alpha$, is smoother still. Using the uniform bounds on the compact set $\mathcal{K}$ (specifically $|\mathcal{A}(\cdot)| \le M_\mathcal{A}$ and $1 - |\psi_k|^\alpha \ge \delta$, together with $|\beta\tan(\pi\alpha/2)| \le M_\mathcal{A}$), we define the bound for the singular part:

$$T_1(u,v) = \frac{\alpha|u||\psi_k u+v|^{\alpha-1}}{\delta} M_\mathcal{A}.$$

Again, we need Assumption 2 and impose $w(u,v) = (\kappa/\pi)\exp(-\kappa(u^2+v^2))$, for $\kappa > 0$, to prove the convergence. We rely on the polar coordinates of the integral of T_1:

$$\begin{aligned} I_1 &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} T_1(u,v)\,\frac{\kappa}{\pi}\exp(-\kappa(u^2+v^2))\,du\,dv \\ &= \frac{\alpha M_\mathcal{A}}{\delta}\frac{\kappa}{\pi}\int_0^{2\pi}\int_0^\infty r|\cos\phi|\cdot r^{\alpha-1}|\psi_k\cos\phi+\sin\phi|^{\alpha-1}e^{-\kappa r^2}\,r\,dr\,d\phi \\ &= \frac{\alpha M_\mathcal{A}}{\delta}\int_0^{2\pi}|\cos\phi||\psi_k\cos\phi+\sin\phi|^{\alpha-1}\,d\phi\cdot I_r(\alpha,\kappa), \end{aligned}$$

with $u = r\cos\phi$, $v = r\sin\phi$. This decomposition reveals that the radial integral converges for $\alpha > -2$:

$$I_r(\alpha,\kappa) := \frac{\kappa}{\pi}\int_0^\infty r^{\alpha+1}e^{-\kappa r^2}\,dr = \frac{\Gamma\left(\frac{\alpha+2}{2}\right)}{2\pi\,\kappa^{\alpha/2}} < \infty. \tag{A.5}$$

The angular integrand is in fact bounded near the zeros ϕ_0 of $\psi_k\cos\phi + \sin\phi$, its exponent $\alpha - 1$ being positive; a fortiori the angular integral converges, and it would already do so for $\alpha > 0$:

$$\int_{\phi_0-\epsilon}^{\phi_0+\epsilon}|\psi_k\cos\phi+\sin\phi|^{\alpha-1}d\phi \sim \int_{-\epsilon}^{\epsilon}|C\tau|^{\alpha-1}d\tau = \frac{2C^{\alpha-1}\epsilon^\alpha}{\alpha} < \infty,$$

with $\epsilon > 0$ an arbitrary small constant and $C = \sqrt{1+\psi_k^2}$. Therefore, for any $\alpha \in (1,2)$, the derivative is bounded by an integrable function:

$$\left|\frac{\partial\varphi}{\partial\psi_k}(u,v;\theta)\right| \le C_\psi|\varphi(u,v;\theta)|\left[\frac{|u||\psi_k u+v|^{\alpha-1}}{\delta} + \frac{|\psi_k u+v|^\alpha}{\delta^2}\right] = C_\psi G_\psi(u,v), \tag{A.6}$$

where C_ψ is a suitable constant. All terms in $G_\psi(u,v)$ are integrable against $w(u,v)$ for $\alpha \in (1,2)$. This completes the first-order derivative analysis, establishing that each component of the gradient $\nabla_\theta \varphi(u,v;\theta)$ admits a w-integrable dominant. We now turn to the second-order derivatives.

Finally, we analyze the second derivatives to establish $C^2(\Theta)$ regularity. The most critical terms arise from the second derivative with respect to ψ_k, specifically from the modulus term $|\psi_k u+v|^\alpha$. Using the decomposition in (A.1), we have

$$\frac{\partial^2}{\partial \psi_k^2}\left[\varsigma_k^\alpha \tilde{\varphi}_k\right] = -\frac{\varsigma_k^\alpha}{1-|\psi_k|^\alpha}\frac{\partial^2}{\partial \psi_k^2}\Big[|\psi_k u+v|^\alpha \mathcal{A}(\psi_k u+v)\Big] + R_k^{(2)}(u,v),$$

where $R_k^{(2)}(u,v)$ collects terms involving first derivatives of the modulus and derivatives of the coefficients, which are less singular. Specifically, $R_k^{(2)}(u,v) = O(|u|^\alpha + |v|^\alpha)$ as $||(u,v)|| \to \infty$ and is locally integrable. The dominant singular term is

$$\frac{\partial^2}{\partial \psi_k^2}\Big[|\psi_k u+v|^\alpha \mathcal{A}(\psi_k u+v)\Big] = \alpha(\alpha-1)u^2|\psi_k u+v|^{\alpha-2}\big(\mathrm{sign}(\psi_k u+v) - i\beta\tan\tfrac{\pi\alpha}{2}\big)\mathrm{sign}(\psi_k u+v),$$

The integrability of this term against $w(u,v)$ determines the $C^2(\Theta)$ regularity. Let us bound the integral of the modulus of this second derivative

$$T_2(u,v) = C|\varphi(u,v;\theta)|u^2|\psi_k u+v|^{\alpha-2}.$$

Using polar coordinates ($u = r\cos\phi, v = r\sin\phi$) and the weight $w(u,v) = \frac{\kappa}{\pi}e^{-\kappa r^2}$, the integral becomes

$$\begin{aligned} I_2 &= \frac{\kappa}{\pi}\int_0^{2\pi}\int_0^\infty r^2\cos^2\phi \cdot r^{\alpha-2}|\psi_k\cos\phi+\sin\phi|^{\alpha-2}e^{-\kappa r^2}\, r\, dr\, d\phi \\ &= I_r(\alpha,\kappa)\int_0^{2\pi}\cos^2\phi|\psi_k\cos\phi+\sin\phi|^{\alpha-2}d\phi. \end{aligned}$$

The radial integral converges for $\alpha > -2$. The angular integral $J_\alpha = \int_0^{2\pi}\cos^2\phi|\psi_k\cos\phi+\sin\phi|^{\alpha-2}d\phi$ presents singularities when $\psi_k\cos\phi+\sin\phi = 0$. Let ϕ_0 be such a singularity. Locally, the integrand behaves like $|\phi-\phi_0|^{\alpha-2}$. Convergence requires

$$\int_{\phi_0-\epsilon}^{\phi_0+\epsilon}|\tau|^{\alpha-2}d\tau < \infty \iff \alpha-2 > -1 \iff \alpha > 1.$$

For $\alpha \in (1,2)$, the angular integral is finite. The other second-order partial derivatives are covered by the same analysis: the mixed partials such as $\partial^2/\partial\psi_k\partial\alpha$ produce factors $u\,|\psi_k u+v|^{\alpha-1}\ln|\psi_k u+v|$, and $\partial^2/\partial\alpha^2$ produces factors $|x|^\alpha(\ln|x|)^2$ (with $x = \psi_j u+v$ or $x = u$); in polar coordinates the corresponding angular exponents are $\alpha-1 > -1$ up to logarithmic corrections, which do not alter the integrability thresholds, so all these terms are w-integrable for $\alpha \in (1,2)$. The remaining terms in the second derivative of the objective function $D_{\mathcal{X}}(\theta)$ involve products of first derivatives (which are integrable because each first-derivative envelope $G_\bullet$ constructed above is in fact square-integrable against w: the most singular case is $|u|^2|\psi_k u+v|^{2(\alpha-1)}$, whose angular exponent $2(\alpha-1) > -1$ remains integrable) or the second derivative analyzed above.

At first order this suffices: since $\alpha-1>0$, the envelopes just constructed are uniform over θ on a neighbourhood, because

$$\sup_{\psi\in[\underline{\psi},\overline{\psi}]}|u|\,|\psi u+v|^{\alpha-1} \;\le\; |u|\big(|u|+|v|\big)^{\alpha-1}\max(1,\overline{\psi})^{\alpha-1} \;\in\; L^1(w),$$

and ordinary dominated convergence gives $D_{\mathcal{X}} \in C^1(\Theta)$ with $\nabla D_{\mathcal{X}} = \iint \partial_\theta[\,\cdot\,]\,w$. At second order, dominated convergence is not available, and this is not an accident of the proof: as shown in Appendix A.1.3, the supremum over any neighbourhood V of θ of $u^2|\psi_k u+v|^{\alpha-2}$ is infinite on a cone of positive Lebesgue measure (the cone swept by the lines $\{\psi u + v = 0\}$ as ψ ranges over the projection of V, which is nondegenerate because $\theta_0 \in \mathrm{Int}(\Theta)$). No locally uniform w-integrable envelope for the second derivatives exists. We therefore replace dominated convergence by the following elementary substitute, which requires only what is available.

Lemma A.1. *Let $h(\theta) := \iint f(u, v, \theta)\, w\, du\, dv$ on an open set $U \subset \mathbb{R}^p$ (we apply it with U an open neighbourhood of the compact Θ, on which all the bounds below hold, this being how $C^2(\Theta)$ is to be read throughout) and assume: (ι) for almost every (u, v), $\theta \mapsto \partial_\theta f(u, v, \theta)$ is absolutely continuous on line segments of U, with almost-everywhere derivative $\partial^2_{\theta\theta'} f$; ($\iota\iota$) $\nabla h(\theta) = \iint \partial_\theta f\, w$ for every θ; ($\iota\iota\iota$) $\theta \mapsto \partial^2_{\theta\theta'} f(\cdot\,; \theta)$ is a continuous map $U \to L^1(w\, du\, dv)$. Then $h \in C^2(U)$ and $\nabla^2 h(\theta) = \Phi(\theta) := \iint \partial^2_{\theta\theta'} f(\cdot\,; \theta)\, w$.*

Proof. Fix $\theta_* \in U$, a direction e and s small, and set $\theta_t := \theta_* + te$. By ($\iota$) and ($\iota\iota$),

$$\nabla h(\theta_s) - \nabla h(\theta_*) = \iint \Big[\int_0^s \partial^2_{\theta\theta'} f(\cdot\,; \theta_t)\, e\, dt\Big] w = \int_0^s \Phi(\theta_t)\, e\, dt,$$

the exchange of integrals being licensed by Fubini, since $t \mapsto \iint \|\partial^2_{\theta\theta'} f(\cdot\,; \theta_t)\|\, w$ is continuous by ($\iota\iota\iota$), hence bounded on the compact segment, so that $\int_0^s \iint \|\partial^2_{\theta\theta'} f(\cdot\,; \theta_t)\|\, w\, dt < \infty$. As ($\iota\iota\iota$) also makes Φ continuous, the fundamental theorem of calculus gives $\partial_s \nabla h(\theta_s)|_{s=0} = \Phi(\theta_*)$, and $h \in C^2$. □

Lemma A.1-(ι) holds here: the only non-smooth dependence is through $\psi \mapsto \alpha\, u\, |\psi u + v|^{\alpha-1} \mathrm{sign}(\psi u + v)$, proportional to $|\psi - \psi^*|^{\alpha-1} \mathrm{sign}(\psi - \psi^*)$, which is absolutely continuous because $\int |\psi - \psi^*|^{\alpha-2} d\psi < \infty$ for $\alpha > 1$. Lemma A.1-($\iota\iota$) is the C^1 step above. Lemma A.1-($\iota\iota\iota$) follows from (A.8). Indeed, with $f = |\varphi_n - \varphi(\cdot\,; \theta)|^2$, (2.20) gives $\partial^2_{\theta\theta'} f = 2\,\mathrm{Re}[\partial_\theta \varphi\, \overline{\partial_{\theta'} \varphi}] - 2\,\mathrm{Re}[\overline{f_\theta}\, \partial^2_{\theta\theta'} \varphi]$, with $f_\theta := \varphi_n - \varphi(\cdot\,; \theta)$. The first term is $L^1(w)$-continuous in θ because $\theta \mapsto \partial_\theta \varphi(\cdot\,; \theta)$ is continuous into $L^2_w(\mathbb{R}^2)$ (established below from the first-order envelopes alone) and $\|fg\|_{L^1(w)} \le \|f\|_{L^2_w} \|g\|_{L^2_w}$. For the second, inserting $\pm\, \partial^2_{\theta\theta'} \varphi(\cdot\,; \theta)$ and using $|f_\theta| \le 2$,

$$\iint \big\|\overline{f_{\theta'}}\, \partial^2_{\theta\theta'} \varphi(\theta') - \overline{f_\theta}\, \partial^2_{\theta\theta'} \varphi(\theta)\big\|\, w \;\le\; \iint |\varphi(\theta') - \varphi(\theta)|\, \big\|\partial^2_{\theta\theta'} \varphi(\theta)\big\|\, w \;+\; 4 \iint \big\|\partial^2_{\theta\theta'} \varphi(\theta') - \partial^2_{\theta\theta'} \varphi(\theta)\big\|\, w,$$

where the first term tends to zero by ordinary dominated convergence (its dominating function $2\|\partial^2_{\theta\theta'} \varphi(\cdot\,; \theta)\|$ is fixed and w-integrable by Assumption 7) and the second by (A.8). Appendix A.1.3 uses only the pointwise derivative computations carried out above (the envelopes T_1, T_2 and $G_\bullet$) and not the present conclusion, so the reference is not circular. Applying Lemma A.1 with $f = |\varphi_n - \varphi(\cdot\,; \theta)|^2$, the objective function is $C^2(\Theta)$ for $\alpha \in (1, 2)$, and Assumption 3 is satisfied under this condition.

A.1.2. Validation of Assumption 6

Assumption 6 requires the random sequence $K(x; \theta)$ defined in (2.17) to be measurable and bounded. Since trigonometric functions and the theoretical characteristic function $\varphi(u, v; \theta)$ are continuous (and thus measurable), the entire integrand in (2.17) is a measurable function of x for each fixed (u, v, θ). By the Fubini theorem, the integral of this function with respect to (u, v) is a measurable function of x. Next, we demonstrate that $K(x; \theta)$ is uniformly bounded with respect to x. From the natural bounds of trigonometric functions, $|\cos(ux_1 + vx_2)| \le 1$ and $|\sin(ux_1 + vx_2)| \le 1$, and since $|\varphi(u, v; \theta)| \le 1$, we have

$$\begin{aligned}
|K(x; \theta)| &\le \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Bigg[\big(|\cos(ux_1 + vx_2)| + |\mathrm{Re}\, \varphi(u, v; \theta)|\big) \left| \frac{\partial \mathrm{Re}\, \varphi(u, v; \theta)}{\partial \theta} \right| \\
&\qquad + \big(|\sin(ux_1 + vx_2)| + |\mathrm{Im}\, \varphi(u, v; \theta)|\big) \left| \frac{\partial \mathrm{Im}\, \varphi(u, v; \theta)}{\partial \theta} \right| \Bigg] w(u, v)\, du\, dv \\
&\le 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\left| \frac{\partial \mathrm{Re}\, \varphi(u, v; \theta)}{\partial \theta} \right| + \left| \frac{\partial \mathrm{Im}\, \varphi(u, v; \theta)}{\partial \theta} \right| \right) w(u, v)\, du\, dv \\
&\le 2\sqrt{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left| \frac{\partial \varphi(u, v; \theta)}{\partial \theta} \right| w(u, v)\, du\, dv := B(\theta),
\end{aligned}$$

where the last inequality follows from the Cauchy-Schwarz inequality: for any complex number $z = a + ib$, we have $|a| + |b| \le \sqrt{2}|z|$ since $(|a| + |b|)^2 \le 2(a^2 + b^2) = 2|z|^2$.

The first-order analysis in the proof of Lemma 2.1 established that for each parameter component θ_i, the integral of the derivatives is finite, i.e.,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|\frac{\partial\varphi(u,v;\theta)}{\partial\theta_i}\right| w(u,v)\,du\,dv < \infty.$$

Furthermore, the smoothness of φ in θ established directly in the proof of Lemma 2.1 shows that, for each fixed (u, v), the gradient $\partial\varphi/\partial\theta$ is continuous in θ; combined with the K-uniform w-integrable envelopes $G_\bullet(u, v)$ constructed there, the dominated convergence theorem yields the continuity of the integral function $B(\theta)$ on the parameter space Θ. Since Θ is compact (Assumption 1), the continuous function $B(\theta)$ is bounded. We thus have

$$\sup_{\theta\in\Theta}\sup_{x}|K(x;\theta)| \le \sup_{\theta\in\Theta} B(\theta) < \infty.$$

This uniform boundedness ensures that Assumption 6 is satisfied. □

A.1.3. Validation of Assumption 7

Assumption 7 has two parts: the nonsingularity of the matrix $\Sigma(\theta_0)$, and an integrability together with a continuity condition on the second derivatives $\partial^2\varphi(u, v; \theta)/\partial\theta\partial\theta'$. We verify the second here, and then show that in the specification of Case 1 (and, by the same argument, in that of Case 2) the first is not an additional restriction either, but a consequence of Definition 2.1 together with $\alpha \in (1, 2)$ and $\varsigma_j > 0$. It is retained as an assumption in the general statement only because the argument does not extend to the MAR(r_j, s_j) aggregates of Case 4, for the reason given at the end of the subsection.

Form of the domination condition. The domination in Assumption 7 bears on the w-integrals of the second derivatives rather than on a pointwise envelope, and this is the form suited to α-stable scores: a single $g \in L^1(w\,du\,dv)$ dominating $\|\partial^2\varphi/\partial\theta\partial\theta'\|$ pointwise for almost every (u, v) would have to be infinite on a set of positive measure. Indeed, the dominant term of $\partial^2\varphi/\partial\psi_k^2$ is

$$\varphi(u,v;\theta)\cdot\frac{\varsigma_k^\alpha\,\alpha(\alpha-1)\,u^2\,|\psi_k u+v|^{\alpha-2}}{1-|\psi_k|^\alpha}\,\mathcal{A}_k(u,v), \qquad \alpha-2<0.$$

Let $[\underline{\psi}, \overline{\psi}] \subset (0, 1)$ denote the range of ψ_k over the compact Θ and consider the cone of positive Lebesgue measure $S := \{(u, v) : u \in [1, 2],\ -v/u \in [\underline{\psi}, \overline{\psi}]\}$. For every $(u, v) \in S$, the value $\psi^* = -v/u$ is admissible and letting $\psi_k \to \psi^*$ drives $|\psi_k u + v| \to 0$ while $|\varphi|$, $|\mathcal{A}_k|$, u^2 and $\varsigma_k^\alpha/(1 - |\psi_k|^\alpha)$ remain bounded away from zero on S. Hence $\sup_{\theta\in\Theta} |\partial^2\varphi/\partial\psi_k^2| = +\infty$ everywhere on S. The w-integrated version of the domination, on the other hand, is finite and continuous in θ, as we now show.

