

On the Stationary Marginal Distributions of Subclasses of Multivariate SETAR Processes of Order One

Soumya Das¹ and Marc G. Genton¹

January 7, 2020

Abstract

In order to introduce more flexibility in process-parameters through a regime-switching behavior, the classical autoregressive (AR) processes have been extended to self-exciting threshold autoregressive (SETAR) processes. However, the stationary marginal distributions of SETAR processes are usually difficult to obtain in explicit forms and, therefore, they lack appropriate characterizations. The stationary marginal distribution of a multivariate (d -dimensional) SETAR process of order one ($\text{MSETAR}_d(1)$) with multivariate normal innovations is shown to belong to the unified skew-normal (SUN) family and characterized under a fairly broad condition. This article also characterizes the stationary marginal distributions of a subclass of the $\text{MSETAR}_d(1)$ processes with symmetric multivariate stable innovations. To characterize the stationary marginal distributions of these processes, the authors show that they belong to specific skew-distribution families, and for a given skew-distribution from the corresponding family, an $\text{MSETAR}_d(1)$ process, with stationary marginal distribution identical to the given skew-distribution, can be associated. Furthermore, this article illustrates a diagnostic of an $\text{MSETAR}_2(1)$ model using the corresponding stationary marginal density.

Keywords: multivariate self-exciting threshold autoregressive model; multivariate skew-normal distribution; skew-symmetric multivariate stable distribution; stationary marginal density.

¹Statistics Program, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia.
E-mail: soumya.das@kaust.edu.sa, marc.genton@kaust.edu.sa

1 Introduction

Due essentially to the various limitations of linear autoregressive models in real applications, nonlinear autoregressive models and their stationary marginal distributions have been of interest to the time series community for several decades; see Tong (1978, 1983, 1990, 2011), and references therein. Among different nonlinear autoregressive models, self-exciting threshold autoregressive (SETAR) models have gathered significant attention because of the simplicity of their definition and, at the same time, enough flexibility of their representation to explain relatively complex time series processes. Following the growing needs of the SETAR models, Arnold and Günther (2001) extended the idea to the multivariate scenario to define multivariate SETAR (MSETAR) models and discussed their importance in applications; see also Chan et al. (2004), Baragona and Cucina (2013), Addo (2014), and Wong et al. (2017).

While there is an extensive study on obtaining explicit forms of the stationary marginal densities of such models (e.g. Anděl et al., 1984; Anděl and Bartoň, 1986; Chan and Tong, 1986; Loges, 2004), little work has been done to characterize them. By and large, our current work is intended to link the stationary marginal densities of certain MSETAR models to the skew-symmetric densities (Genton, 2004; Azzalini and Capitanio, 2014) and to characterize them under certain additional but fairly broad assumptions.

The main motivation of this article comes from an example provided by Chan and Tong (1986) to illustrate their general formula, to derive the stationary marginal densities for a subclass of MSETAR (d -dimensional) processes of order one (MSETAR $_d(1)$), that relied on the ‘symmetry’ of the autoregressive function. In that example, Chan and Tong (1986) obtained the stationary marginal density, up to a multiplicative constant, of the MSETAR $_d(1)$ model,

$$\mathbf{X}_t = -\alpha|\mathbf{X}_{t-1}| + \epsilon_t, \tag{1}$$

with independent and identically distributed (iid) $\mathcal{N}_d(\mathbf{0}, \mathbf{I})$ innovations, as

$$\phi_d(\mathbf{x}\sqrt{1-\alpha^2})\Phi_d(-\alpha\mathbf{x}|\mathbf{I}), \quad \mathbf{x} \in \mathbb{R}^d, \tag{2}$$

where $\alpha \in (-1, 1)$ is a nonzero scalar, $|\mathbf{x}| = (|x_1|, \dots, |x_d|)^T$, $\mathbf{0}$ denotes a d -dimensional null vector, \mathbf{I} is a $d \times d$ identity matrix, $\phi_d(\mathbf{x})$ represents the probability density function (pdf) of the standard normal distribution, and $\Phi_d(\mathbf{x}|\Sigma)$ denotes the cumulative distribution function (cdf) of a normal distribution with mean $\mathbf{0}$ and positive definite covariance matrix Σ . We observed that (2) is of the form of a multivariate skew-normal density, proposed by Arellano-Valle and Genton (2005), whereas the general form of that, up to a multiplicative constant, is given by

$$\phi_{d_1} \{(\Psi + \Lambda\Lambda^T)^{-1/2}(\mathbf{x} - \boldsymbol{\xi})\} \Phi_{d_2} \{\Lambda^T(\Psi + \Lambda\Lambda^T)^{-1}(\mathbf{x} - \boldsymbol{\xi})|\mathbf{I} - \Lambda^T(\Psi + \Lambda\Lambda^T)^{-1}\Lambda\}, \quad \mathbf{x} \in \mathbb{R}^{d_1}, \quad (3)$$

where $\boldsymbol{\xi}$ denotes a d_1 -dimensional location vector, Ψ is a $d_1 \times d_1$ symmetric positive definite dispersion matrix, Λ represents a $d_1 \times d_2$ skewness matrix, and by the square root of a matrix, here and hereafter, we denote the symmetric positive definite square root of the matrix; see also Sahu et al. (2003) and Arellano-Valle et al. (2018). In the following year, Arellano-Valle and Azzalini (2006) generalized (3) to define ‘unified skew-normal’ density as

$$\frac{\phi_{d_1} \{\Psi^{-1/2}(\mathbf{x} - \boldsymbol{\xi})\} \Phi_{d_2} \{\Lambda^T(\mathbf{x} - \boldsymbol{\xi}) + \boldsymbol{\tau}|\boldsymbol{\Gamma}\}}{\Phi_{d_2}(\boldsymbol{\tau}|\boldsymbol{\Gamma} + \Lambda^T\Psi\Lambda)}, \quad \mathbf{x} \in \mathbb{R}^{d_1}, \quad (4)$$

where $\boldsymbol{\tau}$ denotes a d_2 -dimensional extension vector and $\boldsymbol{\Gamma}$ is a $d_2 \times d_2$ positive definite extension matrix. We refer to the distribution with multivariate skew-normal density (3) using the notation $\mathcal{SN}_{d_1, d_2}(\boldsymbol{\xi}, \Psi, \Lambda)$ (or $\mathcal{SN}_d(\boldsymbol{\xi}, \Psi, \Lambda)$ when $d_1 = d_2 = d$) and (4) using the notation $\mathcal{SUN}_{d_1, d_2}(\boldsymbol{\xi}, \Psi, \Lambda, \boldsymbol{\tau}, \boldsymbol{\Gamma})$ (or $\mathcal{SUN}_d(\boldsymbol{\xi}, \Psi, \Lambda, \boldsymbol{\tau}, \boldsymbol{\Gamma})$ when $d_1 = d_2 = d$). Thus, clearly, the density (2) is from the $\mathcal{SN}_d(\mathbf{0}, \mathbf{I}, -\frac{\alpha}{\sqrt{1-\alpha^2}}\mathbf{I})$ distribution.

While there is an obvious motivation for studying stationary marginal distributions that many authors worked on them and found interesting properties of the corresponding processes in the limit, we, in the following paragraph, brief the essence of them.

Largely, stationary marginal distributions, if they exist, can be thought of as the limiting distributions of the corresponding time series processes with iid innovations for any given starting distributions.

In other words, if there is enough evidence to believe that the process has been running for a long time, we can reasonably presume that the observations from that process follow its stationary marginal distribution. Now, if we have data from such a process and we know the exact form of the distribution from which it came, we can always conduct an extensive analysis to explore the data. Following this notion of the stationary marginal density, a model diagnostic is exemplified in Section 4. Moreover, the likelihood function, which may be used to estimate the parameters of a model, comprises the stationary marginal density of the underlying process as a noteworthy factor

$$\mathcal{L}(\text{parameters}|\mathbf{x}_1, \dots, \mathbf{x}_n) = f_{\mathbf{X}_1}(\mathbf{x}_1)f_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{x}_2|\mathbf{x}_1) \cdots f_{\mathbf{X}_n|\mathbf{X}_{n-1}, \dots, \mathbf{X}_1}(\mathbf{x}_n|\mathbf{x}_{n-1}, \dots, \mathbf{x}_1).$$

We, with the sole purpose of linking the stationary marginal densities with the skew-symmetric densities, extend (1) in two directions. First, we generalize the coefficient matrices $\mathcal{H}_1(\mathbf{A}) = \{\mathbf{A} : \mathbf{A} = \alpha \mathbf{I}, \alpha \in (-1, 1), \alpha \neq 0\}$ to $\mathcal{H}_2(\mathbf{A}) = \{\mathbf{A} : \rho(\mathbf{A}) < 1, \mathbf{A} \neq \mathbf{O}\}$, hence

$$\mathbf{X}_t = -\mathbf{A}|\mathbf{X}_{t-1}| + \epsilon_t, \tag{5}$$

keeping the innovation distribution, $\mathcal{N}_d(\mathbf{0}, \Sigma)$, unchanged to show that the stationary marginal density of (5) has the form (4); here \mathbf{O} represents a $d \times d$ zero matrix and $\rho(\mathbf{A})$ denotes the spectral radius of the matrix \mathbf{A} , that is, the maximum magnitude of its eigenvalues. The condition $\rho(\mathbf{A}) < 1$ is sufficient for the geometric ergodicity of (5) (see Addo, 2014) and also needed for the existence of the matrices $\mathbf{P} = \sum_{k=0}^{\infty} \mathbf{A}^k \Sigma (\mathbf{A}^k)^T$ and $\mathbf{L} = \mathbf{A}^T \Sigma^{-1} \mathbf{A} + \mathbf{P}^{-1}$, which are used later in this article. Furthermore, we show that the resulting stationary marginal density has the form (3) when $\mathbf{A} \in \tilde{\mathcal{H}}_2(\mathbf{A}) = \{\mathbf{A} : \rho(\mathbf{A}) < 1, \det(\mathbf{A}) \neq 0, \mathbf{L} \succcurlyeq \mathbf{O}\} \subseteq \mathcal{H}_2(\mathbf{A})$, where $\mathbf{L} \succcurlyeq \mathbf{O}$ means that all the entries of \mathbf{L} are non-negative. In addition, for a given skew-normal density of the form (3) with $d_1 = d_2 = d$, we associate a coefficient matrix $\mathbf{A} \in \tilde{\mathcal{H}}_2(\mathbf{A})$ to (5) so that its stationary marginal density becomes identical to the given density. Clearly, these two results jointly characterize the stationary marginal densities of the processes under consideration.