Integrability at fixed θ. For each fixed $\theta \in \Theta$, the analysis of Lemma 2.1 shows that the most singular contribution to $\|\partial^2\varphi/\partial\theta\partial\theta'\|$ is of the form $C\,u^2|\psi_k u + v|^{\alpha-2}$, all remaining terms being of polynomial growth (possibly with logarithmic factors) and therefore w-integrable. Passing to polar coordinates,

$$\iint u^2|\psi u+v|^{\alpha-2}\,w(u,v)\,du\,dv = I_r(\alpha,\kappa)\int_0^{2\pi}\cos^2\phi\,|\psi\cos\phi+\sin\phi|^{\alpha-2}\,d\phi<\infty \tag{A.7}$$

for $\alpha \in (1, 2)$, where $I_r(\alpha, \kappa) = \Gamma(\frac{\alpha+2}{2})/(2\pi\kappa^{\alpha/2})$, since the angular singularities are of order $\alpha - 2 > -1$. Hence $\iint \|\partial^2\varphi(u, v; \theta)/\partial\theta\partial\theta'\|\,w\,du\,dv < \infty$ for each $\theta \in \Theta$.

$L^1(w)$-continuity of the Hessian. Let $\theta^{(m)} \to \theta$ in Θ. We show

$$\iint\left\|\partial^2_{\theta\theta'}\varphi(\cdot\,;\theta^{(m)})-\partial^2_{\theta\theta'}\varphi(\cdot\,;\theta)\right\|w\,du\,dv\ \longrightarrow\ 0, \tag{A.8}$$

which is the continuity clause of Assumption 7. The integrands converge to 0 almost everywhere: for almost every (u,v) one has $\psi_k u+v\neq 0$ for every k, and at such a point $\theta\mapsto\partial^2_{\theta\theta'}\varphi(u,v;\theta)$ is continuous. Recall that $\underline{\alpha}:=\min\{\alpha:\theta\in\Theta\}>1$, define also $\overline{\alpha}:=\max\{\alpha:\theta\in\Theta\}<2$, and let $C:=\sup_{\theta\in\Theta}C(\theta)<\infty$ collect the θ-uniform bounds ς_k^α, $M_{\mathcal{A}}$, δ^{-1} and $\alpha(\alpha-1)$ of the proof of Lemma 2.1. Write g_θ for the envelope of $\|\partial^2_{\theta\theta'}\varphi(\cdot\,;\theta)\|$,

$$g_\theta(u,v):=C\sum_{k=1}^{J}u^2|\psi_k u+v|^{\alpha-2}\;+\;C\big(1+|u|^\alpha+|v|^\alpha\big)^2\Big(1+\big|\ln|u|\big|+\sum_{k=1}^{J}\big|\ln|\psi_k u+v|\big|\Big)^2,$$

the squared second block accounting for the products of first derivatives that arise in $\partial^2_{\theta\theta'}\varphi=\varphi\,\big(\partial^2_{\theta\theta'}\log\varphi+\partial_\theta\log\varphi\,\partial_{\theta'}\log\varphi\big)$, including the cross terms $|\psi_j u+v|^{\alpha-1}|\psi_\ell u+v|^{\alpha-1}$. That block is w-integrable, the weight being Gaussian and the added singularities only logarithmic, and $\theta\mapsto\iint(\text{second block})\,w$ is continuous by ordinary dominated convergence; only the first block requires the argument below. Put $g_m:=g_{\theta^{(m)}}$, $g:=g_\theta$. The difference is dominated by g_m+g, and $\int(g_m+g)\,w\to 2\int g\,w$ as soon as $\int g_m\,w\to\int g\,w$, so Pratt's lemma (the generalized dominated convergence theorem) applies to the sequence g_m+g. It remains to check $\int g_m\,w\to\int g\,w$, i.e. the joint continuity in (ψ_k,α) of the map

$$(\psi,\alpha)\longmapsto\iint u^2|\psi u+v|^{\alpha-2}\,w(u,v)\,du\,dv=I_r(\alpha,\kappa)\int_0^{2\pi}\cos^2\phi\,|\psi\cos\phi+\sin\phi|^{\alpha-2}\,d\phi$$

(the remaining, logarithmic, terms are handled identically and are less singular). The radial factor $I_r(\alpha,\kappa)$ is continuous in α by (A.5). For the angular factor, an envelope free of α but still depending on ψ would not suffice, since its singular points $\{\phi:\psi\cos\phi+\sin\phi=0\}$ move with ψ and dominated convergence along $\psi^{(m)}\to\psi$ requires a fixed dominating function. We therefore freeze the singularities by a rotation. Writing $\varpi:=\arctan\psi$, so that $\psi\cos\phi+\sin\phi=\sqrt{1+\psi^2}\,\sin(\phi+\varpi)$, and substituting $\phi'=\phi+\varpi$,

$$\int_0^{2\pi}\cos^2\phi\,|\psi\cos\phi+\sin\phi|^{\alpha-2}\,d\phi=\big(1+\psi^2\big)^{\frac{\alpha-2}{2}}\int_0^{2\pi}\cos^2(\phi'-\varpi)\,|\sin\phi'|^{\alpha-2}\,d\phi'. \tag{A.9}$$

The singularities now sit at the fixed points $\phi'\in\{0,\pi\}$; the prefactor and $\cos^2(\phi'-\varpi)$ are jointly continuous and bounded in (ψ,α) on $[\underline{\psi},\overline{\psi}]\times[\underline{\alpha},\overline{\alpha}]$; and $|\sin\phi'|^{\alpha-2}\le|\sin\phi'|^{\underline{\alpha}-2}\in L^1[0,2\pi]$, since $|\sin\phi'|\le 1$ and $\underline{\alpha}-2>-1$, is a dominating function free of both ψ and α. Ordinary dominated convergence therefore gives joint continuity in (ψ,α), whence $\int g_m\,w\to\int g\,w$. (Equivalently, in Cartesian coordinates, the substitution $v'=\psi u+v$ moves the whole ψ-dependence into the smooth Gaussian weight.)

The Hessian step of Knight and Yu (2002) under the integrated domination. We now check, that the integrated condition just verified is enough for the step of the proof of Theorem 2.1 of Knight and Yu (2002) in which their pointwise condition (A6) is used. Let $(\bar\theta_n)$ be any (possibly random) sequence in Θ with $\bar\theta_n\to\theta_0$ almost surely, and set $\eta_n:=\varphi_n-\varphi(\cdot\,;\bar\theta_n)$, so that $\|\eta_n\|_\infty\le 2$ and, almost surely, $\eta_n\to 0$ at every (u,v). The ergodic theorem gives, for each fixed (u,v), a null set depending on (u,v); a single null set is obtained by applying it on a countable dense subset of $\mathbb{R}^2$ and invoking the almost-sure equicontinuity of $\{\varphi_n\}$ on compacts, which holds because $\mathbb{E}|\mathcal{X}_t|<\infty$ for $\alpha>1$. Only convergence for $w\,du\,dv$-almost every (u,v) is used below. Evaluating (2.20) at $\bar\theta_n$, the Hessian splits into two integrals.

For the first integral, $\iint\mathrm{Re}[\partial_\theta\varphi(\cdot\,;\bar\theta_n)\overline{\partial_{\theta'}\varphi(\cdot\,;\bar\theta_n)}]\,w\to\Sigma(\theta_0)$ almost surely, because $\theta\mapsto\partial_\theta\varphi(\cdot\,;\theta)$ is continuous in $L^2_w(\mathbb{R}^2)$: each first-order envelope $G_\bullet$ constructed in the proof of Lemma 2.1 is square-integrable against w (the most singular case being $|u|^2|\psi_k u+v|^{2(\alpha-1)}$, whose angular exponent $2(\alpha-1)>-1$), so dominated convergence applies.

For the second integral, inserting $\pm\,\partial^2_{\theta\theta'}\varphi(\cdot\,;\theta_0)$,

$$\Big|\iint\mathrm{Re}\big[\overline{\eta_n}\,\partial^2_{\theta\theta'}\varphi(\cdot\,;\bar\theta_n)\big]w\Big|\;\le\;2\iint\big\|\partial^2_{\theta\theta'}\varphi(\cdot\,;\bar\theta_n)-\partial^2_{\theta\theta'}\varphi(\cdot\,;\theta_0)\big\|\,w\;+\;\Big|\iint\mathrm{Re}\big[\overline{\eta_n}\,\partial^2_{\theta\theta'}\varphi(\cdot\,;\theta_0)\big]w\Big|.$$

The first term tends to zero by (A.8), applied along $\bar{\theta}_n \to \theta_0$; note that the multiplier has already been eliminated at this stage by $|\eta_n| \le 2$, so what is used is exactly the $L^1(w)$-continuity of Assumption 7 and nothing stronger (in particular no uniformity in η_n, which is fortunate since η_n is random). The second tends to zero by ordinary dominated convergence: $\eta_n \to 0$ pointwise, $|\eta_n| \le 2$, and $\|\partial^2_{\theta\theta'}\varphi(\cdot\,;\theta_0)\| \in L^1(w\,du\,dv)$ by the finiteness clause of Assumption 7.

Consequently $\nabla^2 D_{\mathcal{X}}(\bar{\theta}_n) \to 2\Sigma(\theta_0)$ almost surely, which is condition (b) of the proof of Theorem 2.1 in Knight and Yu (2002). Two features of this argument are worth isolating. First, what is needed is convergence along the consistent sequence, not uniform convergence over Θ: the latter is stronger than the integrated condition delivers, and is not used anywhere in Knight and Yu (2002). Second, the pointwise dominance of (A6) is never invoked; it is replaced by the $L^1(w)$-continuity of the Hessian map alone, and this is the whole content of the adaptation of Knight and Yu (2002) to the α-stable case. The integrability and continuity conditions of Assumption 7 are thus verified, and suffice.

On the nonsingularity of $\Sigma(\theta_0)$. The matrix $\Sigma(\theta_0)$ is the Gram matrix of the score functions in the weighted Hilbert space $L^2_w(\mathbb{R}^2)$ equipped with the inner product

$$\langle f, g\rangle_w = \int_{\mathbb{R}^2} f(u,v)\overline{g(u,v)}\,w(u,v)\,du\,dv.$$

Explicitly, $[\Sigma(\theta_0)]_{ij} = \langle g_i, g_j\rangle_w$ where $g_i = \partial\varphi(u,v;\theta_0)/\partial\theta_i = \varphi(u,v;\theta_0)\,\partial\log\varphi(u,v;\theta_0)/\partial\theta_i$. Since $|\varphi| = e^{\mathrm{Re}\log\varphi} > 0$ everywhere, the linear independence in $L^2_w(\mathbb{R}^2)$ of the family $\{g_i\}$ is equivalent to that of the log-scores $\{\partial\log\varphi/\partial\theta_i\}$, whose expressions are displayed below. Since a Gram matrix is nonsingular if and only if the generating vectors are linearly independent, nonsingularity amounts to the linear independence of the score functions $\{g_{\varsigma_1},\dots,g_{\varsigma_J},g_{\psi_1},\dots,g_{\psi_J},g_\alpha,g_\beta\}$ in $L^2_w(\mathbb{R}^2)$. We work with the reparametrized vector $\theta = (\varsigma_1,\dots,\varsigma_J,\psi_1,\dots,\psi_J,\alpha,\beta)'$ where $\varsigma_j = \sigma\pi_j$, consistently with Assumption 1. The log-characteristic function of the aggregate takes the form

$$\log\varphi(u,v;\theta) = -\sum_{j=1}^{J}\varsigma_j^\alpha\left(\frac{|\psi_j u + v|^\alpha}{1-|\psi_j|^\alpha}\mathcal{A}_j(u,v) + |u|^\alpha\mathcal{B}(u)\right), \tag{A.10}$$

where $\mathcal{A}_j(u,v) = 1 - i\beta\,\mathrm{sign}(\psi_j u + v)\tan(\pi\alpha/2)$ and $\mathcal{B}(u) = 1 - i\beta\,\mathrm{sign}(u)\tan(\pi\alpha/2)$.

The score functions are computed by differentiating (A.10). For each $k \in \{1,\dots,J\}$:

$$\begin{aligned} g_{\varsigma_k}(u,v) &= -\alpha\varsigma_k^{\alpha-1}\left(\frac{|\psi_k u + v|^\alpha}{1-|\psi_k|^\alpha}\mathcal{A}_k(u,v) + |u|^\alpha\mathcal{B}(u)\right),\\ g_{\psi_k}(u,v) &= -\varsigma_k^\alpha\frac{\partial}{\partial\psi_k}\left[\frac{|\psi_k u + v|^\alpha}{1-|\psi_k|^\alpha}\mathcal{A}_k(u,v)\right]. \end{aligned} \tag{A.11}$$

For the global parameters:

$$g_\alpha(u,v) = \sum_{j=1}^{J}\varsigma_j^\alpha\left(\ln(\varsigma_j)\cdot\tilde{\varphi}_j + \frac{\partial\tilde{\varphi}_j}{\partial\alpha}\right), \tag{A.12}$$

$$g_\beta(u,v) = i\tan\left(\frac{\pi\alpha}{2}\right)\sum_{j=1}^{J}\varsigma_j^\alpha\left(\frac{|\psi_j u + v|^\alpha\,\mathrm{sign}(\psi_j u + v)}{1-|\psi_j|^\alpha} + |u|^\alpha\,\mathrm{sign}(u)\right), \tag{A.13}$$

where $\tilde{\varphi}_j(u,v) = -\frac{|\psi_j u+v|^\alpha}{1-|\psi_j|^\alpha}\mathcal{A}_j(u,v) - |u|^\alpha\mathcal{B}(u)$ denotes the unit-scale log-characteristic function of component j. Suppose there exist constants $\{c_{\varsigma_k}\}_{k=1}^J$, $\{c_{\psi_k}\}_{k=1}^J$, c_α, and c_β such that

$$\sum_{k=1}^{J}c_{\varsigma_k}g_{\varsigma_k}(u,v) + \sum_{k=1}^{J}c_{\psi_k}g_{\psi_k}(u,v) + c_\alpha g_\alpha(u,v) + c_\beta g_\beta(u,v) = 0 \quad \text{a.e.} \tag{A.14}$$

We now show that (A.14) forces all the coefficients to vanish, so that $\Sigma(\theta_0)$ is in fact nonsingular for every admissible θ_0 in this specification. The mechanism is a separation of singularities, but it must be applied to the derivatives of the scores: for $\alpha \in (1,2)$ the scores themselves are continuous across the null lines, and it is their transverse derivatives that blow up there.

(ι) Preliminaries: differentiating the null relation. Let

$$\mathcal{L}_k := \{(u,v) \in \mathbb{R}^2 : \psi_k u + v = 0\}, \qquad x_k := \psi_k u + v,$$

and let $\mathcal{C}$ be any connected component of $\mathbb{R}^2 \setminus \big(\{u=0\} \cup \mathcal{L}_1 \cup \cdots \cup \mathcal{L}_J\big)$. By Definition 2.1 the parameters $\psi_1, \dots, \psi_J$ are pairwise distinct, so the lines $\mathcal{L}_1, \dots, \mathcal{L}_J$ are pairwise distinct and these components are nonempty open cones. On $\mathcal{C}$ every x_j keeps a constant sign and none vanishes, so each score g_i is real-analytic there; since (A.14) holds almost everywhere and the g_i are continuous on $\mathcal{C}$, the relation holds identically on $\mathcal{C}$ and may be differentiated in (u,v) any number of times. The coefficients $c_\bullet$ are the same on every component. The coefficients are complex: $\Sigma(\theta_0)$ being the Gram matrix of the g_i in $L^2_w(\mathbb{R}^2)$, one has $c^*\Sigma(\theta_0)c = \iint |\sum_i \bar{c}_i g_i|^2 w\, du\, dv$, so that singularity is equivalent to the vanishing of a nontrivial complex combination. (One could instead take $c \in \mathbb{R}^p$: $\Sigma(\theta_0)$ is real by Remark 2.1, and a real singular matrix has a real null vector. The argument below does not need this and is written for complex coefficients.) It is the transverse derivatives of the scores, not the scores themselves, that separate the parameters. Indeed, for $\alpha \in (1,2)$ the exponent $\alpha - 1$ is strictly positive, so every score is continuous across each $\mathcal{L}_k$ and vanishes there: by (A.4),

$$g_{\psi_k} = -\varsigma_k^\alpha \left[\frac{\alpha\, u\, |x_k|^{\alpha-1}\big(\mathrm{sign}(x_k) - i\beta \tan\frac{\pi\alpha}{2}\big)}{1 - |\psi_k|^\alpha} + \frac{\alpha |\psi_k|^{\alpha-1} \mathrm{sign}(\psi_k)\, |x_k|^\alpha \mathcal{A}_k}{(1 - |\psi_k|^\alpha)^2} \right] \xrightarrow[x_k \to 0]{} 0.$$

Since $\partial_v x_j = 1$ for every j, we differentiate (A.14) with respect to v on $\mathcal{C}$ and let $x_k \to 0$ from within $\mathcal{C}$. *($\iota\iota$) $c_{\psi_k} = 0$ for all k.* Differentiating once,

$$\partial_v g_{\psi_k} = -\, \varsigma_k^\alpha\, \frac{\alpha(\alpha-1)\, u}{1 - |\psi_k|^\alpha}\, |x_k|^{\alpha-2} \mathcal{A}_k(u,v) \;+\; O\big(|x_k|^{\alpha-1}\big), \tag{A.15}$$

which diverges as $|x_k|^{\alpha-2}$, the exponent being negative. All the remaining transverse derivatives stay bounded near $\mathcal{L}_k$:

$$\partial_v g_{\varsigma_j} = O\big(|x_j|^{\alpha-1}\big), \quad \partial_v g_\beta = O\big(|x_j|^{\alpha-1}\big), \quad \partial_v g_\alpha = O\big(|x_j|^{\alpha-1} \ln |x_j|\big),$$

and $\partial_v g_{\psi_\ell}$ for $\ell \neq k$ is bounded near $\mathcal{L}_k$ because $|x_\ell| \to |(\psi_\ell - \psi_k)u| \neq 0$ there. Since $u \neq 0$ on $\mathcal{C}$, $\varsigma_k > 0$ and $|\mathcal{A}_k| \geq 1$, the coefficient in (A.15) is nonzero, and multiplying the differentiated relation by $|x_k|^{2-\alpha}$ before letting $x_k \to 0$ gives $c_{\psi_k} = 0$. This holds for each $k \in \{1, \dots, J\}$. *($\iota\iota\iota$) $c_\alpha = 0$.* With $c_\psi = 0$, (A.14) reduces to

$$\sum_{k=1}^J c_{\varsigma_k} g_{\varsigma_k}(u,v) + c_\alpha g_\alpha(u,v) + c_\beta g_\beta(u,v) = 0 \quad \text{on } \mathcal{C}. \tag{A.16}$$