Second, we generalize the innovation distribution $\mathcal{N}_d(\mathbf{0}, \Sigma)$ to symmetric multivariate stable distributions to derive the stationary marginal density as a skew-symmetric multivariate stable density, keeping the coefficient matrices, $\mathcal{H}_1(\mathbf{A})$, unchanged. The skew-symmetric multivariate stable density is introduced and briefly studied in Section 3. Moreover, for a given skew-symmetric multivariate stable density, we confirm that an $\text{MSETAR}_d(1)$ model of the form (1), with stationary marginal density identical to the given density, can be found. Thus, combining these two results, we characterize the $\text{MSETAR}_d(1)$ models with $\mathbf{A} \in \mathcal{H}_1(\mathbf{A})$ and symmetric multivariate stable innovations. For an extensive study on multivariate stable distributions, see, for example, Press (1972), Nolan (2003), Fallahgoul et al. (2013), and Teimouri et al. (2017).

The structure of the rest of this article is as follows. Section 2 presents the results on stationary marginal densities of $\text{MSETAR}_d(1)$ processes with $\mathbf{A} \in \mathcal{H}_2(\mathbf{A})$ and multivariate normal innovations following a brief review of the theory proposed by Chan and Tong (1986). Section 3 characterizes the stationary marginal densities of $\text{MSETAR}_d(1)$ processes with $\mathbf{A} \in \mathcal{H}_1(\mathbf{A})$ and symmetric multivariate stable innovations following a brief introduction of multivariate stable and skew-symmetric multivariate stable distributions. We provide a simulation study for model diagnostic using the corresponding stationary marginal density in Section 4. Section 5 summarizes the findings of this article, and discusses its significance, limitations and possible future extensions.

2 Stationary marginal densities of $\text{MSETAR}_d(1)$ processes with multivariate normal innovations

2.1 Preliminaries

Chan and Tong (1986) considered the following difference equation:

$$\mathbf{X}_t = \hat{\mathbf{T}}(\mathbf{X}_{t-1}) + \epsilon_t, \quad (6)$$

where ϵ_t 's are iid with pdf f_d ; here f_d is Borel-measurable and $\hat{\mathbf{T}}$ is a 'restriction' on \mathbf{T} , where $\mathbf{T} :$

$\mathbb{R}^d \rightarrow \mathbb{R}^d$ is a Borel-measurable function. To understand the ‘restriction’, consider a subgroup G of all the invertible $d \times d$ matrices with entries over \mathbb{R} that contains invertible linear automorphisms on \mathbb{R}^d , and a Borel set $\mathcal{R} \subseteq \mathbb{R}^d$ such that $\mathcal{R} \cap \langle \mathbf{x} \rangle$ is a singleton set for all $\mathbf{x} \in \mathbb{R}^d$. Here $\langle \mathbf{x} \rangle$ denotes the orbit of \mathbf{x} under G . Then, $\mathbf{T}|_{\mathcal{R}}$ is the restriction of \mathbf{T} on \mathcal{R} , and the extension of $\mathbf{T}|_{\mathcal{R}}$ with respect to G is $\hat{\mathbf{T}}(\mathbf{x}) = \mathbf{T}|_{\mathcal{R}}(\mathbf{y})$, where $\langle \mathbf{y} \rangle = \langle \mathbf{x} \rangle$ and $\mathbf{y} \in \mathcal{R}$.

We know that the stationary marginal density of (6) is the solution to the integral equation in $h_{d,1}(\cdot)$

$$h_{d,1}(\mathbf{y}) = \int_{\mathbb{R}^d} f_d\{\mathbf{y} - \hat{\mathbf{T}}(\mathbf{x})\} h_{d,1}(\mathbf{x}) \, d\mathbf{x}. \quad (7)$$

Under fairly general conditions on G , f_d , and \mathbf{T} , Chan and Tong (1986) showed that if $h_d(\cdot)$ is a non-negative integrable solution to the integral equation in $h_{d,2}(\cdot)$

$$h_{d,2}(\mathbf{y}) = \int_{\mathbb{R}^d} f_d\{\mathbf{y} - \mathbf{T}(\mathbf{x})\} h_{d,2}(\mathbf{x}) \, d\mathbf{x},$$

then $\bar{h}_d(\mathbf{y}) = \int_{\mathcal{R}} f_d\{\mathbf{y} - \mathbf{T}(\mathbf{x})\} h_d(\mathbf{x}) \nu(d\mathbf{x})$ is an integrable solution to (7), where ν is a measure on \mathcal{R} . That is, $c \cdot \bar{h}_d(\cdot)$ is the stationary marginal density of (6), c being a suitable normalizing constant. To illustrate the above theory, Chan and Tong (1986) showed that the stationary marginal density of (1) is (2).

2.2 Stationary marginal density for the case of multivariate normal innovations

Theorem 1. *The MSETAR_d(1) process (5), where $\epsilon_t \stackrel{iid}{\sim} \mathcal{N}_d(\mathbf{0}, \Sigma)$ and $\mathbf{A} \in \mathcal{H}_2(\mathbf{A})$, has the stationary marginal distribution*

$$SUN_d(\mathbf{0}, \mathbf{P}, -\Sigma^{-1}\mathbf{A}\mathbf{L}^{-1}, \mathbf{0}, \mathbf{L}^{-1}),$$

where $\mathbf{P} = \sum_{k=0}^{\infty} \mathbf{A}^k \Sigma (\mathbf{A}^k)^T$ and $\mathbf{L} = \mathbf{A}^T \Sigma^{-1} \mathbf{A} + \mathbf{P}^{-1}$.

Before going to the proof of Theorem 1, we put two lemmas for convenience.

Lemma 1. *The matrix \mathbf{P} exists if $\rho(\mathbf{A}) < 1$.*

Proof of Lemma 1. Consider the following two facts.

1. The geometric series $\sum_{k=0}^{\infty} \mathbf{M}^k$ converges if and only if $\rho(\mathbf{M}) < 1$.
2. The eigenvalues of $\mathbf{M}_1 \otimes \mathbf{M}_2$ are the products of all eigenvalues of \mathbf{M}_1 with all eigenvalues of \mathbf{M}_2 (see Schacke, 2004); here “ \otimes ” represents the Kronecker product. Thus, if $\rho(\mathbf{M}_1) < 1$ and $\rho(\mathbf{M}_2) < 1$, then $\rho(\mathbf{M}_1 \otimes \mathbf{M}_2) < 1$.

Therefore, since $\rho(\mathbf{A}) < 1$, we have

$$\sum_{k=0}^{\infty} (\mathbf{A} \otimes \mathbf{A})^k < \infty.$$

This implies $\boldsymbol{\sigma} \sum_{k=0}^{\infty} (\mathbf{A} \otimes \mathbf{A})^k < \infty$, where $\boldsymbol{\sigma} = \text{vec}(\boldsymbol{\Sigma})$; here $\text{vec}(\boldsymbol{\Sigma})$ represents the vectorization of the matrix $\boldsymbol{\Sigma}$. Lastly, by de-vectorizing, we get $\mathbf{P} = \sum_{k=0}^{\infty} \mathbf{A}^k \boldsymbol{\Sigma} (\mathbf{A}^k)^T < \infty$. □

Lemma 2. *The matrices \mathbf{P} and \mathbf{L} are both positive definite.*

Proof of Lemma 2. Since $\boldsymbol{\Sigma}$ is positive definite, we note that, for all $\mathbf{x} \neq \mathbf{0}$,

$$\mathbf{x}^T \mathbf{P} \mathbf{x} = \sum_{k=0}^{\infty} \mathbf{x}^T \mathbf{A}^k \boldsymbol{\Sigma} (\mathbf{A}^k)^T \mathbf{x} = \mathbf{x}^T \boldsymbol{\Sigma} \mathbf{x} + \sum_{k=1}^{\infty} (\mathbf{z}^{(k)})^T \boldsymbol{\Sigma} \mathbf{z}^{(k)} > 0,$$

where $\mathbf{z}^{(k)} = (\mathbf{A}^k)^T \mathbf{x}$. That is, \mathbf{P} is positive definite and, therefore, \mathbf{P}^{-1} is also positive definite.

Finally, we have, for all $\mathbf{x} \neq \mathbf{0}$,

$$\mathbf{x}^T \mathbf{L} \mathbf{x} = (\mathbf{A} \mathbf{x})^T \boldsymbol{\Sigma}^{-1} \mathbf{A} \mathbf{x} + \mathbf{x}^T \mathbf{P}^{-1} \mathbf{x} > 0,$$

which implies \mathbf{L} is also positive definite. □

Proof of Theorem 1. The stationary marginal distribution of the multivariate AR(1) process

$$\mathbf{X}_t = -\mathbf{A} \mathbf{X}_{t-1} + \boldsymbol{\epsilon}_t \quad \text{or} \quad \mathbf{X}_t = \boldsymbol{\Pi}(\mathbf{B})^{-1} \boldsymbol{\epsilon}_t$$

is $\mathcal{N}_d(\mathbf{0}, \mathbf{P})$, where \mathbf{B} is such that $\mathbf{B}\mathbf{X}_t = \mathbf{X}_{t-1}$ and $\Pi(\mathbf{B}) = \mathbf{I} + \mathbf{A}\mathbf{B}$ denotes the finite lag operator; see Kasparis (2008). Thus, by virtue of the theorem provided by Chan and Tong (1986) (take $\mathbf{T}(\mathbf{x}) = -\mathbf{A}\mathbf{x}$ and $\mathcal{R} = \{\mathbf{x} : x_i \geq 0, 1 \leq i \leq d\}$), the stationary marginal density of (5), up to a multiplicative constant, becomes

$$\begin{aligned}
& \int_{\mathbb{R}_+^d} e^{-\frac{1}{2}(\mathbf{y} + \mathbf{A}\mathbf{x})^T \boldsymbol{\Sigma}^{-1}(\mathbf{y} + \mathbf{A}\mathbf{x}) - \frac{1}{2}\mathbf{x}^T \mathbf{P}^{-1}\mathbf{x}} \, d\mathbf{x} \\
&= e^{-\frac{1}{2}\mathbf{y}^T \boldsymbol{\Sigma}^{-1}\mathbf{y}} \int_{\mathbb{R}_+^d} e^{-\frac{1}{2}(\mathbf{x}^T \mathbf{L}\mathbf{x} + 2\mathbf{x}^T \mathbf{A}^T \boldsymbol{\Sigma}^{-1}\mathbf{y})} \, d\mathbf{x} \quad [\because \mathbf{L} = \mathbf{A}^T \boldsymbol{\Sigma}^{-1} \mathbf{A} + \mathbf{P}^{-1}] \\
&= e^{-\frac{1}{2}\mathbf{y}^T (\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} \mathbf{A} \mathbf{L}^{-1} \mathbf{A}^T \boldsymbol{\Sigma}^{-1}) \mathbf{y}} \int_{\mathbb{R}_+^d} e^{-\frac{1}{2}(\mathbf{x} + \mathbf{L}^{-1} \mathbf{A}^T \boldsymbol{\Sigma}^{-1} \mathbf{y})^T \mathbf{L} (\mathbf{x} + \mathbf{L}^{-1} \mathbf{A}^T \boldsymbol{\Sigma}^{-1} \mathbf{y})} \, d\mathbf{x} \\
&\propto \phi_d(\mathbf{P}^{-1/2} \mathbf{y}) \Phi_d(-\mathbf{L}^{-1} \mathbf{A}^T \boldsymbol{\Sigma}^{-1} \mathbf{y} | \mathbf{L}^{-1}), \tag{8}
\end{aligned}$$

where \mathbb{R}_+^d denotes the set of all d -tuples of positive real numbers and the first factor of (8) is obtained by exploiting the Woodbury matrix identity

$$\boldsymbol{\Sigma}^{-1} - \mathbf{P}^{-1} = \{ \boldsymbol{\Sigma} + \boldsymbol{\Sigma} (\mathbf{P} - \boldsymbol{\Sigma})^{-1} \boldsymbol{\Sigma} \}^{-1} = \{ \boldsymbol{\Sigma} + \boldsymbol{\Sigma} (\mathbf{A} \mathbf{P} \mathbf{A}^T)^{-1} \boldsymbol{\Sigma} \}^{-1} = \boldsymbol{\Sigma}^{-1} \mathbf{A} \mathbf{L}^{-1} \mathbf{A}^T \boldsymbol{\Sigma}^{-1}.$$

We can find the multiplicative constant of the stationary marginal density using Lemma 2.1 of Arellano-Valle and Genton (2005). Furthermore, we note that when $\mathbf{A} = \mathbf{O}$, the stationary marginal density (8) reduces to the density of the innovation distribution, $\mathcal{N}_d(\mathbf{0}, \boldsymbol{\Sigma})$, as one may expect. \square

2.3 Characterization for a subclass

Theorem 2. *The stationary marginal density (8) of the process (5), with additional assumptions $\det(\mathbf{A}) \neq 0$ and $\mathbf{L} \succcurlyeq \mathbf{O}$, can be rewritten as the density of*

$$\mathcal{SN}_d(\mathbf{0}, \boldsymbol{\Sigma}, -\mathbf{A}\mathbf{P}^{1/2}).$$

Proof. Let $\Delta = \Sigma^{-1} \mathbf{A} \mathbf{L}^{-1} \mathbf{A}^T \Sigma^{-1}$ and, note that, by virtue of the additional condition, $\mathbf{L} \succcurlyeq \mathbf{O}$, we can rewrite (8) as

$$\phi_d \left\{ (\Sigma^{-1} - \Delta)^{1/2} \mathbf{y} \right\} \Phi_d \left(-\mathbf{L}^{-1/2} \mathbf{A}^T \Sigma^{-1} \mathbf{y} | \mathbf{I} \right). \quad (9)$$

To get (9) from (8), we use the result that for any matrix $\mathbf{M} \succcurlyeq \mathbf{O}$ and vectors \mathbf{x}_1 and \mathbf{x}_2 such that $\mathbf{x}_1 > \mathbf{x}_2$, the inequality $\mathbf{M} \mathbf{x}_1 \geq \mathbf{M} \mathbf{x}_2$ holds (see Karlin, 2014, p.475). Now, to write (9) in the form of (3) with $d_1 = d_2 = d$, we need to identify the dispersion matrix Ψ and the skewness matrix Λ of (9). To accomplish this, we further assume $\det(\Lambda) \neq 0$, $\mathbf{I} - \Lambda^T (\Psi + \Lambda \Lambda^T)^{-1} \Lambda \succcurlyeq \mathbf{O}$, and $\xi = \mathbf{0}$ of (3); we may assume so as long as we can solve for Ψ and Λ under these additional restrictions. Thus (3) with $d_1 = d_2 = d$ can also be rewritten as

$$\phi_d \left\{ (\Psi + \Lambda \Lambda^T)^{-1/2} \mathbf{x} \right\} \Phi_d \left[\left\{ \mathbf{I} - \Lambda^T (\Psi + \Lambda \Lambda^T)^{-1} \Lambda \right\}^{-1/2} \Lambda^T (\Psi + \Lambda \Lambda^T)^{-1} \mathbf{x} | \mathbf{I} \right]. \quad (10)$$

Now, to write (9) in the form of (10), we compare the arguments of ϕ_d and Φ_d from (9) with (10) to get

$$(\Sigma^{-1} - \Delta)^{1/2} = (\Psi + \Lambda \Lambda^T)^{-1/2} \quad \text{and} \quad (11)$$

$$-\mathbf{L}^{-1/2} \mathbf{A}^T \Sigma^{-1} = \left\{ \mathbf{I} - \Lambda^T (\Psi + \Lambda \Lambda^T)^{-1} \Lambda \right\}^{-1/2} \Lambda^T (\Psi + \Lambda \Lambda^T)^{-1}. \quad (12)$$

From (12), we have (while noting that this step introduces multiple solutions, which we will discard as we proceed)

$$(\Psi + \Lambda \Lambda^T)^{-1} \Lambda \left\{ \mathbf{I} - \Lambda^T (\Psi + \Lambda \Lambda^T)^{-1} \Lambda \right\}^{-1} \Lambda^T (\Psi + \Lambda \Lambda^T)^{-1} = \Sigma^{-1} \mathbf{A} \mathbf{L}^{-1} \mathbf{A}^T \Sigma^{-1},$$

which implies

$$\Delta = (\Psi + \Lambda \Lambda^T)^{-1} \left\{ (\Psi + \Lambda \Lambda^T) (\Lambda \Lambda^T)^{-1} - \mathbf{I} \right\}^{-1} = (\Psi + \Lambda \Lambda^T)^{-1} \Lambda \Lambda^T \Psi^{-1}.$$

Putting the expression of Δ back into (11) gives $\Sigma^{-1} - (\Psi + \Lambda\Lambda^T)^{-1}\Lambda\Lambda^T\Psi^{-1} = (\Psi + \Lambda\Lambda^T)^{-1}$, as both sides of the equation are positive definite. Hence, we have $\Sigma^{-1} = (\Psi + \Lambda\Lambda^T)^{-1}(I + \Lambda\Lambda^T\Psi^{-1}) = \Psi^{-1}$; that is, $\Psi = \Sigma$.

Furthermore, we have $\Sigma^{-1}\mathbf{A}\mathbf{L}^{-1}\mathbf{A}^T\Sigma^{-1} = \Delta = (\Psi + \Lambda\Lambda^T)^{-1}\Lambda\Lambda^T\Psi^{-1}$, which implies

$$\Sigma^{-1}\mathbf{A}\mathbf{L}^{-1}\mathbf{A}^T = (\Psi + \Lambda\Lambda^T)^{-1}\Lambda\Lambda^T = \{(\Lambda\Lambda^T)^{-1}\Psi + I\}^{-1}.$$

Hence,

$$(\Lambda\Lambda^T)^{-1}\Psi + I = (\mathbf{A}^T)^{-1}\mathbf{L}\mathbf{A}^{-1}\Sigma = (\mathbf{A}^T)^{-1}(\mathbf{A}^T\Sigma^{-1}\mathbf{A} + \mathbf{P}^{-1})\mathbf{A}^{-1}\Sigma = I + (\mathbf{A}\mathbf{P}\mathbf{A}^T)^{-1}\Sigma.$$

Therefore, $\Lambda\Lambda^T = \mathbf{A}\mathbf{P}\mathbf{A}^T$; this implies

$$\Lambda = \mathbf{A}\mathbf{P}^{1/2}\mathbf{U}, \tag{13}$$

where \mathbf{U} is a unitary matrix. That is, $\mathbf{U}\mathbf{U}^* = \mathbf{U}^*\mathbf{U} = I$, where ‘*’ denotes the conjugate transpose of the matrix. As the matrix \mathbf{U} is constant with respect to both \mathbf{A} and Σ and the equation (13) shall hold for all plausible \mathbf{A} and Σ , we consider specific choices of \mathbf{A} and Σ to get $\mathbf{U} = -I$, which implies $\Lambda = -\mathbf{A}\mathbf{P}^{1/2}$. \square

Next, we present two corollaries, which exhibit some specific cases of Theorem 2.