Differentiate twice in v. Using $\partial^2_{xx}[|x|^\alpha \mathcal{A}(x)] = \alpha(\alpha-1)|x|^{\alpha-2}\mathcal{A}(x)$ and $\partial^2_{xx}[|x|^\alpha \ln|x|\, \mathcal{A}(x)] = |x|^{\alpha-2}\mathcal{A}(x)[\alpha(\alpha-1)\ln|x| + (2\alpha-1)]$, the term of g_α coming from $\partial_\alpha \tilde{\varphi}_k \ni -|x_k|^\alpha \ln|x_k|\, \mathcal{A}_k/(1 - |\psi_k|^\alpha)$ contributes

$$-\, \varsigma_k^\alpha\, \frac{\alpha(\alpha-1)}{1 - |\psi_k|^\alpha}\, |x_k|^{\alpha-2} \ln|x_k|\, \mathcal{A}_k \;+\; O\big(|x_k|^{\alpha-2}\big),$$

and it is the only term in (A.16) carrying a factor $\ln|x_k|$ at the order $|x_k|^{\alpha-2}$: $\partial^2_{vv} g_{\varsigma_j}$ and $\partial^2_{vv} g_\beta$ are $O(|x_k|^{\alpha-2})$ without a logarithm. Dividing by $|x_k|^{\alpha-2}\ln|x_k|$ and letting $x_k \to 0$, the surviving coefficient is $-c_\alpha\, \varsigma_k^\alpha\, \alpha(\alpha-1)\, \mathcal{A}_k/(1 - |\psi_k|^\alpha)$. Since $|\mathcal{A}_k| \geq 1$, $\varsigma_k > 0$ and $\alpha(\alpha-1) > 0$, it vanishes only if $c_\alpha = 0$. *($\iota\nu$) $c_{\varsigma_k} = c_\beta = 0$.* With $c_\psi = c_\alpha = 0$, only the scale and asymmetry scores remain. Retaining in ∂^2_{vv}(A.16) the coefficient of $\alpha(\alpha-1)|x_k|^{\alpha-2}/(1 - |\psi_k|^\alpha)$ and writing $\varepsilon := \mathrm{sign}(x_k) \in \{-1, +1\}$,

$$-\, c_{\varsigma_k}\, \alpha\, \varsigma_k^{\alpha-1}\Big(1 - i\beta \tan\tfrac{\pi\alpha}{2}\, \varepsilon\Big) \;+\; c_\beta\, i \tan\tfrac{\pi\alpha}{2}\, \varsigma_k^\alpha\, \varepsilon \;=\; 0. \tag{A.17}$$

Both values of ε are admissible, since $\mathcal{L}_k$ is approached from both sides, each side lying in a component on which (A.16) holds with the same constants. Adding the two equations (A.17) eliminates c_β and leaves $-2c_{\varsigma_k}\alpha\varsigma_k^{\alpha-1} = 0$, whence $c_{\varsigma_k} = 0$; substituting back gives $c_\beta\, i\tan(\pi\alpha/2)\,\varsigma_k^\alpha = 0$, and since $\tan(\pi\alpha/2) \neq 0$ for $\alpha \in (1,2)$ and $\varsigma_k > 0$, $c_\beta = 0$. All the coefficients in (A.14) vanish, so the score family is linearly independent in $L^2_w(\mathbb{R}^2)$, the scores belonging to that space by Lemma 2.1, and $\Sigma(\theta_0)$ is nonsingular. The argument uses $\alpha \in (1,2)$ twice, and essentially: once for $\alpha - 2 < 0$, which produces the divergence that separates ψ from the rest, and once for $\tan(\pi\alpha/2) \neq 0$, which separates β from ς. It requires no symmetry restriction, and covers $\beta_0 \neq 0$.

The nonsingularity clause of Assumption 7 is therefore not an additional restriction in Cases 1 and 2: it is a consequence of Definition 2.1 together with $\alpha \in (1,2)$ and $\varsigma_j > 0$. We nonetheless retain it as an assumption in the statement of Assumption 7, because in Case 4 the analogous separation is not available. There the score carries one singular line per lag, $\mathcal{L}_{j,k} = \{u\, d_{j,k} + v\, d_{j,k-1} = 0\}$ with $k \in \mathbb{Z}$, of slope $-d_{j,k}/d_{j,k-1}$; and these slopes are not separated. Indeed $d_{j,k}/d_{j,k-1} \to \zeta_{j,1}$ as $k \to +\infty$ and $d_{j,k}/d_{j,k-1} \to 1/\lambda_{j,1}$ as $k \to -\infty$, $\zeta_{j,1}$ and $\lambda_{j,1}$ denoting the roots of maximal modulus on each side, so that the lines accumulate at the two directions of slope $-\zeta_{j,1}$ and $-1/\lambda_{j,1}$. Near either of them infinitely many $\mathcal{L}_{j,k}$ pile up, and no single line can be approached with every other score bounded, which is what steps $(\iota\iota)$–$(\iota\nu)$ above require. There the nonsingularity remains primitive, as in the empirical-characteristic-function and minimum-distance literature, including applications of Knight and Yu (2002).

□

A.1.4. Validation of Assumption 8

Assumption 8 requires that the sequence $\{K_j\}$ exhibits sufficient temporal dependence decay to apply a central limit theorem for dependent processes. We establish this by showing that the aggregated process $(\mathcal{X}_t)$ satisfies a strong mixing condition with explicit geometric decay rates, invoking Theorem 4.4.1 in Rosenblatt (2000).

Each latent process $(X_{j,t})$, whether purely anticipative AR(1) or mixed MAR(1,1), admits a two-sided infinite moving average representation

$$X_{j,t} = \sum_{k=-\infty}^{+\infty} a_{j,k}\varepsilon_{j,t-k}, \tag{A.18}$$

where $(\varepsilon_{j,t})_{t\in\mathbb{Z}}$ are i.i.d. α-stable innovations with $\alpha \in (1,2)$, and the coefficients $a_{j,k}$ satisfy geometric decay: there exist constants $D_j > 0$ and $\lambda_j \in (0,1)$ such that $|a_{j,k}| \leq D_j\lambda_j^{|k|}$ for all $k \in \mathbb{Z}$. For the MAR(1,1) case, $\lambda_j = \max(|\phi_j|, |\psi_j|)$.

We now verify that the conditions of Theorem 4.4.1 in Rosenblatt (2000) are satisfied. Let $\varrho \in (1,\alpha)$ be a fixed moment exponent. The moment condition (4.4.2) holds since α-stable innovations with $\alpha \in (1,2)$ have finite moments of all orders less than α, i.e., $\mathbb{E}|\varepsilon_{j,t}|^\varrho < \infty$, and zero mean since $\alpha > 1$. The coefficient summability condition (4.4.3) follows from the geometric decay, as $\sum_{k\in\mathbb{Z}}|a_{j,k}| \leq D_j(1+\lambda_j)/(1-\lambda_j) < \infty$. For the invertibility condition (4.4.4), we define $a_j(e^{-i\xi}) = \sum_{k\in\mathbb{Z}} a_{j,k}e^{-ik\xi}$, which is continuous on $[-\pi,\pi]$ since the coefficients are absolutely summable. Note that no spectral density is available here: α-stable processes with $\alpha < 2$ have infinite variance, and the relevant object is the transfer function itself. Since the roots are bounded away from the unit circle, $|\phi_j|, |\psi_j| \leq 1 - \delta' < 1$, the transfer function $a_j(z) = \big((1-\phi_j z)(1-\psi_j z^{-1})\big)^{-1}$ satisfies $a_j(z) \neq 0, \infty$ on $|z| = 1$; its inverse $a_j^{-1}(z) = (1-\phi_j z)(1-\psi_j z^{-1})$ is a Laurent polynomial, hence trivially a bounded operator, and condition (4.4.4) holds without invoking Wiener's theorem. Finally, the density regularity condition (4.4.5) requires the density p of the innovations to satisfy $\int_{\mathbb{R}} |p(\xi + x) - p(\xi)|d\xi \leq c|x|$ for some constant $c > 0$. For α-stable distributions with $\alpha \in (1,2)$, the density p is infinitely differentiable with $p' \in L^1(\mathbb{R})$ (by Fourier inversion: the characteristic function of the innovations has modulus $e^{-|t|^\alpha}$, which decays faster than any polynomial, so $p^{(m)}(x) = \frac{1}{2\pi}\int_{\mathbb{R}}(-it)^m e^{-itx}\,\hat{p}(t)\,dt$ exists and is continuous for every $m \geq 0$; moreover p' is integrable since it is continuous and $p'(x) = O(|x|^{-\alpha-2})$ as $|x| \to \infty$). By the mean value theorem, $|p(\xi+x)-p(\xi)| \le |x| \sup_{t\in[0,1]} |p'(\xi+tx)|$, and integrating over ξ yields

$$\int_{\mathbb{R}} |p(\xi+x)-p(\xi)|\, d\xi \le |x| \int_{\mathbb{R}} |p'(\xi)|\, d\xi = |x| \|p'\|_{L^1} < \infty,$$

which establishes condition (4.4.5) with $c = \|p'\|_{L^1}$.

Applying Theorem 4.4.1 in Rosenblatt (2000), each component process $(X_{j,t})$ is strongly mixing[20] with coefficient satisfying, for k sufficiently large,

$$\alpha_{X_j}(2k) \le \varkappa_j W_j(k,\varrho), \tag{A.19}$$

where $\varkappa_j > 0$ is a constant and $W_j(k,\varrho) = \{\sum_{m=k}^{\infty} d_{j,m,\varrho}^{\varrho/(1+\varrho)}\} \vee \{\sum_{m=k}^{\infty} L(d_{j,m,2})\}$ with $d_{j,m,\mu} = \sum_{|l|>m} |a_{j,l}|^{\mu}$ and $L(u) = \sqrt{u[1 \vee |\ln u|]}$. Given the geometric decay $|a_{j,k}| \le D_j \lambda_j^{|k|}$, we have $d_{j,m,\mu} = O(\lambda_j^{\mu m})$, which yields $\alpha_{X_j}(h) \le C_j \lambda_j^{\gamma h}$ for constants $C_j > 0$ and $\gamma > 0$ depending on ϱ.

Since the latent processes $(X_{1,t}), \ldots, (X_{J,t})$ are mutually independent, the σ-algebra generated by the aggregate $\mathcal{X}_t = \sigma \sum_{j=1}^{J} \pi_j X_{j,t}$ is contained in the product σ-algebra generated by the components. For independent processes, the strong mixing coefficient of the joint process satisfies $\alpha_{(X_1,\ldots,X_J)}(h) \le J \max_{1\le j\le J} \alpha_{X_j}(h)$ (see Doukhan , 1994, Lemma 1.2.1). Since measurable functions of mixing processes inherit the mixing property, we obtain

$$\alpha_{\mathcal{X}}(h) \le J \cdot \max_{1\le j\le J} \alpha_{X_j}(h) \le C\lambda^h, \tag{A.20}$$

where $C = J \max_{1\le j\le J} C_j$ and $\lambda = \max_{1\le j\le J} \lambda_j^{\gamma} \in (0,1)$.

The score term K_t defined in (2.17) is a measurable function of the finite segment $(\mathcal{X}_t, \mathcal{X}_{t+1})$. Since $\sigma(K_s : s \le t) \subset \sigma(\mathcal{X}_s : s \le t+1)$ while $\sigma(K_s : s \ge t+h) \subset \sigma(\mathcal{X}_s : s \ge t+h)$, the two σ-fields are separated by $h-1$ time units, so that $\alpha_K(h) \le \alpha_{\mathcal{X}}(h-1) \le (C/\lambda)\,\lambda^h$: a measurable function of a block of length two inherits the mixing rate up to the block length.

It remains to verify the two conditions of Assumption 8, stated at θ_0 so that $\mathbb{E}[K_0] = 0$. The geometric mixing rate ensures that $\mathbb{E}[K_0|\mathcal{F}_{-m}]$ converges to $\mathbb{E}[K_0] = 0$ in L^2 as $m \to \infty$. For the summability condition, the projection differences $\nu_\ell = \mathbb{E}[K_0|K_{-\ell}, K_{-\ell-1}, \ldots] - \mathbb{E}[K_0|K_{-\ell-1}, K_{-\ell-2}, \ldots]$ satisfy $\mathbb{E}[\nu_\ell' \nu_\ell]^{1/2} = O(\lambda^{\ell/2})$. Indeed, writing $\mathcal{G}_{-\ell} := \sigma(K_{-\ell}, K_{-\ell-1}, \ldots)$ and using $\mathbb{E}[K_0] = 0$ together with the boundedness of K established for Assumption 6, the covariance inequality for bounded random variables gives

$$\big\|\mathbb{E}(K_0 \mid \mathcal{G}_{-\ell})\big\|_2^2 = \mathbb{E}\big[K_0\, \mathbb{E}(K_0 \mid \mathcal{G}_{-\ell})\big] \le 4\, \alpha_K(\ell)\, \|K\|_\infty^2,$$

whence $\|\nu_\ell\|_2 \le \|\mathbb{E}(K_0|\mathcal{G}_{-\ell})\|_2 + \|\mathbb{E}(K_0|\mathcal{G}_{-\ell-1})\|_2 = O(\alpha_K(\ell)^{1/2}) = O(\lambda^{\ell/2})$. This is the only place where the boundedness of the score is used quantitatively. Since $\lambda \in (0,1)$, the series $\sum_{\ell=0}^{\infty} \mathbb{E}[\nu_\ell' \nu_\ell]^{1/2}$ converges, and Assumption 8 is satisfied. □

A.2. Proofs for Case 4: MAR(r,s) Aggregates

A.2.1. Proof of Lemma 2.2

If $\boldsymbol{d}_{j,k}^1(\theta) = 0$, both sides of (2.35) vanish and the identity holds trivially; assume henceforth $\boldsymbol{d}_{j,k}^1(\theta) \ne 0$. We introduce the angle $\upsilon_k \in [0,2\pi)$ satisfying $(\cos\upsilon_k, \sin\upsilon_k) = \boldsymbol{d}_{j,k}^1/\|\boldsymbol{d}_{j,k}^1\|$ and consider the rotation

$$s = \frac{d_{j,k}u + d_{j,k-1}v}{\|\boldsymbol{d}_{j,k}^1\|}, \qquad t = \frac{-d_{j,k-1}u + d_{j,k}v}{\|\boldsymbol{d}_{j,k}^1\|}.$$

[20] $\alpha_\bullet(\cdot)$ with an argument always denotes a mixing coefficient.

This transformation corresponds to a rotation by $-\upsilon_k$, hence has unit Jacobian determinant. The Euclidean norm is preserved, $s^2+t^2=u^2+v^2$, and the linear form simplifies to $\Delta_{j,k}=\|\boldsymbol{d}^1_{j,k}\|s$. Consequently,

$$\iint_{\mathbb{R}^2} F(|\Delta_{j,k}(u,v;\theta)|)\,\frac{\kappa}{\pi}e^{-\kappa(u^2+v^2)}\,du\,dv=\iint_{\mathbb{R}^2} F(\|\boldsymbol{d}^1_{j,k}\||s|)\,\frac{\kappa}{\pi}e^{-\kappa(s^2+t^2)}\,ds\,dt.$$

Specializing to $F(x)=x^\alpha$ and factoring the resulting product of one-dimensional integrals yields (2.35). The constant $C_{w,\alpha}$ is finite provided $\alpha>-1$, which is satisfied since $\alpha\in(1,2)$. □

A.2.2. Proof of Lemma 2.3

Throughout this proof, we apply the rotation from Lemma 2.2 to each index $k\in\mathbb{Z}$ with $\boldsymbol{d}^1_{j,k}(\theta)\neq 0$.

Preliminary bounds on $\|\boldsymbol{d}^1_{j,k}\|$. By the definition $\boldsymbol{d}^1_{j,k}=(d_{j,k},d_{j,k-1})$ and the decay bound (2.27), we have

$$\|\boldsymbol{d}^1_{j,k}\|=\sqrt{d^2_{j,k}+d^2_{j,k-1}}\le\sqrt{2}\max(|d_{j,k}|,|d_{j,k-1}|)\le\sqrt{2}\,C_0\,\rho^{|k|-1}=\sqrt{2}\,\bar{C}_0\rho^{|k|}. \tag{A.21}$$

The corresponding lower bound

$$\|\boldsymbol{d}^1_{j,k}\|\ge c_v\,\rho_1^{|k|},\qquad k\in\mathbb{Z}, \tag{A.22}$$

is provided by Assumption 9; as noted there, distinct roots alone guarantee only that the partial-fraction coefficients $A_{j,l}$, $B_{j,i}$ are nonzero, which does not by itself yield a pointwise-in-k geometric lower bound when the dominant moduli on the two sides differ or when complex-conjugate root pairs occur.

Bounds on the derivatives of the partial-fraction weights. We now justify the constants C_1,C_2 in (2.30). Write $A_{j,l}=(-1)^{r_j}N/D$ with $N=\zeta^{s_j-1}_{j,l}$ and D the product of the r_j+s_j-1 factors making up its denominator. Every such factor is bounded by 2 in modulus, since $|\zeta_{j,l}-\zeta_{j,m}|\le 2\rho$ and $|\lambda_{j,i}\zeta_{j,l}-1|\le 1+\rho^2$, and by 2 again for its first two derivatives, with equality only when the coordinate being differentiated is the imaginary part of a conjugate pair, the one case where a single coordinate displaces both roots at once. A given coordinate enters at most $\bar{n}-1$ of the r_j+s_j-1 factors, so differentiating D term by term gives $|\partial D|\le\bar{n}2^{\bar{n}-1}$ and $|\partial^2 D|\le\bar{n}^2 2^{\bar{n}-1}$; together with the elementary bounds $|\partial N|\le\bar{q}$, $|\partial^2 N|\le\bar{q}^2$ on the numerator, the quotient rule $\partial A=\partial N/D-N\partial D/D^2$, and its second-order counterpart, turns into $|\partial A_{j,l}|\le\mathcal{A}_1$ and $\|\partial^2 A_{j,l}\|\le\mathcal{A}_2$. The rest of (2.30) comes from differentiating the power $\zeta^k_{j,l}$ rather than the weight. A real coordinate moves one root if it parametrizes a real root, two if it is the real or imaginary part of a conjugate pair; hence the factors 2 and 4, combined with $|k|\rho^{|k|-1}\le\rho^{-1}(1+|k|)\rho^{|k|}$ and $|k(k-1)|\rho^{|k|-2}\le 2\rho^{-2}(1+k^2)\rho^{|k|}$. The bounds on $B_{j,i}$ are the mirror image, causal and noncausal sides swapped.