Corollary 1. *The process (5) with $\mathbf{A} = \text{diag}(\alpha_1, \dots, \alpha_d)$ and $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_d^2)$ has the stationary marginal distribution*

$$\mathcal{SN}_d\left(\mathbf{0}, \Sigma, -\mathbf{A}\{(I - \mathbf{A}^2)^{-1}\Sigma\}^{1/2}\right).$$

Proof. In this particular case, $\mathbf{P} = \sum_{k=0}^{\infty} \mathbf{A}^k \Sigma (\mathbf{A}^k)^T = \Sigma [I + \mathbf{A}^2 + \dots] = \Sigma (I - \mathbf{A}^2)^{-1}$. Substituting \mathbf{P} in Theorem 2 results in the expression above. \square

Corollary 2. *The process (5) with $\mathbf{A} = \alpha \mathbf{I}$ has the stationary marginal distribution*

$$\mathcal{SN}_d \left(\mathbf{0}, \boldsymbol{\Sigma}, \frac{-\alpha}{\sqrt{1-\alpha^2}} \boldsymbol{\Sigma}^{1/2} \right).$$

The following theorem, along with Theorem 2, characterizes the stationary marginal densities of the $\text{MSETAR}_d(1)$ processes with $\mathbf{A} \in \tilde{\mathcal{H}}_2(\mathbf{A})$ and multivariate normal innovations.

Theorem 3. *For a given multivariate skew-normal density of the form (3) from $\mathcal{SN}_{d_1, d_2}(\boldsymbol{\xi}, \boldsymbol{\Psi}, \boldsymbol{\Lambda})$ with $d_1 = d_2 = d$, $\boldsymbol{\xi} = \mathbf{0}$, $\det(\boldsymbol{\Lambda}) \neq 0$, and $\mathbf{I} - \boldsymbol{\Lambda}^T(\boldsymbol{\Psi} + \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T)^{-1}\boldsymbol{\Lambda} \succcurlyeq \mathbf{O}$, one can associate the following $\text{MSETAR}_d(1)$ process, which has stationary marginal density identical to the given density,*

$$\mathbf{X}_t = -\boldsymbol{\Lambda}(\boldsymbol{\Psi} + \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T)^{-1/2}|\mathbf{X}_{t-1}| + \boldsymbol{\epsilon}_t,$$

where $\boldsymbol{\epsilon}_t \stackrel{iid}{\sim} \mathcal{N}_d(\mathbf{0}, \boldsymbol{\Psi})$.

Proof. To prove Theorem 3, we assume $\det(\mathbf{A}) \neq 0$ and $\mathbf{L} \succcurlyeq \mathbf{O}$ of the process (5); we may assume so as long as we can find a solution to equations (11) and (12) for \mathbf{A} and $\boldsymbol{\Sigma}$ under these additional conditions. From the proof of Theorem 2, it is clear that $\boldsymbol{\Sigma} = \boldsymbol{\Psi}$. Therefore, it only remains to solve the equation

$$\sum_{k=1}^{\infty} \mathbf{A}^k \boldsymbol{\Sigma} (\mathbf{A}^k)^T = \boldsymbol{\Lambda} \boldsymbol{\Lambda}^T \quad (14)$$

for \mathbf{A} . Equation (14), along with the constraint $\boldsymbol{\Sigma} = \boldsymbol{\Psi}$, implies $\sum_{k=1}^{\infty} \mathbf{A}^k \boldsymbol{\Psi} (\mathbf{A}^k)^T = \boldsymbol{\Lambda} \boldsymbol{\Lambda}^T$. Vectorizing both sides of (14) produces $\sum_{k=1}^{\infty} \mathbf{Q}^k \boldsymbol{\psi} = \mathbf{v}$, where $\mathbf{Q} = \mathbf{A} \otimes \mathbf{A}$, $\boldsymbol{\psi} = \text{vec}(\boldsymbol{\Psi})$, and $\mathbf{v} = \text{vec}(\boldsymbol{\Lambda} \boldsymbol{\Lambda}^T)$. Thus, we have $\{(\mathbf{I} - \mathbf{Q})^{-1} - \mathbf{I}\} \boldsymbol{\psi} = \mathbf{v}$; that is,

$$\mathbf{Q}(\boldsymbol{\psi} + \mathbf{v}) = \mathbf{v}. \quad (15)$$

By de-vectorizing both sides of (15), we get $\mathbf{A}(\Psi + \Lambda\Lambda^T)\mathbf{A}^T = \Lambda\Lambda^T$. Therefore,

$$\mathbf{A} = \Lambda\mathbf{V}(\Psi + \Lambda\Lambda^T)^{-1/2},$$

where \mathbf{V} is a unitary matrix. We take $\mathbf{V} = -\mathbf{I}$ after investigating some particular choices of Ψ and Λ . □

Remark 1. *If the following partial MSETAR $_{d_1 \times d_2}(1)$ process*

$$\mathbf{X}_t = -\mathbf{A}_1|\mathbf{X}_{t-1}^{(1)}| + \epsilon_t, \quad (16)$$

where $\epsilon_t \stackrel{iid}{\sim} \mathcal{N}_{d_1}(\mathbf{0}, \Sigma)$, \mathbf{A}_1 is a $d_1 \times d_2$ non-null rectangular matrix with $d_1 > d_2$, and $\mathbf{X}_{t-1}^{(1)}$ denotes a vector comprising first d_2 components of the d_1 -dimensional vector \mathbf{X}_{t-1} , is of interest, then the stationary marginal density of that can be shown to have the form (4) using Theorem 1 above by choosing $\mathbf{A}_{d_1 \times d_1} = (\mathbf{A}_1 \ \mathbf{O}_{d_1 \times (d_1 - d_2)})$.

Remark 2. *A note on the measure of shape of the stationary marginal densities*

One of the most celebrated tools for quantifying the shape (i.e., skewness and kurtosis) of a multivariate distribution is Mardia's measures of multivariate skewness and kurtosis (Mardia, 1970, 1974), which are defined as $\gamma_{1,d} = E[(\mathbf{X} - \boldsymbol{\mu})^T \Theta^{-1}(\widetilde{\mathbf{X}} - \boldsymbol{\mu})]^3$ and $\gamma_{2,d} = E[(\mathbf{X} - \boldsymbol{\mu})^T \Theta^{-1}(\mathbf{X} - \boldsymbol{\mu})]^2$, respectively, where $\widetilde{\mathbf{X}}$ is an independent copy of \mathbf{X} , $\boldsymbol{\mu} = E[\mathbf{X}]$, and $\Theta = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T]$.

Azzalini et al. (2016) derived the measures of skewness and kurtosis of the multivariate skew-normal density of the form (3) with $d_1 = d_2 = d$ and $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_d)$ as

$$\gamma_{1,d} = \left(\frac{4 - \pi}{2}\right)^2 (\boldsymbol{\kappa}^{(3)})^T \Theta^{(-3)} \boldsymbol{\kappa}^{(3)}$$

and

$$\gamma_{2,d} = 2(\pi - 3) (\boldsymbol{\kappa}^{(2)})^T (\mathbf{I} \odot \Theta^{-1})^2 \boldsymbol{\kappa}^{(2)},$$

where $\boldsymbol{\kappa}^{(l)} = (\kappa_1^l, \dots, \kappa_d^l)^T$ with $\kappa_j = \sqrt{\frac{2}{\pi}}\lambda_j$, $\Theta^{(-3)}$ denotes the element-wise cube of Θ^{-1} , and ' \odot ' represents the Hadamard (component-wise) product.

In our case, when the skewness matrix is diagonal (see Corollary 1), $\kappa_j = -\sqrt{\frac{2}{\pi}}\frac{\alpha_j\sigma_j}{\sqrt{1-\alpha_j^2}}$ for $j = 1, \dots, d$ and $\Theta = \Sigma + (1 - \frac{2}{\pi}) \mathbf{A}^2(\mathbf{I} - \mathbf{A}^2)^{-1}\Sigma$. Thus, the two measures above reduce to

$$\begin{aligned}\gamma_{1,d} &= \left(\frac{4-\pi}{2}\right)^2 \left(\frac{2}{\pi}\right)^3 \sum_{j=1}^d \frac{\alpha_j^3\sigma_j^3}{(1-\alpha_j^2)^{3/2}} \cdot \frac{(1-\alpha_j^2)^3}{\sigma_j^6} \left(1 - \frac{2}{\pi}\alpha_j^2\right)^{-3} \cdot \frac{\alpha_j^3\sigma_j^3}{(1-\alpha_j^2)^{3/2}} \\ &= \frac{2(4-\pi)^2}{\pi^3} \sum_{j=1}^d \alpha_j^6 \left(1 - \frac{2}{\pi}\alpha_j^2\right)^{-3}\end{aligned}\quad (17)$$

and

$$\begin{aligned}\gamma_{2,d} &= 2(\pi-3) \left(\frac{2}{\pi}\right)^2 \sum_{j=1}^d \frac{\alpha_j^2\sigma_j^2}{(1-\alpha_j^2)} \cdot \frac{(1-\alpha_j^2)^2}{\sigma_j^4} \left(1 - \frac{2}{\pi}\alpha_j^2\right)^{-2} \cdot \frac{\alpha_j^2\sigma_j^2}{(1-\alpha_j^2)} \\ &= \frac{8(\pi-3)}{\pi^2} \sum_{j=1}^d \alpha_j^4 \left(1 - \frac{2}{\pi}\alpha_j^2\right)^{-2}.\end{aligned}\quad (18)$$

We note that, in this particular case, $\gamma_{1,d}$ and $\gamma_{2,d}$ do not depend on Σ .