Derivatives with respect to the full parameter vector. The formulas displayed in parts $(\iota\iota)$–$(\iota\iota\iota)$ below treat the coordinates of θ acting through the MA coefficients $d_{j,k}(\theta_j)$, that is, the roots collected in θ_j. The full parameter vector also contains the tail index α and the scaled weights ς_j; both are quantified here, the former first. By (2.32) and Assumption 9, $c_v\rho_1^{|k|}\le\|\boldsymbol{d}^1_{j,k}\|\le\sqrt{2}\bar{C}_0\rho^{|k|}$, so that

$$\big|\ln\|\boldsymbol{d}^1_{j,k}\|\big|\le\Lambda_0+\Lambda_1|k|,\qquad\Lambda_0:=\max\big\{\big|\ln(\sqrt{2}\bar{C}_0)\big|,\ |\ln c_v|\big\},\qquad\Lambda_1:=\ln(1/\rho_1), \tag{A.23}$$

where $\rho_1\le\rho$ has been used. Write

$$L_{w,\alpha}:=\frac{\kappa}{\pi}\iint|s|^\alpha\big|\ln|s|\big|e^{-\kappa(s^2+t^2)}\,ds\,dt,\qquad Q_{w,\alpha}:=\frac{\kappa}{\pi}\iint|s|^\alpha\big(\ln|s|\big)^2e^{-\kappa(s^2+t^2)}\,ds\,dt,$$

both finite for $\alpha\in(1,2)$, with the explicit bounds $L_{w,\alpha}\le e^{-1}(C_{w,\alpha+1}+C_{w,\alpha-1})$ and $Q_{w,\alpha}\le 8e^{-2}(C_{w,\alpha+1}+C_{w,\alpha-1})$ obtained from $|\ln x|\le(e\varepsilon)^{-1}(x^\varepsilon+x^{-\varepsilon})$ with $\varepsilon=1$ and $\varepsilon=1/2$ respectively. Since the rotation of Lemma 2.2 gives $|\Delta_{j,k}|=\|\boldsymbol{d}^1_{j,k}\|\,|s|$, the first-order term $\partial_\alpha|\Delta_{j,k}|^\alpha=|\Delta_{j,k}|^\alpha\ln|\Delta_{j,k}|$ satisfies

$$\iint\big|\partial_\alpha|\Delta_{j,k}|^\alpha\big|\,w\,du\,dv\le\widetilde{C}^A_1\,(1+|k|)\,\rho^{\alpha|k|},\quad\widetilde{C}^A_1:=\big(\sqrt{2}\bar{C}_0\big)^\alpha\max\big\{\Lambda_0C_{w,\alpha}+L_{w,\alpha},\ \Lambda_1C_{w,\alpha}\big\}, \tag{A.24}$$

and the second-order term $\partial^2_{\alpha\alpha}|\Delta_{j,k}|^\alpha = |\Delta_{j,k}|^\alpha(\ln|\Delta_{j,k}|)^2$ satisfies, expanding $(\ln\|\boldsymbol{d}^1_{j,k}\| + \ln|s|)^2$ and using $(\Lambda_0 + \Lambda_1|k|)^2 \le 2\max(\Lambda_0,\Lambda_1)^2(1+k^2)$ together with $\Lambda_0 + \Lambda_1|k| \le 2(\Lambda_0+\Lambda_1)(1+k^2)$,

$$\iint \big|\partial^2_{\alpha\alpha}|\Delta_{j,k}|^\alpha\big|\, w\, du\, dv \le \widetilde{C}^A_2\,(1+k^2)\,\rho^{\alpha|k|}, \tag{A.25}$$

$$\widetilde{C}^A_2 := \big(\sqrt{2}\bar{C}_0\big)^\alpha\Big[2\max(\Lambda_0,\Lambda_1)^2 C_{w,\alpha} + 4(\Lambda_0+\Lambda_1)L_{w,\alpha} + Q_{w,\alpha}\Big].$$

For the mixed derivative $\partial_\theta\partial_\alpha|\Delta_{j,k}|^\alpha = |\Delta_{j,k}|^{\alpha-1}\mathrm{sign}(\Delta_{j,k})\,\partial_\theta\Delta_{j,k}\,\big(\alpha\ln|\Delta_{j,k}|+1\big)$ we work in the rotated polar frame, where $|\Delta_{j,k}| = \|\boldsymbol{d}^1_{j,k}\|\,r|\cos\chi|$ and hence $|\ln|\Delta_{j,k}|| \le |\ln\|\boldsymbol{d}^1_{j,k}\|| + |\ln r| + |\ln|\cos\chi||$. The total radial power is $r^{\alpha+1}$ and the angular exponent is $\alpha - 1 > 0$, so that, with $R_\beta := \frac{\kappa}{\pi}\int_0^\infty r^{\beta+1}e^{-\kappa r^2}dr$, $A_\gamma := \int_0^{2\pi}|\cos\phi|^\gamma d\phi$ and their logarithmic counterparts $R^{\log}_\beta$, $A^{\log}_\gamma$,

$$\iint \big|\partial_\theta\partial_\alpha|\Delta_{j,k}|^\alpha\big|\, w\, du\, dv \le \widetilde{C}^{RA}_2\,(1+|k|)^2\,\rho^{\alpha|k|}, \tag{A.26}$$

$$\widetilde{C}^{RA}_2 := 2\bar{C}_1\big(\sqrt{2}\bar{C}_0\big)^{\alpha-1}\Big[\big(\alpha\max(\Lambda_0,\Lambda_1)+1\big)R_\alpha A_{\alpha-1} + \alpha R^{\log}_\alpha A_{\alpha-1} + \alpha R_\alpha A^{\log}_{\alpha-1}\Big],$$

with $R^{\log}_\alpha \le e^{-1}(R_{\alpha+1} + R_{\alpha-1})$ and $A^{\log}_{\alpha-1} \le e^{-1}(A_\alpha + I_a(\alpha))$. It is essential to keep the logarithm inside the polar decomposition rather than bound it by a power of $|\Delta_{j,k}|$: the estimate $|\Delta|^{\alpha-1}|\ln|\Delta|| \le e^{-1}(|\Delta|^\alpha + |\Delta|^{\alpha-2})$ would reintroduce the singular exponent $\alpha - 2$ and degrade the geometric rate from $\rho^{\alpha|k|}$ to $\rho^{|k|}\rho_1^{(\alpha-2)|k|}$, which need not be summable under Assumption 9. The scale parameters, in contrast, require no new estimate. Since $\Delta_{j,k}(\varsigma_j u, \varsigma_j v) = \varsigma_j\Delta_{j,k}(u,v)$, the scaled log-characteristic function factorises as

$$\log\varphi_{X_j}(\varsigma_j u, \varsigma_j v;\theta_j) = -\,\varsigma_j^\alpha\,S_j(u,v;\theta_j), \qquad S_j(u,v;\theta_j) := \sum_{k\in\mathbb{Z}}\big|\Delta_{j,k}(u,v;\theta_j)\big|^\alpha, \tag{A.27}$$

so that $\log\varphi_{X_j}$ is positively homogeneous of degree α in ς_j. Differentiating (A.27) with respect to ς_j therefore multiplies the tail of S_j by $\alpha\varsigma_j^{\alpha-1}$ and creates no new dependence on (u,v): the ς_j-derivatives reproduce the order-zero envelope, and the mixed ς_j–root and ς_j–α derivatives reproduce the order-one envelope. Quantitatively, writing $V_\gamma := \sup_{\theta\in\Theta}\max_j\varsigma_j^\gamma$ and $\ell_\varsigma := \sup_{\theta\in\Theta}\max_j|\ln\varsigma_j|$, both finite on the compact Θ by Assumption 1, the bounds (2.36)–(2.38) carry over to the scaled arguments upon multiplication by V_α at order zero, by $\alpha V_{\alpha-1} + V_\alpha(1+\ell_\varsigma)$ at order one, and by $\alpha(\alpha-1)V_{\alpha-2} + 2\alpha V_{\alpha-1}(1+\ell_\varsigma) + V_\alpha(1+\ell_\varsigma)^2$ at order two. Summing over the J components multiplies each of these by J, which is where the dependence of the aggregate constants on J enters. The bounds below therefore apply to the full gradient and Hessian.

(ι) *Order 0, proof of* (2.36). Combining (2.35) with the upper bound (A.21) gives

$$\iint |\Delta_{j,k}|^\alpha w\, du\, dv = C_{w,\alpha}\,\|\boldsymbol{d}^1_{j,k}\|^\alpha \le C_{w,\alpha}(\sqrt{2}\bar{C}_0)^\alpha\,\rho^{\alpha|k|},$$

where the last step uses $\|\boldsymbol{d}^1_{j,k}\| \le \sqrt{2}\bar{C}_0\rho^{|k|}$ from (2.32), the constant $\bar{C}_0$ of (2.31) already accounting for the index shift in $\boldsymbol{d}^1_{j,k} = (d_{j,k}, d_{j,k-1})$. Summing over the tail $|k| > M$ and using the exact geometric sum $\sum_{|k|>M}\rho^{\alpha|k|} = 2\rho^{\alpha(M+1)}/(1-\rho^\alpha)$,

$$\iint \big|\log\varphi_{X_j} - \log\varphi^{(M)}_{X_j}\big|\, w\, du\, dv \;\le\; \mathfrak{c}_0\,\rho^{\alpha M}, \qquad \mathfrak{c}_0 := \frac{2\,\rho^\alpha\,C_{w,\alpha}\,\big(\sqrt{2}\,\bar{C}_0\big)^\alpha}{1-\rho^\alpha}, \tag{A.28}$$

which establishes (2.36).

$(\iota\iota)$ *Order 1, proof of* (2.37). Throughout parts $(\iota\iota)$–$(\iota\iota\iota)$, the identification of $\partial_\theta[\log\varphi_{X_j} - \log\varphi^{(M)}_{X_j}]$ with the termwise-differentiated series $-\sum_{|k|>M}\partial_\theta|\Delta_{j,k}|^\alpha$ (and likewise at second order) is justified once and for all: the differentiated series converges locally uniformly in θ, by the very envelopes $(1+|k|)\rho^{\alpha|k|}$ and $(1+k^2)\bar{\rho}^{|k|}$ established below, which are summable over k. For $\boldsymbol{d}^1_{j,k} \ne 0$, the function $\|\boldsymbol{d}^1_{j,k}\|^\alpha$ is twice continuously differentiable in θ with gradient

$$\partial_\theta\|\boldsymbol{d}^1_{j,k}\|^\alpha = \alpha\,\|\boldsymbol{d}^1_{j,k}\|^{\alpha-2}(\boldsymbol{d}^1_{j,k}\cdot\partial_\theta\boldsymbol{d}^1_{j,k}).$$

The integrand $|\Delta_{j,k}|^\alpha$ is absolutely continuous in θ with derivative bounded by an integrable envelope in (u,v), so dominated convergence justifies differentiation under the integral:

$$\iint \partial_\theta |\Delta_{j,k}|^\alpha \, w \, du \, dv = C_{w,\alpha} \, \partial_\theta \|\boldsymbol{d}^1_{j,k}\|^\alpha.$$

For the absolute value, a direct computation yields

$$\big|\partial_\theta |\Delta_{j,k}|^\alpha\big| \le \alpha \, |\Delta_{j,k}|^{\alpha-1} \big(|u||\partial_\theta d_{j,k}| + |v||\partial_\theta d_{j,k-1}|\big).$$

Since $\alpha - 1 \in (0,1)$, the map $x \mapsto x^{\alpha-1}$ is subadditive on $\mathbb{R}_+$, giving

$$|\Delta_{j,k}|^{\alpha-1} \le (|u||d_{j,k}|)^{\alpha-1} + (|v||d_{j,k-1}|)^{\alpha-1}.$$

Applying (2.32) to both factors,

$$\big|\partial_\theta |\Delta_{j,k}|^\alpha\big| \le \alpha \, \bar{C}_0^{\alpha-1} \bar{C}_1 (1+|k|) \rho^{\alpha|k|} \big(|u|^{\alpha-1} + |v|^{\alpha-1}\big)\big(|u| + |v|\big).$$

Expanding the product generates four terms. The two diagonal ones are $|u|^\alpha$ and $|v|^\alpha$; the two cross terms are controlled by Young's inequality $|u|^{\alpha-1}|v| \le \frac{\alpha-1}{\alpha}|u|^\alpha + \frac{1}{\alpha}|v|^\alpha$, applied once in each direction, and contribute at most $|u|^\alpha + |v|^\alpha$ in total. Hence

$$\big|\partial_\theta |\Delta_{j,k}|^\alpha\big| \le \widetilde{C}_1^R \, (1+|k|) \, \rho^{\alpha|k|} \big(|u|^\alpha + |v|^\alpha\big), \qquad \widetilde{C}_1^R := 2\alpha \, \bar{C}_0^{\alpha-1} \bar{C}_1, \tag{A.29}$$

the constant 2 being sharp, as the two sides of the inequality agree in the limit $|u| = |v|$. Since $\iint (|u|^\alpha + |v|^\alpha) w \, du \, dv = 2C_{w,\alpha}$, integration gives $2C_{w,\alpha}\widetilde{C}_1^R(1+|k|)\rho^{\alpha|k|}$ for the root-derivatives, while the derivative with respect to α obeys the same rate with the constant $\widetilde{C}_1^A$ of (A.24). Setting

$$\Xi_1 := \max\big\{2C_{w,\alpha}\widetilde{C}_1^R, \ \widetilde{C}_1^A\big\}$$

and using the two-sided geometric bound $\sum_{|k|>M}(1+|k|)q^{|k|} \le 4(1+M)q^{M+1}/(1-q)^2$ with $q = \rho^\alpha$, we obtain (2.37) with

$$\mathfrak{c}_1 := \frac{4 \, \rho^\alpha \, \Xi_1}{(1-\rho^\alpha)^2}. \tag{A.30}$$

($\iota\iota$) *Order 2, proof of* (2.38). Away from the singular set $\{\Delta_{j,k} = 0\}$, the second derivative reads

$$\begin{aligned}
\partial^2_{\theta\theta'} |\Delta_{j,k}|^\alpha &= \alpha(\alpha-1)|\Delta_{j,k}|^{\alpha-2}(u\partial_\theta d_{j,k} + v\partial_\theta d_{j,k-1})(u\partial_{\theta'} d_{j,k} + v\partial_{\theta'} d_{j,k-1}) \\
&\quad + \alpha|\Delta_{j,k}|^{\alpha-1}\mathrm{sign}(\Delta_{j,k})(u\partial^2_{\theta\theta'} d_{j,k} + v\partial^2_{\theta\theta'} d_{j,k-1}).
\end{aligned}$$

We bound $|\partial^2_{\theta\theta'}|\Delta_{j,k}|^\alpha|$ by $T_{1,k}(u,v) + T_{2,k}(u,v)$, where

$$\begin{aligned}
T_{1,k} &:= \alpha(\alpha-1)|\Delta_{j,k}|^{\alpha-2}(|u||\partial_\theta d_{j,k}| + |v||\partial_\theta d_{j,k-1}|)(|u||\partial_{\theta'} d_{j,k}| + |v||\partial_{\theta'} d_{j,k-1}|), \\
T_{2,k} &:= \alpha|\Delta_{j,k}|^{\alpha-1}(|u||\partial^2_{\theta\theta'} d_{j,k}| + |v||\partial^2_{\theta\theta'} d_{j,k-1}|).
\end{aligned}$$

Bound on $\iint T_{2,k} \, w \, du \, dv$. Applying the subadditivity argument from part (ι) to $|\Delta_{j,k}|^{\alpha-1}$ together with Young's inequality yields

$$T_{2,k}(u,v) \le \widetilde{C}_2^R(|u|^\alpha + |v|^\alpha)(1+k^2)\rho^{\alpha|k|}, \qquad \widetilde{C}_2^R := 2\alpha \, \bar{C}_0^{\alpha-1} \bar{C}_2, \tag{A.31}$$

by the chain of (A.29) with $\bar{C}_2(1+k^2)$ in place of $\bar{C}_1(1+|k|)$. Integrating against w gives $\iint T_{2,k} \, w \, du \, dv \le 2C_{w,\alpha}\widetilde{C}_2^R(1+k^2)\rho^{\alpha|k|}$.

Bound on $\iint T_{1,k}\, w\, du\, dv$. This term involves the $|\Delta_{j,k}|^{\alpha-2}$ singularity. We perform the rotation of Lemma 2.2 and pass to polar coordinates $(u,v) = r(\cos\phi, \sin\phi)$ in the rotated frame, so that $\Delta_{j,k} = \|\boldsymbol{d}^1_{j,k}\| r\cos(\phi - \phi_k^\perp)$ for a suitable angle $\phi_k^\perp$. This gives

$$|\Delta_{j,k}|^{\alpha-2} = r^{\alpha-2}\|\boldsymbol{d}^1_{j,k}\|^{\alpha-2}|\cos(\phi-\phi_k^\perp)|^{\alpha-2}.$$

Each of the two bracketed factors in $T_{1,k}$ satisfies

$$|u||\partial_\theta d_{j,k}| + |v||\partial_\theta d_{j,k-1}| \le r\big(|\partial_\theta d_{j,k}| + |\partial_\theta d_{j,k-1}|\big) \le 2\, r\, \bar{C}_1(1+|k|)\rho^{|k|},$$

since $|u|, |v| \le r$ in the rotated polar frame, together with (2.32). Combining these estimates and integrating,

$$\begin{aligned}\iint T_{1,k}\, w\, du\, dv \le 4\alpha(\alpha-1)\bar{C}_1^2(1+|k|)^2\rho^{2|k|}\|\boldsymbol{d}^1_{j,k}\|^{\alpha-2} \\ \times \frac{\kappa}{\pi}\int_0^\infty r^{\alpha+1}\, e^{-\kappa r^2}\, dr \cdot \int_0^{2\pi} |\cos(\phi-\phi_k^\perp)|^{\alpha-2}\, d\phi.\end{aligned}$$

The radial factor is the normalised integral $I_r(\alpha,\kappa)$ of (A.5), namely $I_r(\alpha,\kappa) = \Gamma\big(\frac{\alpha+2}{2}\big)/(2\pi\kappa^{\alpha/2}) < \infty$. The angular integral $I_a(\alpha) := \int_0^{2\pi}|\cos\psi|^{\alpha-2}\, d\psi = 2\sqrt{\pi}\,\Gamma\big(\frac{\alpha-1}{2}\big)/\Gamma\big(\frac{\alpha}{2}\big)$ is finite because $\alpha - 2 > -1$ (equivalently, $\alpha > 1$) ensures integrability near the zeros of the cosine. Importantly, $I_a(\alpha)$ depends neither on k, j, nor θ. Thus,

$$\iint T_{1,k}\, w\, du\, dv \le 4\alpha(\alpha-1)\bar{C}_1^2\, I_r(\alpha,\kappa)\, I_a(\alpha)\,(1+|k|)^2\rho^{2|k|}\|\boldsymbol{d}^1_{j,k}\|^{\alpha-2}.$$

By the lower bound (A.22) of Assumption 9, $\|\boldsymbol{d}^1_{j,k}\|^{\alpha-2} \le c_v^{\alpha-2}\rho_1^{(\alpha-2)|k|}$, where the inequality reverses because $\alpha - 2 < 0$. Substituting, and recalling $\rho_1^{(\alpha-2)|k|} \le \rho_1^{(\alpha-2)|k|}$ together with the definition $\bar{\rho} = \rho^2\rho_1^{\alpha-2}$,

$$\iint T_{1,k}\, w\, du\, dv \le \widehat{C}_1(1+|k|)^2\rho^{2|k|}\rho_1^{(\alpha-2)|k|} \;\le\; \widehat{C}_1(1+|k|)^2\,\bar{\rho}^{\,|k|},$$

with

$$\widehat{C}_1 := 4\alpha(\alpha-1)\,\bar{C}_1^2\, I_r(\alpha,\kappa)\, I_a(\alpha)\, c_v^{\alpha-2}. \tag{A.32}$$

It remains to collect the four second-order contributions: the singular root term (A.32), the regular root term (A.31), the pure α term (A.25) and the mixed term (A.26). Using $(1+|k|)^2 \le 2(1+k^2)$ for the first and last of these, and $\rho^{\alpha|k|} \le \bar{\rho}^{\,|k|}$ for the second and third (recall $\bar{\rho} \ge \rho^\alpha$), the per-lag bound is $\Xi_2(1+k^2)\bar{\rho}^{\,|k|}$ with

$$\Xi_2 := 2\widehat{C}_1 + 2C_{w,\alpha}\widetilde{C}_2^R + \widetilde{C}_2^A + 2\widetilde{C}_2^{RA}.$$

Summing over $|k| > M$ with the two-sided geometric bound $\sum_{|k|>M}(1+k^2)q^{|k|} \le 24(1+M)^2 q^{M+1}/(1-q)^3$, applied with $q = \bar{\rho} < 1$ by the rate condition of Assumption 9, yields (2.38) with

$$\mathfrak{c}_2 := \frac{24\,\bar{\rho}\,\Xi_2}{(1-\bar{\rho})^3}. \tag{A.33}$$

□

A.2.3. Proof of Lemma 2.4

(ι) The truncated log-characteristic function (2.33) is a finite sum of $2M+1$ terms, each of the form $|ud_{j,k} + vd_{j,k-1}|^\alpha$. The coefficients $d_{j,k}$ are C^∞ rational functions of θ in a neighbourhood of Θ by the partial-fraction decomposition (2.26). For each k and θ, the singular locus $\{ud_{j,k} + vd_{j,k-1} = 0\}$ is a line through the origin in the (u,v)-plane, and the polar-coordinate reduction employed in the proof of Lemma 2.1 applies term by term. This yields w-integrable dominating functions for the first derivatives, uniform over θ on a neighbourhood, so that $D_{\mathcal{X}}^{(M)} \in C^1(\Theta)$ by dominated convergence. At second order no such locally uniform envelope exists: the singular line $\{ud_{j,k}+vd_{j,k-1}=0\}$ moves with θ, so $\sup_{\theta\in V}|\Delta_{j,k}|^{\alpha-2}=+\infty$ on a cone of positive measure, and the finiteness of the sum does not help. We invoke instead Lemma A.1: Lemma A.1-(ι) holds term by term as in the proof of Lemma 2.1; Lemma A.1-($\iota\iota$) is the C^1 step; and Lemma A.1-($\iota\iota\iota$) holds because each of the $2M+1$ terms satisfies (A.8), by the argument of Appendix A.1.3 applied to $|ud_{j,k}+vd_{j,k-1}|^{\alpha}$ in place of $|\psi_k u+v|^{\alpha}$, and a finite sum of $L^1(w)$-continuous maps is $L^1(w)$-continuous. Hence $D_{\mathcal{X}}^{(M)}\in C^2(\Theta)$, establishing Assumption 3 for the truncated criterion.

Assumption 6 follows from the same considerations: boundedness of the score is inherited from $|\varphi_{\mathcal{X}}^{(M)}|\le 1$ and the integrable envelopes $(1+|k|)\rho^{\alpha|k|}$ and $(1+k^2)\bar\rho^{|k|}$ of Lemma 2.3. Assumption 8, being a projection condition on the sequence $\{K_t\}$ rather than a bound on the score, follows instead from the mixing argument of the validation of Assumption 8, which extends to MAR(r_j,s_j) components: the transfer function $a_j(z)=\big(\Phi_j(z)\Psi_j(z^{-1})\big)^{-1}$ has inverse $a_j^{-1}(z)=\Phi_j(z)\Psi_j(z^{-1})$, again a Laurent polynomial with no zero on $|z|=1$, so condition (4.4.4) of Rosenblatt (2000) holds, while the geometric decay of the moving-average coefficients required by (4.4.3) is (2.27), which gives $|a_{j,k}|\le C_0\rho^{|k|}$, i.e. a decay rate equal to ρ for every component. As for the integrability and continuity clauses of Assumption 7, the continuity of the integrated Hessian follows from the argument of Appendix A.1.3 applied to the finite sum. For the non-singularity of $\Sigma^{(M)}(\theta_0)$, we use a perturbation argument: by part ($\iota\iota\iota$) below, the truncated scores converge to the full scores in $L^2_w(\mathbb{R}^2)$ as $M\to\infty$, so that $\Sigma^{(M)}(\theta_0)\to\Sigma(\theta_0)$ entrywise. Since the determinant is a continuous function of the entries and $\Sigma(\theta_0)$ is nonsingular by assumption, there exists $M_0\ge 1$ such that $\det\Sigma^{(M)}(\theta_0)\neq 0$ for all $M\ge M_0$.

($\iota\iota$) Write $\varphi_j:=\varphi_{X_j}(\varsigma_j u,\varsigma_j v;\theta_j)$ and $\varphi_j^{(M)}:=\varphi_{X_j}^{(M)}(\varsigma_j u,\varsigma_j v;\theta_j)$. Since $|\varphi_n|,|\varphi_{\mathcal{X}}|,|\varphi_{\mathcal{X}}^{(M)}|\le 1$, the elementary inequality $|a^2-b^2|\le|a-b|(|a|+|b|)$ gives

$$|D_{\mathcal{X}}^{(M)}(\theta)-D_{\mathcal{X}}(\theta)|\le 4\iint|\varphi_{\mathcal{X}}-\varphi_{\mathcal{X}}^{(M)}|\,w\,du\,dv.$$

A telescoping argument on the product structure of $\varphi_{\mathcal{X}}$, combined with $|e^a-e^b|\le|a-b|$ for $\mathrm{Re}(a),\mathrm{Re}(b)\le 0$, yields

$$|\varphi_{\mathcal{X}}-\varphi_{\mathcal{X}}^{(M)}|\le\sum_{j=1}^{J}|\log\varphi_j-\log\varphi_j^{(M)}|.$$

Integrating and invoking (2.36) with the constant $\mathfrak{c}_0$ produces the order-0 estimate.

For the gradient, observe that $\varphi_{\mathcal{X}}^{(M)}=\exp(\mathcal{L}^{(M)})$ with $\mathcal{L}^{(M)}:=\sum_j\log\varphi_j^{(M)}$, so $\partial_\theta\varphi_{\mathcal{X}}^{(M)}=\varphi_{\mathcal{X}}^{(M)}\partial_\theta\mathcal{L}^{(M)}$. The difference decomposes as

$$\partial_\theta\varphi_{\mathcal{X}}^{(M)}-\partial_\theta\varphi_{\mathcal{X}}=\varphi_{\mathcal{X}}^{(M)}(\partial_\theta\mathcal{L}^{(M)}-\partial_\theta\mathcal{L})+(\varphi_{\mathcal{X}}^{(M)}-\varphi_{\mathcal{X}})\partial_\theta\mathcal{L}.$$

The first term on the right is controlled by (2.37) with constant $\mathfrak{c}_1$. For the second, we use the pointwise bound $|\partial_\theta\mathcal{L}|\le C\big(1+|u|^\alpha+|v|^\alpha\big)\big(1+\ln(1+|u|+|v|)\big)$, which follows from (A.29) for the root coordinates and from $\partial_\alpha|\Delta_{j,k}|^\alpha=|\Delta_{j,k}|^\alpha\ln|\Delta_{j,k}|$ for the tail index, summed over k, together with $|\varphi_{\mathcal{X}}^{(M)}-\varphi_{\mathcal{X}}|\le C'(|u|^\alpha+|v|^\alpha)\rho^{\alpha M}$ for a finite constant C' depending on J and $\mathfrak{c}_0$. Since this envelope is square-, and indeed fourth-power-, integrable against the Gaussian weight w, we obtain $\sup_{\theta\in\Theta}\|G^{(M)}(\theta)-G(\theta)\|=O((1+M)\rho^{\alpha M})$.

The Hessian satisfies

$$\partial^2_{\theta\theta'}\varphi_{\mathcal{X}}^{(M)}=\varphi_{\mathcal{X}}^{(M)}(\partial^2_{\theta\theta'}\mathcal{L}^{(M)}+\partial_\theta\mathcal{L}^{(M)}\partial_{\theta'}\mathcal{L}^{(M)}).$$

The difference $\partial^2_{\theta\theta'}\varphi_{\mathcal{X}}^{(M)}-\partial^2_{\theta\theta'}\varphi_{\mathcal{X}}$ splits into contributions involving $\partial^2_{\theta\theta'}\mathcal{L}^{(M)}-\partial^2_{\theta\theta'}\mathcal{L}$ (controlled by (2.38) with constant K_2, at rate $(1+M)^2\bar\rho^M$) and products of terms already bounded (at rates dominated by $\bar\rho^M$ since $\rho^\alpha\le\bar\rho$). Assembling these estimates yields $\sup_{\theta\in\Theta}\|H^{(M)}(\theta)-H(\theta)\|=O((1+M)^2\bar\rho^M)$.

$(\iota\iota\iota)$ The same decomposition, squared and integrated rather than inserted into $\nabla D_{\mathcal{X}}$, gives the L^2_w statement. Summing (A.29) over the discarded lags yields the pointwise bound

$$\left|\partial_\theta \mathcal{L}^{(M)} - \partial_\theta \mathcal{L}\right| \;\le\; C''\,(1+M)\,\rho^{\alpha M}\big(|u|^\alpha + |v|^\alpha\big),$$

with C'' finite and independent of θ, so the first term of the decomposition contributes $O\big((1+M)^2\rho^{2\alpha M}\big)$ after squaring and integrating, because $\iint(|u|^\alpha+|v|^\alpha)^2 w\,du\,dv < \infty$ and $|\varphi^{(M)}_{\mathcal{X}}| \le 1$. The second term is bounded by $|\varphi^{(M)}_{\mathcal{X}} - \varphi_{\mathcal{X}}|\,|\partial_\theta\mathcal{L}| \le C'C(|u|^\alpha+|v|^\alpha)^2\rho^{\alpha M}$, whose square is w-integrable since $\iint(|u|^\alpha+|v|^\alpha)^4 w\,du\,dv < \infty$, and contributes $O(\rho^{2\alpha M})$. Adding the two gives $(\iota\iota\iota)$.

$(\iota\nu)$ Regularity of the untruncated criterion. By (ι) each $D^{(M)}_{\mathcal{X}}$ is of class $C^2(\Theta)$, with $\nabla D^{(M)}_{\mathcal{X}} = G^{(M)}$ and $\nabla^2 D^{(M)}_{\mathcal{X}} = H^{(M)}$, and by $(\iota\iota)$ we have $D^{(M)}_{\mathcal{X}} \to D_{\mathcal{X}}$ pointwise while $G^{(M)} \to G$ and $H^{(M)} \to H$ uniformly on Θ. Applying twice the classical theorem on the differentiation of uniform limits (on each open ball contained in Θ, which is convex) yields $D_{\mathcal{X}} \in C^2(\Theta)$ with $\nabla D_{\mathcal{X}} = G$ and $\nabla^2 D_{\mathcal{X}} = H$. Assumption 3 therefore holds for the untruncated criterion, and Lemma A.1 is not needed for it: it was needed for each fixed M, in (ι), and the passage to the limit is now a matter of uniform convergence. Assumption 7. Decomposing

$$\partial^2_{\theta\theta'}\varphi^{(M)}_{\mathcal{X}} - \partial^2_{\theta\theta'}\varphi_{\mathcal{X}} = \big(\varphi^{(M)}_{\mathcal{X}} - \varphi_{\mathcal{X}}\big)\big(\partial^2_{\theta\theta'}\mathcal{L}^{(M)} + \partial_\theta\mathcal{L}^{(M)}\partial_{\theta'}\mathcal{L}^{(M)}\big) + \varphi_{\mathcal{X}}\big(\partial^2_{\theta\theta'}\mathcal{L}^{(M)} - \partial^2_{\theta\theta'}\mathcal{L}\big) + \varphi_{\mathcal{X}}\big(\partial_\theta\mathcal{L}^{(M)}\partial_{\theta'}\mathcal{L}^{(M)} - \partial_\theta\mathcal{L}\,\partial_{\theta'}\mathcal{L}\big),$$

the middle term is bounded in $L^1(w)$ by (2.38); the third is bounded by $|\partial_\theta\mathcal{L}^{(M)} - \partial_\theta\mathcal{L}|\,(|\partial_{\theta'}\mathcal{L}^{(M)}| + |\partial_{\theta'}\mathcal{L}|)$ and controlled by the pointwise bounds of $(\iota\iota\iota)$, the Gaussian weight integrating the resulting product of two polynomial-times-logarithmic factors; and the first is bounded, using $|\varphi^{(M)}_{\mathcal{X}} - \varphi_{\mathcal{X}}| \le C'(|u|^\alpha+|v|^\alpha)\rho^{\alpha M}$, by

$$C'\rho^{\alpha M}\iint(|u|^\alpha+|v|^\alpha)\Big(\big\|\partial^2_{\theta\theta'}\mathcal{L}^{(M)}\big\| + \big|\partial_\theta\mathcal{L}^{(M)}\big|\,\big|\partial_{\theta'}\mathcal{L}^{(M)}\big|\Big)w\,du\,dv,$$

an integral finite uniformly in M and θ: the polar estimates of the proof of Lemma 2.3 give radial exponent $2\alpha+1$ and angular exponent $\alpha - 2 > -1$ for the singular part, and the summation over k converges by (2.38). Hence

$$\sup_{\theta\in\Theta}\iint\big\|\partial^2_{\theta\theta'}\varphi^{(M)}_{\mathcal{X}}(\cdot\,;\theta) - \partial^2_{\theta\theta'}\varphi_{\mathcal{X}}(\cdot\,;\theta)\big\|\,w\,du\,dv = O\big((1+M)^2\bar\rho^{M}\big) \longrightarrow 0. \tag{A.34}$$

For each M the map $\theta \mapsto \partial^2_{\theta\theta'}\varphi^{(M)}_{\mathcal{X}}(\cdot\,;\theta)$ is continuous into $L^1(w\,du\,dv)$ with finite norm (this is Lemma A.1-$(\iota\iota)$, verified term by term for the $2M+1$ summands in part (ι)) and a uniform limit of continuous functions being continuous, (A.34) transfers both properties to $\varphi_{\mathcal{X}}$. Also, for each M the map $\theta \mapsto \partial_\theta\varphi^{(M)}_{\mathcal{X}}(\cdot\,;\theta)$ is continuous into $L^2_w(\mathbb{R}^2)$, being a finite sum of terms dominated by the θ-uniform, w-square-integrable envelope (A.29); by $(\iota\iota\iota)$ the convergence $\partial_\theta\varphi^{(M)}_{\mathcal{X}} \to \partial_\theta\varphi_{\mathcal{X}}$ holds in L^2_w uniformly in θ, and a uniform limit of continuous maps is continuous. This is the property used by the first integral of the Hessian step of Appendix A.1.3, hence by Proposition 2.2 through (2.22). The integrability and continuity clauses of Assumption 7 therefore hold for the untruncated characteristic function. For Assumption 6, summing the pointwise envelope (A.29) over all lags gives

$$\big\|\partial_\theta \log\varphi_{\mathcal{X}}(u,v;\theta)\big\| \;\le\; J\,\widetilde{C}^R_1\Big(\sum_{k\in\mathbb{Z}}(1+|k|)\rho^{\alpha|k|}\Big)\big(|u|^\alpha+|v|^\alpha\big) \;=\; C\big(|u|^\alpha+|v|^\alpha\big),$$

uniformly in θ, whence $\|\partial_\theta\varphi_{\mathcal{X}}\| \le |\varphi_{\mathcal{X}}|\,C(|u|^\alpha+|v|^\alpha) \in L^1(w)$ and, by (2.17), $\sup_{x,\theta}\|K(x;\theta)\| \le 2\iint\|\partial_\theta\varphi_{\mathcal{X}}\|\,w < \infty$. Assumption 8 holds by the mixing argument of the validation of Assumption 8, extended to $\mathrm{MAR}(r_j,s_j)$ components as indicated there. □

A.2.4. Proof of Proposition 2.2

Consistency. By Lemma 2.4$(\iota\iota)$, $\sup_{\theta\in\Theta}|D^{(M_n)}_{\mathcal{X}}(\theta) - D_{\mathcal{X}}(\theta)| = O(\rho^{\alpha M_n}) = o(1)$, so $D^{(M_n)}_{\mathcal{X}}$ converges uniformly to $D_{\mathcal{X}}$ on Θ (a deterministic approximation). Moreover, by the ergodic theorem (Assumption 4), $\varphi_n(u,v) \to \varphi_{\mathcal{X}}(u,v)$ almost surely for each (u, v), and the consistency part of Theorem 2.1 of Knight and Yu (2002), which holds under their conditions (A1)–(A4) and (A8), verified here, delivers the uniform convergence in probability of $D_{\mathcal{X}}$ to the population objective D_0 over Θ. Combining the two uniform convergences with identifiability (Assumption 5) implies $\hat{\theta}_n^{(M_n)} \to \theta_0$ in probability.