Now, to illustrate the shape of the stationary marginal density of (1) with $\boldsymbol{\epsilon}_t \stackrel{iid}{\sim} \mathcal{N}_d(\mathbf{0}, \mathbf{I})$, we choose $d = 2$ and $\alpha_1 = \alpha_2 = \alpha$. Therefore, the Mardia's measures (17) and (18) simplify to

$$\gamma_{1,2} = \frac{4(4-\pi)^2}{\pi^3} \alpha^6 \left(1 - \frac{2}{\pi}\alpha^2\right)^{-3} \quad \text{and} \quad \gamma_{2,2} = \frac{16(\pi-3)}{\pi^2} \alpha^4 \left(1 - \frac{2}{\pi}\alpha^2\right)^{-2}.$$

Figure 1 displays the contour plots of the bivariate stationary marginal densities of (1) for (a) $\alpha = -0.9$ (upper left), (b) $\alpha = -0.5$ (upper right), (c) $\alpha = 0.5$ (lower left), and (d) $\alpha = 0.9$ (lower right) and, Figure 2 plots $\gamma_{1,2}$ and $\gamma_{2,2}$ against α which varies from -1 to 1 excluding the boundary values. From Figure 1, we observe that as α increases the stationary marginal density turns from positively (right) skewed into negatively (left) skewed. On the other hand, in Figure 2, the magnitudes of Mardia's measures of multivariate skewness and kurtosis increase as α departs from 0. As $\alpha \rightarrow \pm 1$, $\gamma_{1,2} \rightarrow 1.981132$ and $\gamma_{2,2} \rightarrow 1.73836$ independently of the choice of Σ as long as that is diagonal. These are the maximum attainable values of $\gamma_{1,2}$ and $\gamma_{2,2}$ for a density of the form (3) with $d_1 = d_2 = d$

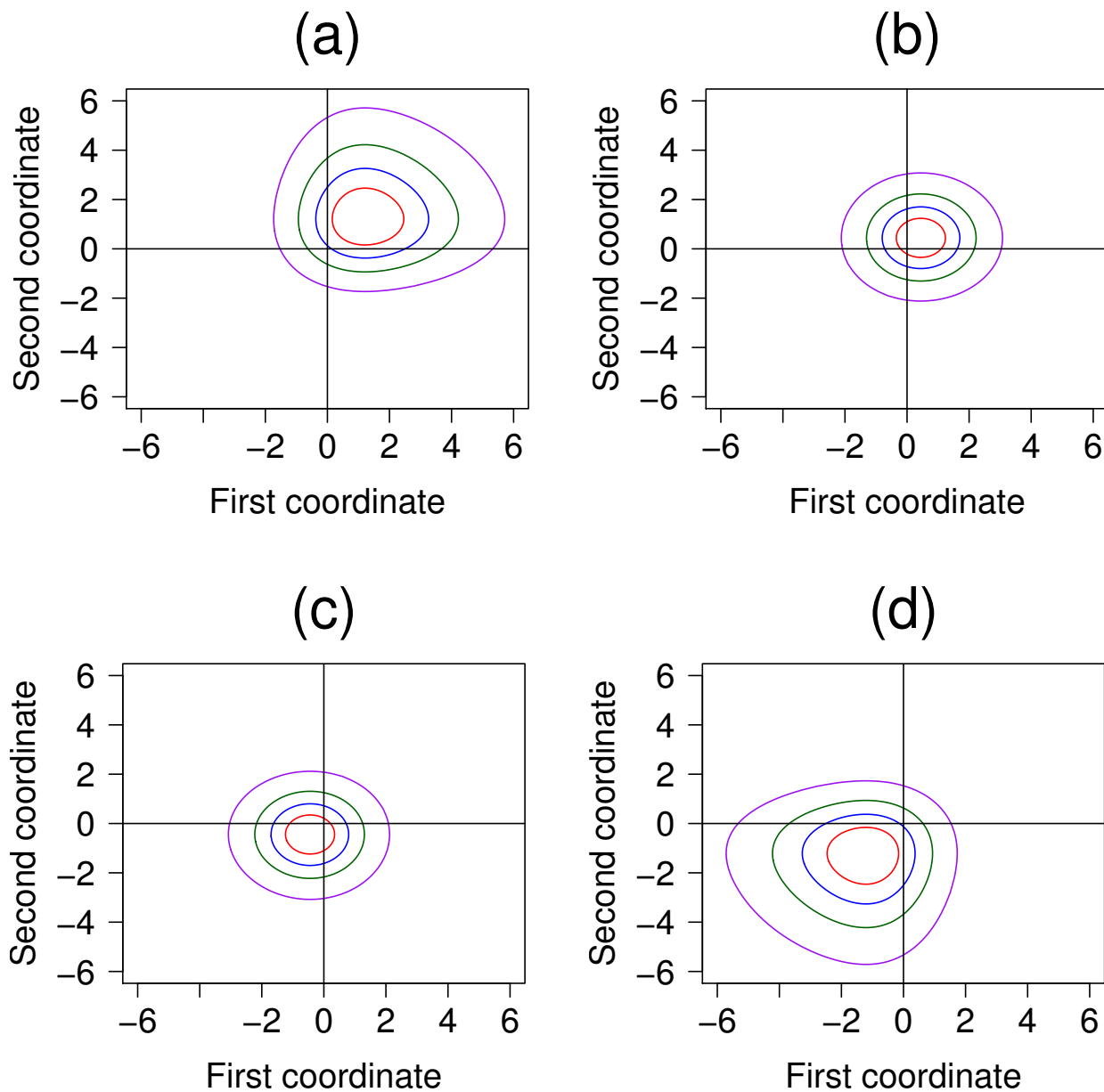


Figure 1: Contour plots of the stationary marginal densities of (1) with $\mathcal{N}_d(\mathbf{0}, \mathbf{I})$ innovations. The first and second coordinates of the arguments of the bivariate stationary marginal densities are plotted along the x and y axes, respectively. The contours (smallest to largest) cover approximately 25% (red), 50% (blue), 75% (green), and 95% (purple) of the stationary marginal densities. In panel (a): $\alpha = -0.9$; in (b): $\alpha = -0.5$; in (c): $\alpha = 0.5$; and in (d): $\alpha = 0.9$

and diagonal Λ ; see Azzalini et al. (2016). Going back to Figure 1, we note that the contour plots in panels (a) and (d) correspond to stationary marginal densities, which have the same values for $\gamma_{1,2}$ and $\gamma_{2,2}$; on the other hand, the contour plots in the panels (b) and (c) correspond to the stationary

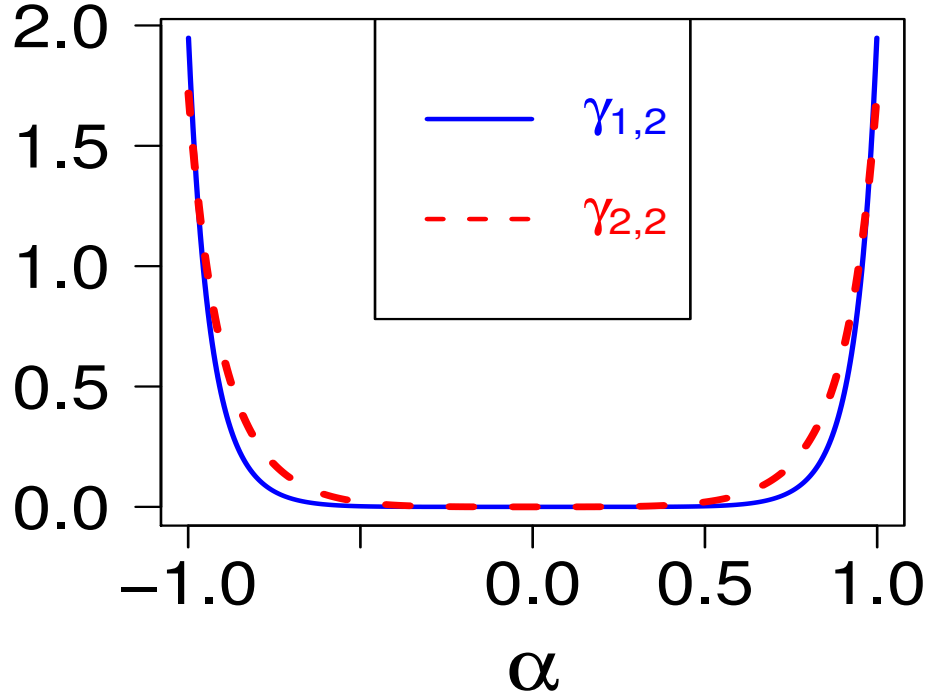


Figure 2: Plot of Mardia's measures of bivariate skewness (solid blue curve) and kurtosis (dashed red curve) of the stationary marginal densities of (1) with $\mathcal{N}_d(\mathbf{0}, \mathbf{I})$ innovations, against the coefficient α (from -1 to 1 , excluding the boundary values) of the model

marginal densities with the same values for $\gamma_{1,2}$ and $\gamma_{2,2}$.

3 Characterization of the stationary marginal densities of $\text{MSETAR}_d(1)$ processes with symmetric multivariate stable innovations

We start by recalling the notion of multivariate stable distributions; for a more detailed review, readers are referred to Nolan (2003) and Teimouri et al. (2017). Let \mathbf{X}_1 and \mathbf{X}_2 be independent copies of a $d \times 1$ random vector \mathbf{X} . Then, \mathbf{X} is said to be multivariate stable in \mathbb{R}^d if

$$a_1 \mathbf{X}_1 + a_2 \mathbf{X}_2 \stackrel{D}{=} a \mathbf{X} + \mathbf{b} \quad (19)$$

for any positive scalars a_1 and a_2 , some positive scalar $a = (a_1^\beta + a_2^\beta)^{1/\beta}$, and some real vector \mathbf{b} ; here ‘ $\stackrel{\text{D}}{=}$ ’ means equal in distribution. The distribution is said to be strictly multivariate stable if (19) holds with $\mathbf{b} = \mathbf{0}$. Here, the parameter $\beta \in (0, 2]$ is called the *index of stability* or *characteristic exponent*; see Nolan (2003).

In most scenarios, we expect the innovations to have a distribution that is centrally symmetric about $\mathbf{0}$. Therefore, hereafter we limit our study to symmetric (about $\mathbf{0}$), strictly multivariate stable distributions and omit the term “strictly” for convenience.