Asymptotic normality. The choice (2.39) gives $\bar{\rho}^{M_n} \le n^{-c_\star \log(1/\bar{\rho})}$ and

$$(1 + M_n)^2 \bar{\rho}^{M_n} = O\big((\log n)^2 n^{-c_\star \log(1/\bar{\rho})}\big) = o(n^{-1/2}),$$

since $c_\star \log(1/\bar{\rho}) > 1/2$; because $\rho^\alpha \le \bar{\rho}$, the same choice also yields $(1 + M_n)\rho^{\alpha M_n} = o(n^{-1/2})$. A first-order Taylor expansion of the gradient condition $\nabla D_{\mathcal{X}}^{(M_n)}(\hat{\theta}_n^{(M_n)}) = 0$ around θ_0 gives

$$0 = \nabla D_{\mathcal{X}}^{(M_n)}(\theta_0) + \nabla^2 D_{\mathcal{X}}^{(M_n)}(\bar{\theta}_n)(\hat{\theta}_n^{(M_n)} - \theta_0),$$

where the intermediate points $\bar{\theta}_n^{(i)}$, $i = 1, \ldots, p$, one per coordinate of the gradient, lie on the segment joining θ_0 and $\hat{\theta}_n^{(M_n)}$ (the mean value theorem applying row by row for a vector-valued map); each converges to θ_0, so the limit $2\Sigma(\theta_0)$ below is obtained for every row. Lemma 2.4($\iota\iota$) yields

$$\begin{aligned}
\nabla D_{\mathcal{X}}^{(M_n)}(\theta_0) &= \nabla D_{\mathcal{X}}(\theta_0) + R_n^{(1)}, \quad |R_n^{(1)}| = O((1 + M_n)\rho^{\alpha M_n}) = o(n^{-1/2}),\\
\nabla^2 D_{\mathcal{X}}^{(M_n)}(\bar{\theta}_n) &= \nabla^2 D_{\mathcal{X}}(\bar{\theta}_n) + R_n^{(2)}, \quad \sup_{\theta \in \Theta} |R_n^{(2)}| = O((1 + M_n)^2 \bar{\rho}^{M_n}) = o(1).
\end{aligned}$$

Consistency and continuity ensure $\nabla^2 D_{\mathcal{X}}(\bar{\theta}_n) \to 2\,\Sigma(\theta_0)$ in probability, by (2.22) of Remark 2.1 applied along the sequence $\bar{\theta}_n \to \theta_0$. Substituting into the Taylor expansion and rearranging,

$$\sqrt{n}(\hat{\theta}_n^{(M_n)} - \theta_0) = -\tfrac{1}{2}\,\Sigma(\theta_0)^{-1}\sqrt{n}\,\nabla D_{\mathcal{X}}(\theta_0) + o_P(1) \;=\; \Sigma(\theta_0)^{-1}\sqrt{n}\,\bar{K}_n + o_P(1),$$

the last equality by the score identity (2.21). Since $\sqrt{n}\,\bar{K}_n \xrightarrow{d} N(0, \Omega(\theta_0))$ by the central limit theorem of Theorem 2.1 of Knight and Yu (2002), whose Assumptions (A5)–(A7) hold here by Assumptions 6–8, the conclusion follows.
□

A.3. Proof of Lemma 3.1

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 $(\Gamma_j, \boldsymbol{\mu}_j^0)$. 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 $a = 2/\pi$

$$\begin{aligned}
\mathbb{E}\Big[e^{i\langle \boldsymbol{u}, \boldsymbol{X_t}\rangle}\Big] &= \mathbb{E}\Big[e^{i\langle \boldsymbol{u}, \sigma \sum_{j=1}^J \pi_j \boldsymbol{X}_{j,t}\rangle}\Big] = \prod_{j=1}^J \mathbb{E}\Big[e^{i\langle \sigma\pi_j \boldsymbol{u}, \boldsymbol{X}_{j,t}\rangle}\Big]\\
&= \prod_{j=1}^J \exp\left\{ -\int_{S_{m+h+1}} \Big(|\langle \sigma\pi_j \boldsymbol{u}, \boldsymbol{s}\rangle| + ia\langle \sigma\pi_j \boldsymbol{u}, \boldsymbol{s}\rangle \ln|\langle \sigma\pi_j \boldsymbol{u}, \boldsymbol{s}\rangle|\Big)\Gamma_j(d\boldsymbol{s}) + i\langle \sigma\pi_j \boldsymbol{u}, \boldsymbol{\mu_j}^0\rangle \right\}\\
&= \exp\left\{ -\int_{S_{m+h+1}} \Big(|\langle \boldsymbol{u}, \boldsymbol{s}\rangle| + ia\langle \boldsymbol{u}, \boldsymbol{s}\rangle \ln|\langle \boldsymbol{u}, \boldsymbol{s}\rangle|\Big) \sum_{j=1}^J \sigma^\alpha \pi_j^\alpha \Gamma_j(d\boldsymbol{s}) \right.\\
&\qquad \left. + i\sum_{j=1}^J \left(\langle \boldsymbol{u}, \sigma\pi_j \boldsymbol{\mu_j}^0\rangle - a\sigma\pi_j \ln(\sigma\pi_j) \int_{S_{m+h+1}} \langle \boldsymbol{u}, \boldsymbol{s}\rangle \Gamma_j(d\boldsymbol{s})\right) \right\}.
\end{aligned}$$

Focusing on the shift vector, we have

$$\sum_{j=1}^J \left(\langle \boldsymbol{u}, \sigma\pi_j \boldsymbol{\mu_j}^0\rangle - a\sigma\pi_j \ln(\sigma\pi_j) \int_{S_{m+h+1}} \langle \boldsymbol{u}, \boldsymbol{s}\rangle \Gamma_j(d\boldsymbol{s})\right) = \langle \boldsymbol{u}, \sum_{j=1}^J \sigma\pi_j(\boldsymbol{\mu}_j^0 - a\ln(\sigma\pi_j)\tilde{\boldsymbol{\mu}}_j)\rangle,$$

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, 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 (3.5),

$$\begin{aligned}
\sum_{j=1}^J \sigma\pi_j(\boldsymbol{\mu}_j^0 - a\ln(\sigma\pi_j)\tilde{\boldsymbol{\mu}}_j) &= \sum_{j=1}^J \sigma\pi_j\Big(- a\beta_j \sum_{k\in\mathbb{Z}} \boldsymbol{d}_{j,k} \ln\|\boldsymbol{d}_{j,k}\|_e - a\ln(\sigma\pi_j)\beta_j \sum_{k\in\mathbb{Z}} \boldsymbol{d}_{j,k}\Big) \\
&= -a\sum_{j=1}^J \sum_{k\in\mathbb{Z}} \sigma\pi_j\beta_j \boldsymbol{d}_{j,k}\ln\|\sigma\pi_j\boldsymbol{d}_{j,k}\|_e \\
&:= \boldsymbol{\mu}^0.
\end{aligned}$$

Therefore,

$$\mathbb{E}\Big[e^{i\langle \boldsymbol{u}, \boldsymbol{X}_t\rangle}\Big] = \exp\Bigg\{ - \int_{S_{m+h+1}} \Big(|\langle \boldsymbol{u},\boldsymbol{s}\rangle| + ia\langle \boldsymbol{u},\boldsymbol{s}\rangle \ln|\langle \boldsymbol{u},\boldsymbol{s}\rangle|\Big) \sum_{j=1}^J \sigma^\alpha \pi_j^\alpha \Gamma_j(d\boldsymbol{s}) + i\langle \boldsymbol{u}, \boldsymbol{\mu}^0\rangle\Bigg\},$$

and the random vector $\boldsymbol{X}_t$ is 1-stable with spectral measure

$$\sum_{j=1}^J \sigma^\alpha\pi_j^\alpha\Gamma_j = \sigma^\alpha \sum_{j=1}^J \sum_{\vartheta\in S_1}\sum_{k\in\mathbb{Z}} w_{j,\vartheta}\pi_j^\alpha \|\boldsymbol{d}_{j,k}\|_e^\alpha \delta_{\left\{\frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|_e}\right\}},$$

and shift vector as announced in the lemma.

A.4. Proof of Lemma 3.2

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, $j = 1, \ldots, J$, their spectral measures on the Euclidean unit sphere. Let Γ be the spectral measure of $\boldsymbol{X}_t$. By Lemma 3.1, we have:

$$\Gamma = \sigma^\alpha \sum_{j=1}^J \pi_j^\alpha \Gamma_j.$$

Thus, by Proposition 1 of DFT, 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

$$\begin{aligned}
\Gamma(K^{\|\cdot\|}) = 0 \quad &\iff \quad \sigma^\alpha \sum_{j=1}^J \pi_j^\alpha \Gamma_j(K^{\|\cdot\|}) = 0 \\
&\iff \quad \Gamma_j(K^{\|\cdot\|}) = 0, \ \ \forall\, j = 1, \ldots, J,
\end{aligned}$$

where the last equivalence follows from the fact that $\sigma^\alpha > 0$ and $\pi_j^\alpha > 0$ for all $j = 1, \ldots, J$. Given that the Γ_j's are the spectral measures of paths of non-aggregated moving averages, we can apply the arguments from the proof of Theorem 1 in DFT. Specifically, for each j, the condition $\Gamma_j(K^{\|\cdot\|}) = 0$ is equivalent to the representability condition (3.4) holding for the sequence $(d_{j,k})_k$ with parameter m. Therefore, $\boldsymbol{X}_t$ is representable on $C_{m+h+1}^{\|\cdot\|}$ if and only if (3.4) holds with m for all sequences $(d_{j,k})_k$, $j = 1, \ldots, J$. For the case $\alpha = 1$ and $\boldsymbol{X}_t$ asymmetric, we need to consider the additional condition involving the shift vector $\boldsymbol{\mu}^0$. From Lemma 3.1, we have:

$$\boldsymbol{\mu}^0 = -\mathbb{1}_{\{\alpha=1\}} \frac{2\sigma}{\pi} \sum_{j=1}^J \sum_{k\in\mathbb{Z}} \pi_j\beta_j \boldsymbol{d}_{j,k} \ln\|\sigma\pi_j \boldsymbol{d}_{j,k}\|_e.$$

By Proposition 1 of DFT, when $\alpha = 1$ and $\boldsymbol{X}_t$ is asymmetric, representability on $C_{m+h+1}^{\|\cdot\|}$ requires both:

1. $\Gamma(K^{\|\cdot\|}) = 0$, which as shown above is equivalent to (3.4) holding for all sequences $(d_{j,k})_k$;
2. The additional condition (3.6) must hold.

To verify condition (3.6), we need to show:

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

However, in the context of stable aggregates, this condition must be interpreted in terms of the aggregated coefficients. Since $\boldsymbol{X}_t = \sigma \sum_{j=1}^{J} \pi_j \boldsymbol{X}_{j,t}$, the effective coefficients are combinations of the individual sequences $(d_{j,k})_k$. The condition (3.6) in the aggregated case becomes:

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

where $\boldsymbol{d}_k$ now refers to the k-th vector in the aggregated representation. Given the linearity of the aggregation and the fact that the condition must hold for each component individually (as each $\boldsymbol{X}_{j,t}$ must satisfy the representability conditions), the condition (3.6) for the aggregate is satisfied if and only if it holds for all sequences $(d_{j,k})_k$, $j = 1, \dots, J$, with the same parameters m and h.

A.5. Proof of Proposition 3.1

If $\alpha \neq 1$, we have by Theorem 1 and the proof of Proposition 3 of DFT,

$$\begin{aligned}
(\mathcal{X}_t) \text{ past-representable } &\iff \exists\, m \geq 0,\ (3.4) \text{ holds with } m \text{ for all sequences } (d_{j,k})_k \\
&\iff \forall\, j = 1, \dots, J,\ m_{0,j} < +\infty \\
&\iff \forall\, j = 1, \dots, J,\ (X_{j,t}) \text{ past-representable.}
\end{aligned}$$

For a given series $(d_{j,k})_k$, (3.4) holds with $m \geq m_{0,j}$ and does not hold with $m < m_{0,j}$. Regarding the last statement, we know that for $(\mathcal{X}_t)$ (m,h)-past-representable, (3.4) 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 if $m < \max_j m_{0,j}$. In the case where $\alpha = 1$, again by Theorem 1 of DFT and denoting generically by $\boldsymbol{X}_t$ a vector $(\mathcal{X}_{t-m}, \dots, \mathcal{X}_t, \mathcal{X}_{t+1}, \dots, \mathcal{X}_{t+h})$ of size $m + h + 1$,

$$\begin{aligned}
&\mathcal{X}_t \text{ past-representable} \\
&\iff \exists\, m \geq 0, h \geq 1, \begin{cases} \boldsymbol{X}_t\ \mathcal{S}1\mathcal{S} \text{ and } (3.4) \text{ holds with } m \text{ for all sequences } (d_{j,k})_k \\ \text{or} \\ \boldsymbol{X}_t \text{ asymmetric and (3.4)-(3.6) hold with } m, h \text{ for all sequences } (d_{j,k})_k \end{cases} \\
&\iff \forall\, j = 1, \dots, J,\ m_{0,j} < +\infty, \text{ and } \exists\, m \geq 0, h \geq 1, \begin{cases} \boldsymbol{X}_t\ \mathcal{S}1\mathcal{S} \\ \text{or} \\ \boldsymbol{X}_t \text{ asymmetric and (3.6) hold} \\ \text{with } m, h \text{ for all sequences } (d_{j,k})_k \end{cases}
\end{aligned}$$

We conclude again by noting that the necessary condition (3.4) holds for $m \geq \max_j m_{0,j}$ and is violated for $m < \max_j m_{0,j}$. Now, for part $(\iota\iota)$, let $\|\cdot\|$ be a semi-norm satisfying (3.3) and assume that $\mathcal{X}_t$ is (m,h)-past-representable for some $m \geq 0$, $h \geq 1$. We need to establish the spectral representation of the vector $\boldsymbol{X_t} = (\mathcal{X}_{t-m}, \dots, \mathcal{X}_t, \mathcal{X}_{t+1}, \dots, \mathcal{X}_{t+h})$ on $C_{m+h+1}^{\|\cdot\|}$. From Lemma 3.1, we know that the spectral representation $(\Gamma, \boldsymbol{\mu}^0)$ of $\boldsymbol{X}_t$ on the Euclidean unit sphere

S_{m+h+1} is given by:

$$\Gamma = \sigma^\alpha \sum_{j=1}^{J} \sum_{\vartheta \in S_1} \sum_{k \in \mathbb{Z}} w_{j,\vartheta} \pi_j^\alpha \|\boldsymbol{d}_{j,k}\|_e^\alpha \delta_{\left\{ \frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|_e} \right\}} \tag{A.35}$$

$$\boldsymbol{\mu}^0 = \begin{cases} \mathbf{0}, & \text{if } \alpha \neq 1 \\ -\frac{2\sigma}{\pi} \sum_{j=1}^{J} \sum_{k \in \mathbb{Z}} \pi_j \beta_j \boldsymbol{d}_{j,k} \ln \|\sigma \pi_j \boldsymbol{d}_{j,k}\|_e, & \text{if } \alpha = 1 \end{cases}$$

To obtain the spectral representation on $C_{m+h+1}^{\|\cdot\|}$, we apply the transformation established in DFT for changing from Euclidean to semi-norm representations. By Lemma 3.2, since $\mathcal{X}_t$ is (m,h)-past-representable, the vector $\boldsymbol{X}_t$ is representable on $C_{m+h+1}^{\|\cdot\|}$. The transformation from the Euclidean representation to the semi-norm representation proceeds as follows. Let $K^{\|\cdot\|} := \{\boldsymbol{s} \in S_{m+h+1} : \|\boldsymbol{s}\| = 0\}$ be the kernel of the semi-norm on the Euclidean unit sphere. Since $\boldsymbol{X}_t$ is representable on $C_{m+h+1}^{\|\cdot\|}$, we have $\Gamma(K^{\|\cdot\|}) = 0$. Define the projection mapping $T_{\|\cdot\|} : S_{m+h+1} \setminus K^{\|\cdot\|} \to C_{m+h+1}^{\|\cdot\|}$ by:

$$T_{\|\cdot\|}(\boldsymbol{s}) = \frac{\boldsymbol{s}}{\|\boldsymbol{s}\|} \tag{A.36}$$

By Proposition 2 of DFT, the spectral measure on the semi-norm unit cylinder is given by:

$$\Gamma^{\|\cdot\|}(A) = \int_{T_{\|\cdot\|}^{-1}(A)} \|\boldsymbol{s}\|_e^{-\alpha} \Gamma(d\boldsymbol{s}) \tag{A.37}$$

for any Borel set $A \subset C_{m+h+1}^{\|\cdot\|}$. Since the original spectral measure Γ from (A.35) is concentrated on atoms of the form $\{\vartheta \boldsymbol{d}_{j,k}/\|\boldsymbol{d}_{j,k}\|_e\}$, and since $\|\vartheta \boldsymbol{d}_{j,k}/\|\boldsymbol{d}_{j,k}\|_e\|_e = 1$, the transformation yields:

$$\Gamma^{\|\cdot\|}(A) = \sum_{j=1}^{J} \sum_{\vartheta \in S_1} \sum_{k \in \mathbb{Z}} w_{j,\vartheta} \pi_j^\alpha \sigma^\alpha \|\boldsymbol{d}_{j,k}\|_e^\alpha \cdot 1^{-\alpha} \cdot \mathbb{1}_A \left(\frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|} \right) \tag{A.38}$$

where we use the fact that $\|\vartheta \boldsymbol{d}_{j,k}/\|\boldsymbol{d}_{j,k}\|_e\|_e = 1$ and $T_{\|\cdot\|}(\vartheta \boldsymbol{d}_{j,k}/\|\boldsymbol{d}_{j,k}\|_e) = \vartheta \boldsymbol{d}_{j,k}/\|\boldsymbol{d}_{j,k}\|$. Applying this transformation to (A.35), we obtain:

$$\Gamma^{\|\cdot\|} = \sigma^\alpha \sum_{j=1}^{J} \sum_{\vartheta \in S_1} \sum_{k \in \mathbb{Z}} w_{j,\vartheta} \pi_j^\alpha \|\boldsymbol{d}_{j,k}\|^\alpha \delta_{\left\{ \frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|} \right\}} \tag{A.39}$$

For the shift vector in the case $\alpha = 1$, the transformation yields:

$$\boldsymbol{\mu}^{\|\cdot\|} = -\frac{2\sigma}{\pi} \sum_{j=1}^{J} \sum_{k \in \mathbb{Z}} \pi_j \beta_j \boldsymbol{d}_{j,k} \ln \|\sigma \pi_j \boldsymbol{d}_{j,k}\| \tag{A.40}$$

This completes the proof that the spectral representation $(\Gamma^{\|\cdot\|}, \boldsymbol{\mu}^{\|\cdot\|})$ of $\boldsymbol{X}_t$ on $C_{m+h+1}^{\|\cdot\|}$ is given by (3.5) with the Euclidean norm $\|\cdot\|_e$ replaced by the semi-norm $\|\cdot\|$, and with the scale parameter σ explicitly included in all relevant terms.

A.6. Proof of Corollary 3.1

The equivalence between $(\iota\iota)$ and $(\iota\iota\iota)$ follows from Corollary 2 of DFT. From the proof of the Corollary in DFT, we also know that, for any j, if $m_{0,j} < +\infty$, then (3.6) holds for the sequence $(d_{j,k})_k$ for any $m \geq m_{0,j}$. For the aggregated process $\mathcal{X}_t = \sigma \sum_{j=1}^{J} \pi_j X_{j,t}$ with $\sigma > 0$, the effective moving average coefficients for each component j become $\sigma \pi_j d_{j,k}$ rather than $d_{j,k}$. However, the past-representability conditions depend only on the pattern of zeros and non-zeros in the coefficient sequences, not on their scaling. Specifically, for condition (3.4), we require:

$$\forall k \in \mathbb{Z}, \quad \Big[(\sigma \pi_j d_{j,k+m}, \ldots, \sigma \pi_j d_{j,k}) = \mathbf{0} \quad \Longrightarrow \quad \forall \ell \leq k-1, \quad \sigma \pi_j d_{j,\ell} = 0 \Big].$$

Since $\sigma > 0$ and $\pi_j > 0$ for all j, this is equivalent to:

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

Thus, the past-representability condition for the aggregated process is unchanged by the scaling factor σ. For the additional condition (3.6) when $\alpha = 1$ and the process is asymmetric, we need:

$$\sum_{k \in \mathbb{Z}} \|\sigma\pi_j \boldsymbol{d}_{j,k}\|_e \left| \ln \Big(\|\sigma\pi_j \boldsymbol{d}_{j,k}\| / \|\sigma\pi_j \boldsymbol{d}_{j,k}\|_e \Big) \right| < +\infty.$$

Since $\|\sigma\pi_j \boldsymbol{d}_{j,k}\|_e = \sigma\pi_j \|\boldsymbol{d}_{j,k}\|_e$ and the norm scales homogeneously, this becomes:

$$\sum_{k \in \mathbb{Z}} \sigma\pi_j \|\boldsymbol{d}_{j,k}\|_e \left| \ln \Big(\|\boldsymbol{d}_{j,k}\| / \|\boldsymbol{d}_{j,k}\|_e \Big) \right| < +\infty.$$

Since $\sigma\pi_j > 0$ is a finite constant, this condition is equivalent to:

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

which is precisely condition (3.6) for the unscaled sequences. Therefore:

$$\begin{aligned} \sup_j m_{0,j} < +\infty &\implies \text{(3.6) holds for any sequence } (d_{j,k})_k \text{ for any } m \geq m_{0,j} \\ &\implies \text{(3.6) holds for any sequence } (\sigma\pi_j d_{j,k})_k \text{ for any } m \geq \max_j m_{0,j}. \end{aligned}$$

Thus, $(\iota\iota\iota)$ implies (ι). The reciprocal is clear. Regarding the last statement, notice that if $\mathcal{X}_t$ is (m,h)-past-representable for some $m < \max_j m_{0,j}$, there would then exist some j such that $m < m_{0,j}$. Hence, (3.4) would not hold with m for the particular sequence $(\sigma\pi_j d_{j,k})_k$, which is impossible by Lemma 3.2, since the past-representability depends only on the zero pattern, not the scaling.

A.7. Proof of Proposition 3.2

By Proposition 2 of DFT, the asymptotic conditional tail property states that for any Borel sets $A, B \subset C^{\|\cdot\|}_{m+h+1}$ with $\Gamma^{\|\cdot\|}\Big(\partial(A \cap B)\Big) = \Gamma^{\|\cdot\|}(\partial B) = 0$, and $\Gamma^{\|\cdot\|}(B) > 0$,

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

Setting $B = B(V) = V \times \mathbb{R}^h$, we have

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

From Proposition 3.1 $(\iota\iota)$, the spectral representation $(\Gamma^{\|\cdot\|}, \boldsymbol{\mu}^{\|\cdot\|})$ of the vector $\boldsymbol{X_t} = (\mathcal{X}_{t-m}, \dots, \mathcal{X}_t, \mathcal{X}_{t+1}, \dots, \mathcal{X}_{t+h})$ on $C^{\|\cdot\|}_{m+h+1}$ is given by equation (3.5) with the Euclidean norm $\|\cdot\|_e$ replaced by the semi-norm $\|\cdot\|$. From Lemma 3.1, the spectral measure can be written as:

$$\Gamma^{\|\cdot\|} = \sigma^\alpha \sum_{j=1}^{J} \sum_{\vartheta \in S_1} \sum_{k \in \mathbb{Z}} w_{j,\vartheta} \pi_j^\alpha \|\boldsymbol{d}_{j,k}\|^\alpha \delta_{\left\{ \frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|} \right\}},$$

where $\boldsymbol{d}_{j,k} = (d_{j,k+m}, \dots, d_{j,k}, d_{j,k-1}, \dots, d_{j,k-h})$, $w_{j,\vartheta} = (1+\vartheta\beta_j)/2$, and if $\boldsymbol{d}_{j,k} = \mathbf{0}$, the term vanishes by convention from the sums.

Now, we compute the numerator and denominator separately, we start by the numerator: $\Gamma^{\|\cdot\|}(A \cap B(V))$ Since $B(V) = V \times \mathbb{R}^h = \Big\{\boldsymbol{s} \in C^{\|\cdot\|}_{m+h+1} : \quad f(\boldsymbol{s}) \in V\Big\}$, we have:

$$A \cap B(V) = \{\boldsymbol{s} \in A : \quad f(\boldsymbol{s}) \in V\}.$$

The spectral measure $\Gamma^{\|\cdot\|}$ charges only the points of the form $\dfrac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|}$ for $(\vartheta, j, k) \in S_1 \times \{1, \dots, J\} \times \mathbb{Z}$. Therefore:

$$
\begin{aligned}
\Gamma^{\|\cdot\|}(A \cap B(V)) &= \sigma^\alpha \sum_{j=1}^{J} \sum_{\vartheta \in S_1} \sum_{k \in \mathbb{Z}} w_{j,\vartheta} \pi_j^\alpha \|\boldsymbol{d}_{j,k}\|^\alpha \delta_{\left\{\frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|}\right\}}(A \cap B(V)) \\
&= \sigma^\alpha \sum_{\substack{(\vartheta,j,k): \\ \frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|} \in A \cap B(V)}} w_{j,\vartheta} \pi_j^\alpha \|\boldsymbol{d}_{j,k}\|^\alpha \\
&= \sigma^\alpha \sum_{\substack{(\vartheta,j,k): \\ \frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|} \in A \text{ and } \frac{\vartheta f(\boldsymbol{d}_{j,k})}{\|\boldsymbol{d}_{j,k}\|} \in V}} w_{j,\vartheta} \pi_j^\alpha \|\boldsymbol{d}_{j,k}\|^\alpha .
\end{aligned}
$$

This can be written as:

$$
\Gamma^{\|\cdot\|}(A \cap B(V)) = \Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|} \in A : \quad \frac{\vartheta f(\boldsymbol{d}_{j,k})}{\|\boldsymbol{d}_{j,k}\|} \in V\right\}\right).
$$

For the denominator $\Gamma^{\|\cdot\|}(B(V))$, we proceed as follows:

$$
\begin{aligned}
\Gamma^{\|\cdot\|}(B(V)) &= \sigma^\alpha \sum_{\substack{(\vartheta,j,k): \\ \frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|} \in B(V)}} w_{j,\vartheta} \pi_j^\alpha \|\boldsymbol{d}_{j,k}\|^\alpha \\
&= \sigma^\alpha \sum_{\substack{(\vartheta,j,k): \\ \frac{\vartheta f(\boldsymbol{d}_{j,k})}{\|\boldsymbol{d}_{j,k}\|} \in V}} w_{j,\vartheta} \pi_j^\alpha \|\boldsymbol{d}_{j,k}\|^\alpha .
\end{aligned}
$$

This can be written as:

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

Note that the factor σ^α appears in both the numerator and denominator, and therefore cancels out in the ratio:

$$
\begin{aligned}
\frac{\Gamma^{\|\cdot\|}(A \cap B(V))}{\Gamma^{\|\cdot\|}(B(V))} &= \frac{\sigma^\alpha \sum_{\substack{(\vartheta,j,k): \\ \frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|} \in A \text{ and } \frac{\vartheta f(\boldsymbol{d}_{j,k})}{\|\boldsymbol{d}_{j,k}\|} \in V}} w_{j,\vartheta} \pi_j^\alpha \|\boldsymbol{d}_{j,k}\|^\alpha}{\sigma^\alpha \sum_{\substack{(\vartheta,j,k): \\ \frac{\vartheta f(\boldsymbol{d}_{j,k})}{\|\boldsymbol{d}_{j,k}\|} \in V}} w_{j,\vartheta} \pi_j^\alpha \|\boldsymbol{d}_{j,k}\|^\alpha} \\
&= \frac{\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|} \in A : \quad \frac{\vartheta f(\boldsymbol{d}_{j,k})}{\|\boldsymbol{d}_{j,k}\|} \in V\right\}\right)}{\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|} \in C^{\|\cdot\|}_{m+h+1} : \quad \frac{\vartheta f(\boldsymbol{d}_{j,k})}{\|\boldsymbol{d}_{j,k}\|} \in V\right\}\right)}.
\end{aligned}
$$

This establishes the desired result. The conclusion follows by considering the points of $B(V)$ and $A \cap B(V)$ that are charged by the spectral measure $\Gamma^{\|\cdot\|}$ given in equation (3.12). The presence of the scale parameter σ^α does not affect the asymptotic conditional probabilities as it appears multiplicatively in both the numerator and denominator of the ratio, thus canceling out in the final expression.

A.8. Proof of Lemma 3.3

By Proposition 3.1 and setting general scale parameter $\sigma > 0$, we have

$$
\Gamma^{\|\cdot\|} = \sum_{j=1}^{J} \sum_{\vartheta \in S_1} \sum_{k \in \mathbb{Z}} w_{j,\vartheta} \sigma^\alpha \pi_j^\alpha \|\boldsymbol{d}_{j,k}\|^\alpha \delta_{\left\{\frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|}\right\}},
$$

with $\boldsymbol{d}_{j,k} = (\psi_j^{k+m}\mathbb{1}_{\{k+m\geq 0\}}, \dots, \psi_j^{k-h}\mathbb{1}_{\{k-h\geq 0\}})$ for any $j = 1, \dots, J$ and $k \in \mathbb{Z}$. Thus, for any $j \in \{1, \dots, J\}$

$$\boldsymbol{d}_{j,k} = \begin{cases} \boldsymbol{0}, & \text{if} \quad k \leq -m-1, \\ (\psi_j^{k+m}, \dots, \psi_j, 1, 0, \dots, 0), & \text{if} \quad -m \leq k \leq h, \\ \psi_j^{k-h}\boldsymbol{d}_{j,h}, & \text{if} \quad k \geq h. \end{cases}$$

Therefore,

$$\Gamma^{\|\cdot\|} = \sum_{j=1}^{J}\sum_{\vartheta\in S_1} w_{j,\vartheta}\sigma^\alpha\pi_j^\alpha\left[\sum_{k=-m}^{h-1}\|\boldsymbol{d}_{j,k}\|^\alpha\delta_{\left\{\frac{\vartheta\boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|}\right\}} + \sum_{k=h}^{+\infty}|\psi_j|^{\alpha(k-h)}\|\boldsymbol{d}_{j,h}\|^\alpha\delta_{\left\{\frac{\vartheta\psi_j^{k-h}\boldsymbol{d}_{j,h}}{|\psi_j|^{k-h}\|\boldsymbol{d}_{j,h}\|}\right\}}\right].$$

Moreover,

$$\begin{aligned}\sum_{j=1}^{J}\sum_{\vartheta\in S_1} w_{j,\vartheta}\sigma^\alpha\pi_j^\alpha &\sum_{k=h}^{+\infty}|\psi_j|^{\alpha(k-h)}\|\boldsymbol{d}_{j,h}\|^\alpha\delta_{\left\{\operatorname{sign}(\psi_j)^{k-h}\frac{\vartheta\boldsymbol{d}_{j,h}}{\|\boldsymbol{d}_{j,h}\|}\right\}} \\ &= \sum_{j=1}^{J}\sum_{\vartheta\in S_1}\sigma^\alpha\pi_j^\alpha\|\boldsymbol{d}_{j,h}\|^\alpha\frac{1}{2}\left[\sum_{k=h}^{+\infty}|\psi_j|^{\alpha(k-h)} + \vartheta\beta_j\sum_{k=h}^{+\infty}(\psi_j^{<\alpha>})^{k-h}\right]\delta_{\left\{\frac{\vartheta\boldsymbol{d}_{j,h}}{\|\boldsymbol{d}_{j,h}\|}\right\}} \\ &= \sum_{j=1}^{J}\sum_{\vartheta\in S_1}\sigma^\alpha\pi_j^\alpha\frac{1}{1-|\psi_j|^\alpha}\|\boldsymbol{d}_{j,h}\|^\alpha\bar{w}_{j,\vartheta}\delta_{\left\{\frac{\vartheta\,\boldsymbol{d}_{j,h}}{\|\boldsymbol{d}_{j,h}\|}\right\}}.\end{aligned}$$

Finally, noticing that for $k = -m$ and any $j \in \{1, \dots, J\}$, $\boldsymbol{d}_{j,k} = (1, 0, \dots, 0)$,

$$\begin{aligned}\Gamma^{\|\cdot\|} &= \sum_{j=1}^{J}\sum_{\vartheta\in S_1}\sigma^\alpha\pi_j^\alpha\left[w_{j,\vartheta}\sum_{k=-m}^{h-1}\|\boldsymbol{d}_{j,k}\|^\alpha\delta_{\left\{\frac{\vartheta\boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|}\right\}} + \frac{\bar{w}_{j,\vartheta}}{1-|\psi_j|^\alpha}\|\boldsymbol{d}_{j,h}\|^\alpha\delta_{\left\{\frac{\vartheta\,\boldsymbol{d}_{j,h}}{\|\boldsymbol{d}_{j,h}\|}\right\}}\right] \\ &= \sum_{j=1}^{J}\sum_{\vartheta\in S_1}\sigma^\alpha\pi_j^\alpha\left[w_{j,\vartheta}\left(\delta_{\{(\vartheta,0,\dots,0)\}} + \sum_{k=-m+1}^{h-1}\|\boldsymbol{d}_{j,k}\|^\alpha\delta_{\left\{\frac{\vartheta\boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|}\right\}}\right) + \frac{\bar{w}_{j,\vartheta}}{1-|\psi_j|^\alpha}\|\boldsymbol{d}_{j,h}\|^\alpha\delta_{\left\{\frac{\vartheta\,\boldsymbol{d}_{j,h}}{\|\boldsymbol{d}_{j,h}\|}\right\}}\right] \\ &= \sum_{\vartheta\in S_1}\left[w_\vartheta\delta_{\{(\vartheta,0,\dots,0)\}} + \sum_{j=1}^{J}\sigma^\alpha\pi_j^\alpha\left(w_{j,\vartheta}\sum_{k=-m+1}^{h-1}\|\boldsymbol{d}_{j,k}\|^\alpha\delta_{\left\{\frac{\vartheta\boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|}\right\}} + \frac{\bar{w}_{j,\vartheta}}{1-|\psi_j|^\alpha}\|\boldsymbol{d}_{j,h}\|^\alpha\delta_{\left\{\frac{\vartheta\,\boldsymbol{d}_{j,h}}{\|\boldsymbol{d}_{j,h}\|}\right\}}\right)\right],\end{aligned}$$

where we have used the definition $w_\vartheta = \sum_{j=1}^{J}\sigma^\alpha\pi_j^\alpha w_{j,\vartheta}$.

A.9. Proof of Proposition 3.3

Lemma A.2. *Let $\Gamma^{\|\cdot\|}$ be the spectral measure given in Lemma 3.3 with $\sigma > 0$ and assume that the ψ_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

$$I_0 = \begin{cases}\left\{\dfrac{\vartheta_0\boldsymbol{d}_{j_0,k_0}}{\|\boldsymbol{d}_{j_0,k_0}\|}\right\}, & \textit{if} \quad -m+1 \leq k_0 \leq -1 \\[2ex] \left\{\dfrac{\vartheta_0\boldsymbol{d}_{0,k_0}}{\|\boldsymbol{d}_{0,k_0}\|}\right\} = \{(\vartheta_0, 0, \dots, 0)\}, & \textit{if} \quad k_0 = -m.\end{cases}$$

For $m = 0$, then

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

Proof. The key observation is that the parameter $\sigma > 0$ appears as a multiplicative factor in the spectral measure $\Gamma^{\|\cdot\|}$ but does **not** affect the normalized directions $\vartheta' \boldsymbol{d}_{j',k'}/\|\boldsymbol{d}_{j',k'}\|$ or their projections $\vartheta' f(\boldsymbol{d}_{j',k'})/\|\boldsymbol{d}_{j',k'}\|$. This is because σ only scales the overall magnitude of the spectral measure but does not change the geometric structure of the charged points on the unit cylinder. More precisely, from Lemma 3.3, the spectral measure takes the form:

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

The factor σ^α multiplies the entire spectral measure uniformly, but the support of $\Gamma^{\|\cdot\|}$ (i.e., the set of points where $\Gamma^{\|\cdot\|}$ assigns positive mass) consists exactly of the normalized directions:

$$\operatorname{supp}(\Gamma^{\|\cdot\|}) = \left\{ (\vartheta, 0, \ldots, 0), \frac{\vartheta \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|} : \vartheta \in S_1, j \in \{1, \ldots, J\}, k \in \{-m+1, \ldots, h\} \right\}$$

Since the condition defining I_0 involves only the equality of normalized projections:

$$\frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})}{\|\boldsymbol{d}_{j_0,k_0}\|}$$

and since these normalized directions are independent of σ, the analysis proceeds exactly as in the case $\sigma = 1$.

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

If $k' \in \{-m, \ldots, -1\}$, the $(m+1)$-th component of $f(\boldsymbol{d}_{j',k'})$ is zero, whereas the $(m+1)$-th component of $f(\boldsymbol{d}_{j_0,k_0})$ is $\psi_{j_0}^{k_0} \neq 0$. This geometric relationship is unaffected by σ.