Before introducing the skew-symmetric multivariate stable distributions, we pause to note a construction of skew-symmetric distributions; see Section 4.2 of Arellano-Valle et al. (2006). Consider two univariate (for simplicity) iid random variables, X and Y , with a symmetric density function $\frac{1}{\sigma} f\left(\frac{\cdot}{\sigma}\right)$ and define $Z \stackrel{\text{D}}{=} (X - kY | Y > 0)$. Then, the pdf of Z ,

$$g(z; \sigma^2, k) = \frac{\text{d}}{\text{d}z} P[X - kY \leq z | Y > 0],$$

where k is a scalar, introduces skewness. Clearly, as k increases from negative values to positive values, the distribution of Z turns from positively (right) skewed into negatively (left) skewed; when $k = 0$, the density function $g(z; \sigma^2, k)$ reduces to $\frac{1}{\sigma} f\left(\frac{z}{\sigma}\right)$. Furthermore, $g(z; \sigma^2, k)$, with the aid of Lebesgue’s dominated convergence theorem, simplifies to

$$\begin{aligned} g(z; \sigma^2, k) &= \frac{\text{d}}{\text{d}z} P[X - kY \leq z | Y > 0] \\ &= \frac{2}{\sigma^2} \frac{\text{d}}{\text{d}z} \int_0^\infty \int_{-\infty}^{z+ky} f\left(\frac{x}{\sigma}\right) \text{d}x f\left(\frac{y}{\sigma}\right) \text{d}y \\ &= \frac{2}{\sigma^2} \int_0^\infty f\left(\frac{z+ky}{\sigma}\right) f\left(\frac{y}{\sigma}\right) \text{d}y. \end{aligned}$$

As we generalize this construction to the multivariate (d -dimensional) case by taking $|\Sigma|^{-1/2} f_d(\Sigma^{-1/2} \mathbf{x})$ to be the pdf of a symmetric multivariate stable distribution with location parameter $\mathbf{0}$ and scale (or dispersion) parameter Σ , $\mathcal{SMS}_d(\mathbf{0}, \Sigma)$, we introduce the pdf of skew-symmetric multivariate stable

distribution (skew- $\mathcal{SMS}_d(\mathbf{0}, \Sigma, k)$) as

$$g_d(\mathbf{z}; \Sigma, k) = \frac{2^d}{|\Sigma|} \int_{\mathbb{R}_+^d} f_d \{ \Sigma^{-1/2}(\mathbf{z} + k\mathbf{y}) \} f_d(\Sigma^{-1/2}\mathbf{y}) \, d\mathbf{y}. \quad (20)$$

Some immediate properties of the skew- \mathcal{SMS} distribution are as follows:

1. $g_d(\mathbf{z}; \Sigma, 0) = |\Sigma|^{-1/2} f_d(\Sigma^{-1/2}\mathbf{z})$ for all \mathbf{z} ;
2. If $\mathbf{Z} \sim \text{skew-}\mathcal{SMS}_d(\mathbf{0}, \Sigma, k)$, then $-\mathbf{Z} \sim \text{skew-}\mathcal{SMS}_d(\mathbf{0}, \Sigma, -k)$;
3. If $\mathbf{Z}_1 \sim \text{skew-}\mathcal{SMS}_d(\mathbf{0}, \Sigma, k)$ and $\mathbf{Z}_2 \sim \mathcal{SMS}_d(\mathbf{0}, \Sigma)$ are independent random vectors, then $\mathbf{Z} = \mathbf{Z}_1 + \mathbf{Z}_2 \sim 2^{1/\beta} \text{skew-}\mathcal{SMS}_d(\mathbf{0}, \Sigma, 2^{-1/\beta}k)$.

Proof of Property 3. The cdf of the random vector \mathbf{Z} is

$$P(\mathbf{Z} \leq \mathbf{z}) = c \int_{\mathbf{z}_1} P(\mathbf{Z}_2 \leq \mathbf{z} - \mathbf{z}_1) \left[\int_{\mathbb{R}_+^d} f_d \{ \Sigma^{-1/2}(\mathbf{z}_1 + k\mathbf{y}) \} f_d(\Sigma^{-1/2}\mathbf{y}) \, d\mathbf{y} \right] d\mathbf{z}_1,$$

where c is a suitable normalizing constant. Thus, the pdf of \mathbf{Z} becomes

$$\begin{aligned} & c |\Sigma|^{-1/2} \int_{\mathbf{z}_1} f_d \{ \Sigma^{-1/2}(\mathbf{z} - \mathbf{z}_1) \} \left[\int_{\mathbb{R}_+^d} f_d \{ \Sigma^{-1/2}(\mathbf{z}_1 + k\mathbf{y}) \} f_d(\Sigma^{-1/2}\mathbf{y}) \, d\mathbf{y} \right] d\mathbf{z}_1 \\ &= c |\Sigma|^{-1/2} \int_{\mathbb{R}_+^d} f_d(\Sigma^{-1/2}\mathbf{y}) \left[\int_{\mathbf{z}_1} f_d \{ \Sigma^{-1/2}(\mathbf{z} - \mathbf{z}_1) \} f_d \{ \Sigma^{-1/2}(\mathbf{z}_1 + k\mathbf{y}) \} \, d\mathbf{z}_1 \right] d\mathbf{y} \\ &\propto |\Sigma|^{-1/2} \int_{\mathbb{R}_+^d} f_d(\Sigma^{-1/2}\mathbf{y}) f_{d,\mathbf{W}} \{ \Sigma^{-1/2}(\mathbf{z} + k\mathbf{y}) \} \, d\mathbf{y}, \end{aligned} \quad (21)$$

where $|\Sigma|^{-1/2} f_{d,\mathbf{W}}(\Sigma^{-1/2}\mathbf{w})$ is the density function of $\mathbf{W} \equiv \mathbf{W}_1 + \mathbf{W}_2$, with \mathbf{W}_1 and \mathbf{W}_2 being independent copies of \mathbf{Z}_2 . Furthermore, using (19), we have $\mathbf{W}_1 + \mathbf{W}_2 \stackrel{D}{=} 2^{1/\beta} \mathbf{Z}_2$. Therefore, the density of \mathbf{W} becomes $2^{-d/\beta} |\Sigma|^{-1/2} f_d(2^{-1/\beta} \Sigma^{-1/2} \mathbf{w})$, which proves the desired property when substituted into (21). \square

Remark 3. *The proof above changes the order of the integration using Tonelli's theorem (see Tonelli, 1909); we briefly summarize the theorem here. Consider two σ -finite measure spaces, $(\Omega_1, \mathcal{F}_1, \mu)$ and $(\Omega_2, \mathcal{F}_2, \nu)$, and a non-negative measurable function, $h_d : (\Omega_1 \times \Omega_2) \rightarrow [0, \infty]$. Then,*

$$\int_{\Omega_1} \left\{ \int_{\Omega_2} h_d(\mathbf{x}, \mathbf{y}) d\nu(\mathbf{y}) \right\} d\mu(\mathbf{x}) = \int_{\Omega_2} \left\{ \int_{\Omega_1} h_d(\mathbf{x}, \mathbf{y}) d\mu(\mathbf{x}) \right\} d\nu(\mathbf{y})$$

holds.

Remark 4. *Similarly to the generalization (4) by Arellano-Valle and Azzalini (2006), (20) can also be further generalized, up to a multiplicative constant, as*

$$g_{d_1, d_2}(\mathbf{z}; \boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \boldsymbol{\tau}, \boldsymbol{\Gamma}) = \int_{\mathbb{R}_+^{d_2}} f_{d_1} \{ \mathbf{z} + \boldsymbol{\Lambda} \mathbf{y} | \boldsymbol{\xi}, \boldsymbol{\Sigma} \} f_{d_2}(\mathbf{y} + \boldsymbol{\tau} | \boldsymbol{\Gamma}) d\mathbf{y},$$

and the corresponding distribution can be denoted by skew- $\mathcal{SMS}_{d_1, d_2}(\boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \boldsymbol{\tau}, \boldsymbol{\Gamma})$. Here, $\boldsymbol{\xi}$ denotes a d_1 -dimensional location vector, $\boldsymbol{\Sigma}$ is a $d_1 \times d_1$ symmetric positive definite dispersion matrix, $\boldsymbol{\Lambda}$ represents a $d_1 \times d_2$ skewness matrix, $\boldsymbol{\tau}$ denotes a d_2 -dimensional extension vector, and $\boldsymbol{\Gamma}$ is a $d_2 \times d_2$ positive definite extension matrix.

Theorem 4. *The $\text{MSETAR}_d(1)$ process (1), where $\boldsymbol{\epsilon}_t \stackrel{iid}{\sim} \mathcal{SMS}_d(\mathbf{0}, \boldsymbol{\Sigma})$ and $\alpha \in (-1, 1)$ is a nonzero scalar, has stationary marginal distribution*

$$\text{skew-}\mathcal{SMS}_d(\mathbf{0}, \boldsymbol{\Sigma}, \alpha(1 - |\alpha|^\beta)^{-1/\beta}).$$

Proof. We note that the multivariate autoregressive model

$$\mathbf{X}_t = -\alpha \mathbf{X}_{t-1} + \boldsymbol{\epsilon}_t \tag{22}$$

can be expressed as $\mathbf{X}_t = (\mathbf{I} + \alpha \mathbf{B})^{-1} \boldsymbol{\epsilon}_t$. Thus, if $\boldsymbol{\epsilon}_t \stackrel{iid}{\sim} \mathcal{SMS}_d(\mathbf{0}, \boldsymbol{\Sigma})$, then, using the definition of

multivariate stable distribution as in (19), we have

$$\mathbf{X}_t = (\mathbf{I} + \alpha \mathbf{B})^{-1} \boldsymbol{\epsilon}_t = \boldsymbol{\epsilon}_t - \alpha \boldsymbol{\epsilon}_{t-1} - \alpha^2 \boldsymbol{\epsilon}_{t-2} - \cdots \stackrel{\text{D}}{=} g(\alpha) \boldsymbol{\epsilon}_1,$$

where $g(\alpha) = (1 + |\alpha|^\beta + |\alpha|^{2\beta} + \cdots)^{1/\beta} = (1 - |\alpha|^\beta)^{-1/\beta}$. Therefore, (22) has the stationary marginal density $\{g(\alpha)\}^{-d} |\boldsymbol{\Sigma}|^{-1/2} f_d [\{g(\alpha)\}^{-1} \boldsymbol{\Sigma}^{-1/2} \mathbf{x}]$. Hence, using the theorem provided by Chan and Tong (1986), the stationary marginal density, up to a multiplicative constant, of (1) reduces to the form of a skew- \mathcal{SMS} density:

$$\begin{aligned} & \{g(\alpha)\}^{-d} |\boldsymbol{\Sigma}|^{-1/2} \int_{\mathbb{R}_+^d} f_d \{\boldsymbol{\Sigma}^{-1/2} (\mathbf{y} + \alpha \mathbf{x})\} f_d [\{g(\alpha)\}^{-1} \boldsymbol{\Sigma}^{-1/2} \mathbf{x}] d\mathbf{x} \\ &= |\boldsymbol{\Sigma}|^{-1/2} \int_{\mathbb{R}_+^d} f_d [\boldsymbol{\Sigma}^{-1/2} \{\mathbf{y} + \alpha g(\alpha) \mathbf{x}\}] f_d (\boldsymbol{\Sigma}^{-1/2} \mathbf{x}) d\mathbf{x}. \end{aligned} \quad (23)$$

We note that when $\alpha = 0$, the stationary marginal density (23) reduces to the density of the innovation distribution $\mathcal{SMS}_d(\mathbf{0}, \boldsymbol{\Sigma})$. \square

Theorem 5 below, along with Theorem 4, characterizes the stationary marginal densities of $\text{MSETAR}_d(1)$ processes with $\mathbf{A} \in \mathcal{H}_1(\mathbf{A})$ and \mathcal{SMS} innovations.

Theorem 5. *For a given density from skew- $\mathcal{SMS}_d(\mathbf{0}, \boldsymbol{\Sigma}, k)$ with nonzero scalar k , one can associate the following $\text{MSETAR}_d(1)$ process, which has stationary marginal density identical to the given density,*

$$\mathbf{X}_t = -k \left(\frac{1}{1 + |k|^\beta} \right)^{1/\beta} |\mathbf{X}_{t-1}| + \boldsymbol{\epsilon}_t,$$

where $\boldsymbol{\epsilon}_t \stackrel{iid}{\sim} \mathcal{SMS}_d(\mathbf{0}, \boldsymbol{\Sigma})$.

Proof. From (23), it is clear that having a solution to the equation $\alpha g(\alpha) = k$ of α suffices to prove the theorem. A simple calculation shows that $\alpha = k \left(\frac{1}{1 + |k|^\beta} \right)^{1/\beta}$ satisfies the equation $\alpha \frac{1}{(1 - |\alpha|^\beta)^{1/\beta}} = k$. \square

Remark 5. *The multivariate Cauchy distribution with location parameter $\mathbf{0}$ ($\beta = 1$) and the multivariate normal distribution with mean vector $\mathbf{0}$ ($\beta = 2$) are two examples of SMS distributions with closed-form expressions of their densities. For the multivariate normal case, the stationary marginal density of (1) obtained in Theorem 4 coincides with the form obtained in Corollary 2 of Section 2.*

4 Simulation study for model diagnostic using stationary marginal density

We simulate $n = 5000$ data points $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ from (5) with $\epsilon_t \stackrel{iid}{\sim} \mathcal{N}_d(\mathbf{0}, \Sigma)$, where

$$\mathbf{A} = \begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.5 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} 4 & 2 \\ 2 & 10 \end{bmatrix},$$

and “burned-in” an initial $l = 500$ of them. We note that \mathbf{A} is non-null, $\rho(\mathbf{A}) < 1$, and Σ is positive definite. Now, to estimate the parameters of (5), we maximize the likelihood,

$$\begin{aligned} \mathcal{L}(\mathbf{A}, \Sigma | \mathbf{x}_{l+1}, \dots, \mathbf{x}_n) &= f_{\mathbf{X}_{l+1}}(\mathbf{x}_{l+1}) f_{\mathbf{X}_{l+2} | \mathbf{X}_{l+1}}(\mathbf{x}_{l+2} | \mathbf{x}_{l+1}) \cdots f_{\mathbf{X}_n | \mathbf{X}_{n-1}}(\mathbf{x}_n | \mathbf{x}_{n-1}) \\ &= \frac{\phi_d(\mathbf{P}^{-1/2} \mathbf{x}_{l+1}) \Phi_d(-\mathbf{L}^{-1} \mathbf{A}^T \Sigma^{-1} \mathbf{x}_{l+1} | \mathbf{L}^{-1})}{\Phi_d(\mathbf{0} | \mathbf{L}^{-1} + \mathbf{L}^{-1} \mathbf{A}^T \Sigma^{-1} \mathbf{P} \Sigma^{-1} \mathbf{A} \mathbf{L}^{-1})} \prod_{j=l+2}^n |\Sigma|^{-1/2} \phi_d \{ \Sigma^{-1/2} (\mathbf{x}_j + \mathbf{A} \mathbf{x}_{j-1}) \}, \end{aligned}$$

with respect to \mathbf{A} and Σ to get the respective maximum likelihood estimates as

$$\hat{\mathbf{A}} = \begin{bmatrix} 0.89 & 0.18 \\ 0.09 & 0.49 \end{bmatrix} \quad \text{and} \quad \hat{\Sigma} = \begin{bmatrix} 3.91 & 2.03 \\ 2.03 & 10.40 \end{bmatrix}.$$

The normalizing constant of the stationary marginal density $f_{\mathbf{X}_{l+1}}(\mathbf{x}_{l+1})$, $\Phi_d(\mathbf{0} | \mathbf{L}^{-1} + \mathbf{L}^{-1} \mathbf{A}^T \Sigma^{-1} \mathbf{P} \Sigma^{-1} \mathbf{A} \mathbf{L}^{-1})$, is derived using Lemma 2.1 of Arellano-Valle and Genton (2005). Next, we estimate the stationary marginal density by plugging in $\hat{\mathbf{A}}$ and $\hat{\Sigma}$ to get $\hat{f}(\cdot)$.

Finally, we recall the understanding from the Introduction that if a model fits a data significantly well, then the data can be assumed to be drawn from the corresponding stationary marginal distribution.

That is, the contour plot of the estimated stationary marginal density shall reasonably approximate the scatter plot of the data. So, we present the scatter plot of the simulated data and the contour plot of $\hat{f}(\cdot)$ in the same frame to diagnose the underlying model; see the left panel of Figure 3. As, in the current scenario, we have simulated the data from (5) itself, the contour plot approximates the scatter plot reasonably well.

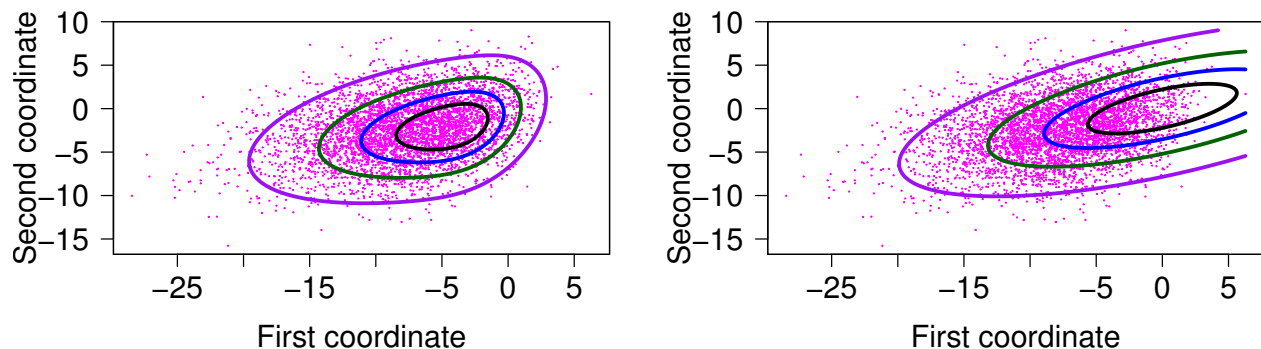


Figure 3: Scatter plot (magenta) of the simulated data and the contour plots of the stationary marginal densities corresponding to (5) (left) and (24) (right). The first and second coordinates of the arguments of the bivariate stationary marginal densities are plotted along the x and y axes, respectively. The contours (smallest to largest) cover approximately 25% (black), 50% (blue), 75% (darkgreen), and 95% (purple) of the corresponding stationary marginal densities.

In a data application framework, if the chosen model does not describe the data well, we would get a significantly different contour plot of the estimated stationary marginal density compared to the scatter plot of the data. For example, if instead we had fitted the classical vector autoregressive model

$$\mathbf{X}_t = -\tilde{\mathbf{A}}\mathbf{X}_{t-1} + \boldsymbol{\epsilon}_t, \quad (24)$$

where $\boldsymbol{\epsilon}_t \stackrel{iid}{\sim} \mathcal{N}_d(\mathbf{0}, \tilde{\boldsymbol{\Sigma}})$, to the data simulated above, we needed to maximize the likelihood

$$\begin{aligned} \mathcal{L}(\tilde{\mathbf{A}}, \tilde{\boldsymbol{\Sigma}} | \mathbf{x}_{l+1}, \dots, \mathbf{x}_n) &= f_{\mathbf{X}_{l+1}}(\mathbf{x}_{l+1}) f_{\mathbf{X}_{l+2} | \mathbf{X}_{l+1}}(\mathbf{x}_{l+2} | \mathbf{x}_{l+1}) \cdots f_{\mathbf{X}_n | \mathbf{X}_{n-1}}(\mathbf{x}_n | \mathbf{x}_{n-1}) \\ &= |\tilde{\mathbf{P}}|^{-1/2} \phi_d \left(\tilde{\mathbf{P}}^{-1/2} \mathbf{x}_{l+1} \right) \prod_{j=l+2}^n |\tilde{\boldsymbol{\Sigma}}|^{-1/2} \phi_d \left\{ \tilde{\boldsymbol{\Sigma}}^{-1/2} (\mathbf{x}_j + \tilde{\mathbf{A}} \mathbf{x}_{j-1}) \right\}, \end{aligned}$$

where $\tilde{\mathbf{P}} = \sum_{k=0}^{\infty} \tilde{\mathbf{A}}^k \tilde{\Sigma} \left(\tilde{\mathbf{A}}^k \right)^T$, with respect to the parameters $\tilde{\mathbf{A}}$ and $\tilde{\Sigma}$, and to estimate the stationary marginal density similarly as for the model (5). But, for this scenario, the contour plot of the estimated stationary marginal density does not approximate the scatter plot of the data well; see the right panel of Figure 3.