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

$$\begin{aligned} f(\boldsymbol{d}_{j',k'}) &= (\psi_{j'}^{k'+m}, \ldots, \psi_{j'}^{k'+1}, \psi_{j'}^{k'}), \\ f(\boldsymbol{d}_{j_0,k_0}) &= (\psi_{j_0}^{k_0+m}, \ldots, \psi_{j_0}^{k_0+1}, \psi_{j_0}^{k_0}), \end{aligned}$$

and by (3.3) we also have that

$$\begin{aligned} \|\boldsymbol{d}_{j',k'}\| &= \|(\psi_{j'}^{k'+m}, \ldots, \psi_{j'}^{k'+1}, \psi_{j'}^{k'}, \overbrace{0, \ldots, 0}^{h})\|, \\ \|\boldsymbol{d}_{j_0,k_0}\| &= \|(\psi_{j_0}^{k_0+m}, \ldots, \psi_{j_0}^{k_0+1}, \psi_{j_0}^{k_0}, \underbrace{0, \ldots, 0}_{h})\|. \end{aligned}$$

The key observation is that these norms and the resulting normalized directions are independent of σ. Thus,

$$\begin{aligned} \frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} &= \frac{\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})}{\|\boldsymbol{d}_{j_0,k_0}\|} \\ &\iff \frac{\vartheta' \psi_{j'}^{k'} f(\boldsymbol{d}_{j',0})}{|\psi_{j'}|^{k'} \|\boldsymbol{d}_{j',0}\|} = \frac{\vartheta_0 \psi_{j_0}^{k_0} f(\boldsymbol{d}_{j_0,0})}{|\psi_{j_0}|^{k_0} \|\boldsymbol{d}_{j_0,0}\|} \\ &\iff \frac{\vartheta' \psi_{j'}^{\ell}}{\|\boldsymbol{d}_{j',0}\|} = \frac{\vartheta_0 \psi_{j_0}^{\ell}}{\|\boldsymbol{d}_{j_0,0}\|}, \qquad \ell = 0, \ldots, m \\ &\iff \vartheta' \vartheta_0 \frac{\|\boldsymbol{d}_{j_0,0}\|}{\|\boldsymbol{d}_{j',0}\|} = \left(\frac{\psi_{j_0}}{\psi_{j'}} \right)^\ell, \quad \ell = 0, \ldots, m \\ &\iff \psi_{j'} = \psi_{j_0} \quad \text{and} \quad \vartheta' \vartheta_0 = 1 \\ &\iff j' = j_0 \quad \text{and} \quad \vartheta' = \vartheta_0, \end{aligned}$$

because the ψ_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

$$\frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})}{\|\boldsymbol{d}_{j_0,k_0}\|} \implies k' = k_0.$$

$$\begin{aligned} f(\boldsymbol{d}_{j',k'}) &= (\overbrace{\psi_{j'}^{k'+m}, \dots, \psi_{j'}, 1, 0, \dots, 0}^{m+1}), \\ f(\boldsymbol{d}_{j_0,k_0}) &= (\underbrace{\psi_{j_0}^{k_0+m}, \dots, \psi_{j_0}, 1, 0, \dots, 0}_{m+1}), \end{aligned}$$

and we also have that

$$\begin{aligned} \|\boldsymbol{d}_{j',k'}\| &= \|(\overbrace{\psi_{j'}^{k'+m}, \dots, \psi_{j'}, 1, 0, \dots, 0}^{m+1}, \overbrace{0, \dots, 0}^{h})\|, \\ \|\boldsymbol{d}_{j_0,k_0}\| &= \|(\underbrace{\psi_{j_0}^{k_0+m}, \dots, \psi_{j_0}, 1, 0, \dots, 0}_{m+1}, \underbrace{0, \dots, 0}_{h})\|. \end{aligned}$$

As $k' = k_0 \leq -1$, the condition becomes:

$$\begin{aligned} &\frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})}{\|\boldsymbol{d}_{j_0,k_0}\|} \\ &\qquad \iff \frac{\vartheta' \psi_{j'}^{\ell}}{\|\boldsymbol{d}_{j',k_0}\|} = \frac{\vartheta_0 \psi_{j_0}^{\ell}}{\|\boldsymbol{d}_{j_0,k_0}\|}, \quad \ell = 0, \dots, m + k_0, \text{ and } k' = k_0 \\ &\qquad \iff \vartheta' \vartheta_0 \frac{\|\boldsymbol{d}_{j_0,k_0}\|}{\|\boldsymbol{d}_{j',k_0}\|} = \left(\frac{\psi_{j_0}}{\psi_{j'}}\right)^{\ell}, \quad \ell = 0, \dots, m + k_0, \text{ and } k' = k_0. \end{aligned}$$

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

$$\begin{aligned} &\vartheta' \vartheta_0 \frac{\|\boldsymbol{d}_{j_0,k_0}\|}{\|\boldsymbol{d}_{j',k_0}\|} = \left(\frac{\psi_{j_0}}{\psi_{j'}}\right)^{\ell}, \quad \ell = 0, 1, \dots, m + k_0, \text{ and } k' = k_0 \\ &\qquad\qquad \iff \vartheta' = \vartheta_0 \text{ and } j' = j_0 \text{ and } k' = k_0. \end{aligned}$$

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

$$\begin{aligned} &\vartheta' \vartheta_0 \frac{\|\boldsymbol{d}_{j_0,k_0}\|}{\|\boldsymbol{d}_{j',k_0}\|} = \left(\frac{\psi_{j_0}}{\psi_{j'}}\right)^{\ell}, \quad \ell = 0, \text{ and } k' = k_0 = -m \text{ and } j' = j_0 = 0, \\ &\qquad \iff \vartheta' = \vartheta_0 \text{ and } k' = k_0 = -m \text{ and } j' = j_0 = 0 \end{aligned}$$

Case $\boldsymbol{m = 0}$

If $k_0 \in \{1, \dots, h\}$ then $f(\boldsymbol{d}_{j_0,k_0}) = \psi_{j_0}^{k_0}$ and by (3.3), $\|\boldsymbol{d}_{j_0,k_0}\| = |\psi_{j_0}|^{k_0}$. 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

$$\frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})}{\|\boldsymbol{d}_{j_0,k_0}\|} \iff \vartheta' = \vartheta_0.$$

Proof. By Proposition 3.2,

$$\mathbb{P}_x^{\|\cdot\|}\Big(\boldsymbol{X_t}, A_{\vartheta,j,k}\Big|B(V_0)\Big) \underset{x\to\infty}{\longrightarrow} \frac{\Gamma^{\|\cdot\|}\left(\left\{\dfrac{\vartheta' \boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|} \in A_{\vartheta,j,k} : \dfrac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} \in V_0\right\}\right)}{\Gamma^{\|\cdot\|}\left(\left\{\dfrac{\vartheta' \boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|} \in C_{m+h+1}^{\|\cdot\|} : \dfrac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} \in V_0\right\}\right)}. \tag{A.41}$$

Focusing on the denominator, we have by (3.15)

$$\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta'\boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|}\in C^{\|\cdot\|}_{m+h+1}:\quad\frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|}\in V_0\right\}\right)=\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta'\boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|}\in C^{\|\cdot\|}_{m+h+1}:\quad\frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|}=\frac{\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})}{\|\boldsymbol{d}_{j_0,k_0}\|}\right\}\right)$$

We will now distinguish the cases arising from the application of Lemma A.2. Recall that we assume for this proposition that the ψ_j's are positive. Thus, $\operatorname{sign}(\psi_j)=1$ and $\bar{\beta}_j=\beta_j\dfrac{1-|\psi_j|^\alpha}{1-\psi_j^{<\alpha>}}=\beta_j$ and $\bar{w}_{j,\vartheta}=w_{j,\vartheta}$ in (3.14) for all j's and $\vartheta\in\{-1,+1\}$.

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

By Lemma A.2,

$$\begin{aligned}
\Gamma^{\|\cdot\|}&\left(\left\{\frac{\vartheta'\boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|}\in C^{\|\cdot\|}_{m+h+1}:\quad\frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|}=\frac{\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})}{\|\boldsymbol{d}_{j_0,k_0}\|}\right\}\right)\\
&=\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta_0\boldsymbol{d}_{j_0,k'}}{\|\boldsymbol{d}_{j_0,k'}\|}:\quad 0\leq k'\leq h\right\}\right)\\
&=\sigma^\alpha\pi^\alpha_{j_0}\left[w_{j_0,\vartheta_0}\sum_{k'=0}^{h-1}\|\boldsymbol{d}_{j_0,k'}\|^\alpha+\frac{\bar{w}_{j_0,\vartheta_0}}{1-|\psi_{j_0}|^\alpha}\|\boldsymbol{d}_{j_0,h}\|^\alpha\right]
\end{aligned}$$

By (3.3), for $k'\in\{0,1,\ldots,h\}$

$$\begin{aligned}
\|\boldsymbol{d}_{j_0,k'}\|&=\|(\psi_{j_0}^{k'+m},\ldots,\psi_{j_0}^{k'+1},\psi_{j_0}^{k'},\underbrace{0,\ldots,0}_{h})\|\\
&=|\psi_{j_0}|^{k'-h}\|(\psi_{j_0}^{m+h},\ldots,\psi_{j_0}^{h+1},\psi_{j_0}^{h},\underbrace{0,\ldots,0}_{h})\|\\
&=|\psi_{j_0}|^{k'-h}\|\boldsymbol{d}_{j_0,h}\|.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta'\boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|}\in C^{\|\cdot\|}_{m+h+1}:\quad\frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|}=\frac{\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})}{\|\boldsymbol{d}_{j_0,k_0}\|}\right\}\right)\\
&=\sigma^\alpha\pi^\alpha_{j_0}w_{j_0,\vartheta_0}\|\boldsymbol{d}_{j_0,h}\|^\alpha\left[\sum_{k'=0}^{h-1}|\psi_{j_0}|^{\alpha(k'-h)}+\frac{1}{1-|\psi_{j_0}|^\alpha}\right]\\
&=\sigma^\alpha\pi^\alpha_{j_0}w_{j_0,\vartheta_0}\|\boldsymbol{d}_{j_0,h}\|^\alpha\frac{|\psi_{j_0}|^{-\alpha h}}{1-|\psi_{j_0}|^\alpha}.
\end{aligned}$$

Similarly for the numerator in (A.41), by (3.16),

$$\begin{aligned}
\Gamma^{\|\cdot\|}&\left(\left\{\frac{\vartheta'\boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|}\in A_{\vartheta,j,k}:\quad\frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|}\in V_0\right\}\right)\\
&=\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta_0\boldsymbol{d}_{j_0,k'}}{\|\boldsymbol{d}_{j_0,k'}\|}\in A_{\vartheta,j,k}:\quad 0\leq k'\leq h\right\}\right)\\
&=\begin{cases}\Gamma^{\|\cdot\|}\left(\left\{\dfrac{\vartheta_0\boldsymbol{d}_{j_0,k}}{\|\boldsymbol{d}_{j_0,k}\|}\right\}\right), & \text{if}\quad j=j_0\quad\text{and}\quad\vartheta=\vartheta_0,\\ \Gamma^{\|\cdot\|}(\emptyset), & \text{if}\quad j\neq j_0\quad\text{or}\quad\vartheta\neq\vartheta_0,\end{cases}\\
&=\begin{cases}\sigma^\alpha\pi^\alpha_{j_0}w_{j_0,\vartheta_0}\|\boldsymbol{d}_{j_0,h}\|^\alpha|\psi_{j_0}|^{\alpha(k-h)}\delta_{\{\vartheta_0\}}(\vartheta)\delta_{\{j_0\}}(j), & \text{if}\quad 0\leq k\leq h-1,\\ \sigma^\alpha\pi^\alpha_{j_0}w_{j_0,\vartheta_0}\|\boldsymbol{d}_{j_0,h}\|^\alpha\dfrac{1}{1-|\psi_{j_0}|^\alpha}\delta_{\{\vartheta_0\}}(\vartheta)\delta_{\{j_0\}}(j), & \text{if}\quad k=h.\end{cases}
\end{aligned}$$

The σ^α terms cancel out in the ratio.

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

We have by Lemma A.2

$$\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta' \boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|} \in C^{\|\cdot\|}_{m+h+1}: \quad \frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})}{\|\boldsymbol{d}_{j_0,k_0}\|}\right\}\right) = \Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta_0 \boldsymbol{d}_{j_0,k_0}}{\|\boldsymbol{d}_{j_0,k_0}\|}\right\}\right).$$

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

$$\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta_0 \boldsymbol{d}_{j_0,k_0}}{\|\boldsymbol{d}_{j_0,k_0}\|}\right\}\right) = \sigma^\alpha \pi^\alpha_{j_0} w_{j_0,\vartheta_0} \|\boldsymbol{d}_{j_0,k_0}\|^\alpha,$$

and

$$\begin{aligned}\Gamma^{\|\cdot\|}&\left(\left\{\frac{\vartheta' \boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|} \in A_{\vartheta,j,k}: \quad \frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} \in V_0\right\}\right) \\ &= \Gamma^{\|\cdot\|}\left(A_{\vartheta,j,k} \cap \left\{\frac{\vartheta_0 \boldsymbol{d}_{j_0,k_0}}{\|\boldsymbol{d}_{j_0,k_0}\|}\right\}\right) \\ &= \begin{cases} \Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta_0 \boldsymbol{d}_{j_0,k_0}}{\|\boldsymbol{d}_{j_0,k_0}\|}\right\}\right), & \text{if } j = j_0 \text{ and } \vartheta = \vartheta_0, \text{ and } k = k_0, \\ \Gamma^{\|\cdot\|}(\emptyset), & \text{if } j \neq j_0 \text{ or } \vartheta \neq \vartheta_0 \text{ or } k \neq k_0, \end{cases} \\ &= \sigma^\alpha \pi^\alpha_{j_0} w_{j_0,\vartheta_0} \|\boldsymbol{d}_{j_0,k_0}\|^\alpha \delta_{\{\vartheta_0\}}(\vartheta)\delta_{\{j_0\}}(j)\delta_{\{k_0\}}(k).\end{aligned}$$

If $k_0 = -m$, then $\boldsymbol{d}_{j_0,k_0} = \boldsymbol{d}_{0,-m} = (1, 0, \ldots, 0)$, and

$$\Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta_0 \boldsymbol{d}_{j_0,k_0}}{\|\boldsymbol{d}_{j_0,k_0}\|}\right\}\right) = \Gamma^{\|\cdot\|}\Big(\{\vartheta_0(1, 0, \ldots, 0)\}\Big) = \sigma^\alpha w_{\vartheta_0},$$

and

$$\begin{aligned}\Gamma^{\|\cdot\|}&\left(\left\{\frac{\vartheta' \boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|} \in A_{\vartheta,j,k}: \quad \frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} \in V_0\right\}\right) \\ &= \Gamma^{\|\cdot\|}\left(A_{\vartheta,j,k} \cap \left\{\frac{\vartheta_0 \boldsymbol{d}_{j_0,k_0}}{\|\boldsymbol{d}_{j_0,k_0}\|}\right\}\right) \\ &= \begin{cases} \Gamma^{\|\cdot\|}\Big(A_{\vartheta,j,k} \cap \{\vartheta_0(1, 0, \ldots, 0)\}\Big), & \text{if } \vartheta = \vartheta_0, \text{ and } k = k_0 = -m, \text{ and } j = j_0 = 0 \\ \Gamma^{\|\cdot\|}(\emptyset), & \text{if } \vartheta \neq \vartheta_0 \text{ or } k \neq k_0, \text{ or } j \neq j_0 \end{cases} \\ &= \sigma^\alpha w_{\vartheta_0} \delta_{\{\vartheta_0\}}(\vartheta)\delta_{\{j_0\}}(j)\delta_{\{k_0\}}(k).\end{aligned}$$

Again, the σ^α terms cancel out in the ratio.

Case $m = 0$

By Lemma A.2, as the ψ_j's are positive

$$\begin{aligned}\Gamma^{\|\cdot\|}&\left(\left\{\frac{\vartheta' \boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|} \in C^{\|\cdot\|}_{m+h+1}: \quad \frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})}{\|\boldsymbol{d}_{j_0,k_0}\|}\right\}\right) \\ &= \Gamma^{\|\cdot\|}\left(\left\{\frac{\vartheta_0 \boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|} \in C^{\|\cdot\|}_{m+h+1}: \quad (j', k') \in \{1, \ldots, J\} \times \{0, \ldots, h\} \cup \{(0, 0)\}\right\}\right)\end{aligned}$$

Given that $w_{\vartheta_0} = \sum_{j'=1}^J \pi_{j'}^\alpha w_{j',\vartheta_0}$ and $\|\boldsymbol{d}_{j',k'}\| = |\psi_{j'}|^{k'}$, for any $1 \le j' \le J$, $1 \le k' \le h$,

$$\begin{aligned}
\Gamma^{\|\cdot\|}&\left(\left\{\frac{\vartheta' \boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|} \in C_{m+h+1}^{\|\cdot\|} : \quad \frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} = \frac{\vartheta_0 f(\boldsymbol{d}_{j_0,k_0})}{\|\boldsymbol{d}_{j_0,k_0}\|}\right\}\right) \\
&= \sigma^\alpha w_{\vartheta_0} + \sigma^\alpha \sum_{j'=1}^J \pi_{j'}^\alpha w_{j',\vartheta_0} \left[\sum_{k'=1}^{h-1} \|\boldsymbol{d}_{j',k'}\|^\alpha + \frac{\|\boldsymbol{d}_{j',h}\|^\alpha}{1-|\psi_{j'}|^\alpha}\right] \\
&= \sigma^\alpha \sum_{j'=1}^J \pi_{j'}^\alpha w_{j',\vartheta_0} \left[1 + \sum_{k'=1}^{h-1} |\psi_{j'}|^{\alpha k'} + \frac{|\psi_{j'}|^{\alpha h}}{1-|\psi_{j'}|^\alpha}\right] \\
&= \sigma^\alpha \sum_{j'=1}^J \pi_{j'}^\alpha w_{j',\vartheta_0} \left[\frac{1-|\psi_{j'}|^{\alpha h}}{1-|\psi_{j'}|^\alpha} + \frac{|\psi_{j'}|^{\alpha h}}{1-|\psi_{j'}|^\alpha}\right] \\
&= \sigma^\alpha \sum_{j'=1}^J \pi_{j'}^\alpha w_{j',\vartheta_0} \frac{1}{1-|\psi_{j'}|^\alpha}.
\end{aligned}$$

Similarly, by (3.16),

$$\begin{aligned}
\Gamma^{\|\cdot\|}&\left(\left\{\frac{\vartheta' \boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|} \in A_{\vartheta,j,k} : \quad \frac{\vartheta' f(\boldsymbol{d}_{j',k'})}{\|\boldsymbol{d}_{j',k'}\|} \in V_0\right\}\right) \\
&= \Gamma^{\|\cdot\|}\left(A_{\vartheta,j,k} \cap \left\{\frac{\vartheta_0 \boldsymbol{d}_{j',k'}}{\|\boldsymbol{d}_{j',k'}\|} \in C_{m+h+1}^{\|\cdot\|} : \quad (j',k') \in \{1,\dots,J\} \times \{0,\dots,h\} \cup \{(0,0)\}\right\}\right) \\
&= \begin{cases} \Gamma^{\|\cdot\|}\left(\left\{\dfrac{\vartheta_0 \boldsymbol{d}_{j,k}}{\|\boldsymbol{d}_{j,k}\|}\right\}\right), & \text{if } \vartheta = \vartheta_0, \\ \Gamma^{\|\cdot\|}(\emptyset), & \text{if } \vartheta \neq \vartheta_0, \end{cases} \\
&= \begin{cases} \sigma^\alpha \sum_{j'=1}^J \pi_{j'}^\alpha w_{j',\vartheta_0} \delta_{\{\vartheta_0\}}(\vartheta), & \text{if } \quad k = 0, \\ \sigma^\alpha \pi_j^\alpha w_{j,\vartheta_0} |\psi_j|^{\alpha k} \delta_{\{\vartheta_0\}}(\vartheta), & \text{if } \quad 1 \le k \le h-1, \\ \sigma^\alpha \pi_j^\alpha w_{j,\vartheta_0} \dfrac{|\psi_j|^{\alpha h}}{1-|\psi_j|^\alpha} \delta_{\{\vartheta_0\}}(\vartheta), & \text{if } \quad k = h. \end{cases}
\end{aligned}$$

The conclusion follows.