5 Discussion

In this study, we derived, explicitly, the stationary marginal density of an $\text{MSETAR}_d(1)$ process with coefficient matrix $\mathbf{A} \in \mathcal{H}_2(\mathbf{A})$ and normal innovations. We also characterized the stationary marginal densities of $\mathbf{X}_t = -\mathbf{A}|\mathbf{X}_{t-1}| + \epsilon_t$ with $\{\epsilon_t \stackrel{iid}{\sim} \mathcal{N}_d(\mathbf{0}, \Sigma) \ \& \ \mathbf{A} \in \tilde{\mathcal{H}}_2(\mathbf{A})\} \cup \{\epsilon_t \stackrel{iid}{\sim} \mathcal{SMS}_d(\mathbf{0}, \Sigma) \ \& \ \mathbf{A} \in \mathcal{H}_1(\mathbf{A})\}$.

Currently, threshold models are widely implemented in various fields, such as actuarial science, ecology, economics, epidemiology, and finance; see, for more details, Tong (2011). Many authors also studied processes with \mathcal{SUN} innovations (see Maleki and Arellano-Valle, 2017), mixture of scale mixtures of Gaussian innovations (see Maleki and Nematollahi, 2017), and mixture of scale-mixtures of skew-normal innovations (see Zarrin et al., 2019), which allowed them to fit a flexible model to accommodate skewness, heavy tails, multimodality, and stationarity simultaneously. Therefore, characterizing the stationary marginal density of (1) adds one more feather in the cap while noting the enormity of the progress in the class of multivariate stable distributions and the likelihood of encountering such distributions in practical situations. The stationary marginal density of (5) reduces to the form of a multivariate skew-normal density, which is a well-established class of densities; various properties, measures, and testing procedures for this class are studied by recent authors (e.g. Sahu et al., 2003; Arellano-Valle and Azzalini, 2006; Azzalini et al., 2016). Also, we note that, for $d = 1$, the multivariate stable distributions reduce to univariate stable (or sum stable) distributions (see Nolan, 2003). Besides the trivial example of the normal distribution, two examples of this univariate class of distributions that have closed form expressions are the Cauchy and the Lévy distributions. Thus, the stationary

marginal densities of MSETAR₁(1) (or SETAR(1)) processes with these as the innovation-distributions can easily be characterized following the results provided in this article. Furthermore, we remarked that the stationary marginal density of (16) has the form (4). Such models can be used by statisticians to reduce effort, time, and cost as these models predict the future \mathbf{X}_t using only partial information of the present \mathbf{X}_{t-1} ; in some situations, the trajectory $\{\mathbf{X}_t\}$ may be transformed using a linear transformation to get the principal components $\{\mathbf{Y}_t\}$ of $\{\mathbf{X}_t\}$, and the first d_2 components of \mathbf{Y}_{t-1} , which include a satisfactory percentage of the total variation, may be considered to minimize the loss of information.

Though we did not study them here, similar characterizations for other ‘symmetric’ autoregressive functions and higher order MSETAR processes may be discovered. Furthermore, we could extend the norming terms of the multivariate stable law to matrices. That is, we might consider the distribution of the random vector $\mathbf{X} \in \{\mathbf{X} : \mathbf{E}_1\mathbf{X}_1 + \mathbf{E}_2\mathbf{X}_2 \stackrel{D}{=} \mathbf{E}\mathbf{X} + \mathbf{f}; \mathbf{X}_1 \stackrel{D}{=} \mathbf{X}_2 \stackrel{D}{=} \mathbf{X}\}$, with certain conditions on the $d \times d$ matrices \mathbf{E}_1 , \mathbf{E}_2 , and \mathbf{E} , as the distribution of the innovations, and call this new class of distributions as ‘matrix stable’ distributions. Thus, having constructed the class of ‘matrix stable’ distributions, the stationary marginal density of (5) with ‘matrix stable’ innovations could be characterized following similar steps. Apart from the main theories, we also defined a skew-symmetric multivariate stable distribution in this article; the properties of this new class of distributions may be studied in the future.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

References

- Addo PM. 2014. Multivariate self-exciting threshold autoregressive models with exogenous input. *arXiv preprint arXiv: 1407.7738*.
- Anděl J, Bartoň T. 1986. A note on the threshold AR(1) model with Cauchy innovations. *Journal of Time Series Analysis* 7(1): 1–5.

- Anděl J, Netuka I, Zvára K. 1984. On threshold autoregressive processes. *Kybernetika* **20**(2): 89–106.
- Arellano-Valle RB, Branco MD, Genton MG. 2006. A unified view on skewed distributions arising from selections. *Canadian Journal of Statistics* **34**(4): 581–601.
- Arellano-Valle RB, Ferreira CS, Genton MG. 2018. Scale and shape mixtures of multivariate skew-normal distributions. *Journal of Multivariate Analysis* **166**: 98–110.
- Arellano-Valle RB, Genton MG. 2005. On fundamental skew distributions. *Journal of Multivariate Analysis* **96**(1): 93–116.
- Arellano-Valle RB, Azzalini A. 2006. On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics* **33**(3): 561–574.
- Arnold M, Günther R. 2001. Adaptive parameter estimation in multivariate self-exciting threshold autoregressive models. *Communications in Statistics-Simulation and Computation* **30**(2): 257–275.
- Azzalini A, Browne RP, Genton MG, McNicholas PD. 2016. On nomenclature for, and the relative merits of, two formulations of skew distributions. *Statistics and Probability Letters* **110**: 201–206.
- Azzalini A, Capitanio A. 2014. *The Skew-Normal and Related Families*. UK: Cambridge University Press.
- Baragona R, Cucina D. 2013. Multivariate self-exciting threshold autoregressive modeling by genetic algorithms. *Jahrbücher für Nationalökonomie und Statistik* **233**(1): 3–21.
- Chan K, Tong H. 1986. A note on certain integral equations associated with non-linear time series analysis. *Probability Theory and Related Fields* **73**(1): 153–158.
- Chan WS, Wong AC, Tong H. 2004. Some nonlinear threshold autoregressive time series models for actuarial use. *North American Actuarial Journal* **8**(4): 37–61.
- Fallahgoul H, Hashemiparast S, Fabozzi FJ, Kim YS. 2013. Multivariate stable distributions and generating densities. *Applied Mathematics Letters* **26**(3): 324–329.

- Genton MG. 2004. *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. CRC Press.
- Karlin S. 2014. *A First Course in Stochastic Processes*. Academic press.
- Kasparis I. 2008. A simple proof for the invertibility of the lag polynomial operator. *Discussion Paper, 2008-06, Department of Economics, University of Cyprus*.
- Loges W. 2004. The stationary marginal distribution of a threshold AR(1) process. *Journal of Time Series Analysis* **25**(1): 103–125.
- Maleki M, Arellano-Valle RB. 2017. Maximum a-posteriori estimation of autoregressive processes based on finite mixtures of scale-mixtures of skew-normal distributions. *Journal of Statistical Computation and Simulation* **87**(6): 1061–1083.
- Maleki M, Nematollahi A. 2017. Autoregressive models with mixture of scale mixtures of Gaussian innovations. *Iranian Journal of Science and Technology, Transactions A: Science* **41**(4): 1099–1107.
- Mardia KV. 1970. Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57**(3): 519–530.
- Mardia KV. 1974. Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhyā: The Indian Journal of Statistics, Series B* 115–128.
- Nolan J. 2003. *Stable Distributions: Models for Heavy-Tailed Data*. Birkhauser New York.
- Press S. 1972. Multivariate stable distributions. *Journal of Multivariate Analysis* **2**(4): 444–462.
- Sahu SK, Dey DK, Branco MD. 2003. A new class of multivariate skew distributions with applications to Bayesian regression models. *Canadian Journal of Statistics* **31**(2): 129–150.
- Schacke K. 2004. On the Kronecker product. *Master's thesis, University of Waterloo*.

- Teimouri M, Rezakhah S, Mohammadpour A. 2017. *U*-statistic for multivariate stable distributions. *Journal of Probability and Statistics* **2017**.
- Tonelli L. 1909. Sull'integrazione per parti. *Atti della Accademia Nazionale dei Lincei* **18**(2): 246–253.
- Tong H. 1978. On a threshold model in pattern recognition and signal processing. *Chen. Sijhoff and Noordhoff, Amsterdam*.
- Tong H. 1983. *Threshold Models in Non-Linear Time Series Analysis*. Springer Science & Business Media.
- Tong H. 1990. *Non-Linear Time Series: A Dynamical System Approach*. Oxford University Press.
- Tong H. 2011. Threshold models in time series analysis—30 years on. *Statistics and Its Interface* **4**(2): 107–118.
- Wong SF, Tong H, Siu TK, Lu Z. 2017. A new multivariate nonlinear time series model for portfolio risk measurement: the threshold copula-based TAR approach. *Journal of Time Series Analysis* **38**(2): 243–265.
- Zarrin P, Maleki M, Khodadai Z, Arellano-Valle RB. 2019. Time series models based on the unrestricted skew-normal process. *Journal of Statistical Computation and Simulation* **89**(1): 38–51.